AEROJEPA: LEARNING SEMANTIC LATENT REPRESENTATIONS FOR SCALABLE 3D AERODYNAMIC FIELD MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

High-fidelity CFD is essential for aerodynamic design, but repeated simulations are computationally expensive, motivating surrogate models for rapid evaluation across geometries and operating conditions. Most existing surrogates are designed for direct field regression, requiring the evaluation of millions of field points even when only an aerodynamic quantity or a localized region is needed, while their internal representations are not intended for direct use in downstream tasks. We introduce *AeroJEPA*, a framework inspired by joint-embedding predictive architectures that represents the problem in two distinct latent spaces: context tokens encode geometry, while predicted tokens encode the aerodynamic state. Both representations remain directly accessible for downstream tasks, such as linear readouts of design variables and aerodynamic quantities without decoding and integrating the full field. When spatial detail is needed, a continuous implicit decoder evaluates the field only at the requested coordinates while reusing the encoded geometry. We evaluate AeroJEPA on HiLiftAeroML, with multi-million-point fields, and SuperWing, which spans a broad family of transonic wings. Compared with state-of-the-art direct-regression surrogates, AeroJEPA trades peak full-field accuracy for compact, reusable representations. In our selective-decoding experiment, however, AeroJEPA substantially outperforms the evaluated direct-regression surrogates while avoiding predictions over the remainder of the aircraft. The learned representations further support controlled interpolation, concept-vector arithmetic, and preliminary constrained latent-space optimization. These results show how predictive representations can support aerodynamic analysis with or without full-field reconstruction.

1 INTRODUCTION

Aerodynamic design requires repeated evaluation of complex flow fields across different geometries and operating conditions. High-fidelity computational fluid dynamics (CFD) remains the primary tool for obtaining these fields, but its computational cost limits rapid design-space exploration and optimization. This has long motivated surrogate modeling for aerodynamic analysis (Forrester et al., 2006; 2008; Yondo et al., 2018), and more recently the development of neural surrogates, neural operators, and continuous field representations (Azizzadenesheli et al., 2024; Catalani et al., 2024; Duvall & Duraisamy, 2025).

Most existing approaches are designed primarily as direct geometry-to-field regressors. This is effective when the complete field is the desired output, but it has two practical limitations. First, obtaining a localized field or an integrated aerodynamic quantity can still require predicting millions of values over the complete CFD discretization. Second, although these models contain intermediate features, those representations are generally not explicitly organized or evaluated for reuse in downstream engineering tasks. A model that instead exposes compact representations of geometry and aerodynamics could support coefficient prediction, design-variable recovery, interpolation, and optimization (Vishwasrao et al., 2026; Vinuesa et al., 2026), without requiring full-field decoding at every stage. When spatial detail is required, the same representation could be decoded only at the coordinates relevant to the analysis.

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.

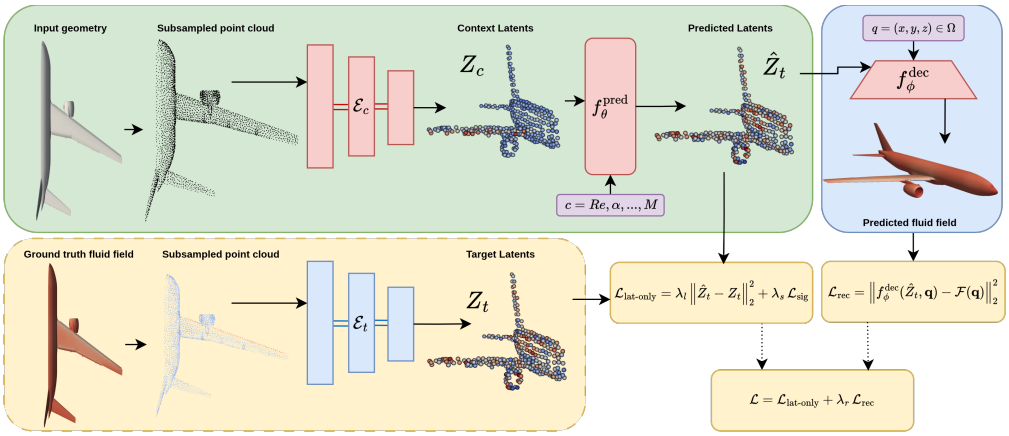

Figure 1: Overview of AeroJEPA. The context encoder maps the geometry point cloud to context tokens Z_c, which represent the geometry independently of the operating condition. The predictor conditions Z_c on the operating variables c (e.g., α, Re, and Mach number) to produce predicted tokens $\hat{Z}_t$ representing the aerodynamic state. During end-to-end training, the target encoder maps the ground-truth flow to target tokens Z_t, providing a flow-derived target through $\mathcal{L}_{\text{lat}}$; SIGReg $\mathcal{L}_{\text{sig}}$ regularizes the latent distribution, and the implicit decoder anchors the predicted tokens to the physical field through $\mathcal{L}_{\text{rec}}$. At inference, the target encoder is discarded. Context and predicted tokens can be used directly by lightweight downstream heads, while the decoder is evaluated only when and where field values are requested.

This objective lies at the intersection of aerodynamic surrogate modeling and predictive representation learning. Recent aerodynamic methods have demonstrated compact latent models, transformer-based surrogates, neural operators, and implicit decoders for high-dimensional CFD data (Solera-Rico et al., 2024; Francés-Belda et al., 2024; Eivazi et al., 2022; Choy et al., 2025; Wu et al., 2024a; Adams et al., 2025; Zou et al., 2026). These methods establish the value of compressed or continuous field representations, but their primary objective is typically field reconstruction or direct regression. Joint-Embedding Predictive Architectures (JEPAs), in contrast, learn by predicting a target representation from a context representation (Assran et al., 2023; Sobal et al., 2022). Recent methods such as LeJEPA and LeWorldModel use explicit latent regularization to stabilize this process, while related work has extended predictive representation learning to 3D, multimodal, and scientific settings (Balestriero & LeCun, 2025; Maes et al., 2026; Hu et al., 2024; Perera et al., 2025; Chen et al., 2025; Qu et al., 2026; Yee & Koh, 2026). However, this paradigm has not been developed around the distinct geometry and aerodynamic representations required for 3D aerodynamic analysis.

We introduce *AeroJEPA*, a framework inspired by joint-embedding predictive architectures that represents an aerodynamic problem in two distinct latent spaces. The context encoder produces tokens describing geometry, while the predictor combines these tokens with the operating conditions to produce tokens describing the corresponding aerodynamic state. This separation makes both representations explicitly accessible: lightweight heads can recover design variables from geometric information and aerodynamic quantities from predicted flow information without reconstructing and integrating the complete field. When local or full-field values are required, a continuous implicit neural representation (INR) decoder evaluates the predicted latent state at arbitrary query coordinates. The geometry is encoded once and its context can be reused across requested coordinates, subregions, and operating conditions.

We evaluate these capabilities on two complementary datasets. HiLiftAeroML contains high-fidelity boundary-layer fields with millions of points over realistic high-lift aircraft configurations, providing a demanding test of large-field and selective decoding. SuperWing spans a broad parametric family of transonic wings and enables a detailed evaluation of representation structure and downstream use. The results expose a deliberate trade-off: compared with direct-regression surrogates, AeroJEPA can sacrifice peak full-field accuracy in exchange for compact geometry and flow representations that are directly reusable. This reflects a broader trade-off between reconstruction fidelity and representation utility; we therefore evaluate AeroJEPA along both axes rather than solely as a field regressor. In a

selective-decoding experiment restricted to an outboard-wing region, however, AeroJEPA outperforms the evaluated direct-regression surrogates while avoiding predictions over the remainder of the aircraft. Across the two datasets, the context and predicted representations support linear readouts, controlled interpolation, concept-vector arithmetic, and preliminary constrained latent-space optimization.

Our contributions are threefold. First, we formulate an aerodynamic predictive-representation architecture that explicitly separates geometry context from the condition-dependent aerodynamic state. Second, we show that these frozen representations support downstream engineering readouts and structured latent-space operations without requiring full-field reconstruction. Third, we combine these representations with a continuous decoder that supports both complete field reconstruction and selective evaluation at requested coordinates, and characterize the resulting trade-off against direct-regression surrogates on two complementary aerodynamic benchmark datasets.

2 METHOD

AeroJEPA is a predictive representation-learning framework inspired by joint-embedding predictive architectures and tailored to 3D aerodynamic problems. It represents the input geometry and the corresponding aerodynamic state in two distinct token spaces. The geometry representation can be used directly for geometry-related downstream tasks, while the predicted aerodynamic representation supports physical readouts and optional continuous field decoding.

2.1 PROBLEM FORMULATION

We consider the problem of predicting steady-state 3D fluid fields around complex aerodynamic geometries. Let a given geometric design be defined by its boundary surface $\partial\Omega$, which interacts with a fluid over a spatial domain $\Omega \subset \mathbb{R}^3$. The geometry is discretized and represented as an unstructured point cloud of boundary condition (BC) points, $\mathcal{P} = \{x_i\}_{i=1}^{N_c}$, where each $x_i \in \partial\Omega$.

The flow is governed by a vector of operating or freestream conditions, denoted by c. Depending on the dataset, this vector may include quantities such as angle of attack α, Reynolds number Re, and Mach number. Let $\mathcal{D} \subseteq \overline{\Omega}$ denote the surface or volumetric query domain, and let $\mathcal{F} : \mathcal{D} \to \mathbb{R}^{d_f}$ denote the corresponding aerodynamic field, such as pressure, velocity, or wall-shear quantities.

Given $\mathcal{P}$ and c, our objective is to learn two complementary representations: a compact context representation of the geometry and a predicted representation of the corresponding aerodynamic state. The geometry representation should support geometry-related downstream tasks, while the predicted representation should support aerodynamic readouts without decoding and integrating the full field. When spatial information is required, the predicted representation should also retain sufficient information to evaluate the physical field at any requested query coordinate $q = (x, y, z) \in \mathcal{D}$.

2.2 THE AEROJEPA FRAMEWORK

The core design of AeroJEPA separates geometry encoding, aerodynamic-state prediction, and field decoding through a tokenized interface. The context encoder maps the geometry point cloud to context tokens Z_c. Conditioned on the operating variables, the predictor transforms these tokens into predicted aerodynamic tokens $\hat{Z}_t$. During training, the target encoder maps the ground-truth flow to target tokens Z_t. At inference, the target encoder is discarded, leaving the geometry-conditioned path that produces Z_c and $\hat{Z}_t$.

The framework admits several possible training strategies. In the end-to-end configuration used for the primary results, the context encoder, target encoder, predictor, and decoder are jointly optimized using latent-matching, latent-regularization, and reconstruction losses. In a two-stage alternative, inspired by a staged representation-learning strategy (Li et al., 2025), a flow autoencoder is first trained for reconstruction; its target encoder and decoder are subsequently frozen while the context encoder and predictor learn to match the frozen flow representation from geometry and operating conditions. A reconstruction-only alternative removes the target encoder and latent matching and trains the geometry-conditioned path using field reconstruction as its supervision. A decoder-free configuration could also be formulated using only latent matching and regularization, but this possibility is not

evaluated in this work. The frozen-target two-stage strategy is described as an alternative training procedure but is likewise not included in the experimental comparisons.

We use end-to-end training for the primary models because it learns the geometry representation, aerodynamic representation, and field decoder in a single optimization procedure. The frozen-target alternative requires a separate flow-autoencoder pretraining stage before the geometry-to-latent predictor can be trained, introducing an additional sequential training cost. End-to-end training also allows the target representation and geometry-conditioned predictor to co-adapt instead of requiring the predictor to reproduce a fixed latent space. On HiLiftAeroML, we assess the effect of latent matching through a matched comparison with reconstruction-only training in Appendix D.4.

At inference, context tokens can be used for geometry-related tasks, while predicted tokens can be used for aerodynamic readouts conditioned on the operating state. If field values are required, the predicted tokens are passed to the implicit decoder, which is evaluated only at the requested coordinates. A geometry can therefore be encoded once and reused across operating conditions, downstream readouts, and selective field queries.

2.3 Encoders: transforming physical space to latent space

AeroJEPA converts irregular geometry and flow point clouds into fixed-size token sets. This creates a compact interface whose size does not depend on the number of points in the original CFD field.

Context Encoder The context encoder, $\mathcal{E}_c$, receives an unstructured point-cloud representation of the geometry without mesh connectivity. Each case is first uniformly subsampled to a manageable input size, typically 8×10^3–131×10^3 points depending on the dataset. Farthest-point sampling (FPS) is then applied within this input cloud to select spatially distributed tokenizer centers. For surface-only settings, point coordinates are used as input features. For volumetric settings, the signed-distance-function value is additionally supplied. Other geometry-derived attributes such as surface normals could also be incorporated, but were not used during the development of this work.

Given the subsampled geometry cloud $\mathcal{P}$, the context encoder produces

$$Z_c = \mathcal{E}_c(\mathcal{P}) \in \mathbb{R}^{M \times d}, \tag{1}$$

where M is the number of spatial tokens and d is the token width. In the main experiments, HiLiftAeroML uses 3072 tokens of width 64, while SuperWing uses 512 tokens of width 128.

Target Encoder The target encoder, $\mathcal{E}_t$, is used only during training and operates on an independently sampled point cloud containing coordinates and ground-truth flow values. Depending on the dataset, this cloud may contain surface quantities such as C_p and wall-shear components, or volumetric quantities such as velocity and pressure. Target points are sampled independently from the geometry input, so the framework does not require pointwise correspondence between the context and target point clouds.

The target encoder produces the flow-derived tokens

$$Z_t = \mathcal{E}_t(\mathcal{F}) \in \mathbb{R}^{M \times d}, \tag{2}$$

which provide the latent-space target during end-to-end training.

In both encoders, tokenizer centers are selected by FPS within the subsampled input cloud. Local neighborhoods around these centers are aggregated with a lightweight message-passing network (Gilmer et al., 2017), and the resulting token set is refined with a point-transformer backbone. This maps irregular point clouds of varying size to fixed-size geometry and aerodynamic token sets.

2.4 Latent predictor network

The predictive core of AeroJEPA is the predictor network, f_θ^{pred}, parameterized by θ. Given the geometry context tokens Z_c and operating conditions c, it predicts the aerodynamic tokens

$$\hat{Z}_t = f_\theta^{\text{pred}}(Z_c, c). \tag{3}$$

The operating conditions are injected within the predictor, allowing one geometry encoding to support multiple operating states. The prediction stage produces a fixed-size aerodynamic representation whose cost is independent of the number of field coordinates subsequently queried. The predicted tokens can be passed either to lightweight downstream heads or to the field decoder.

2.5 Implicit neural representation (INR) decoder

To map the fluid latent states back into the physical domain, we employ an INR Decoder, f_ϕ^{dec}, parameterized by ϕ. Unlike voxel grids or fixed meshes, INRs provide a continuous, resolution-independent mapping by querying exact spatial coordinates.

The decoder takes as input a spatial query point $q = (x, y, z) \in \Omega$ alongside the predicted latents $\hat{Z}_t$. During training, the reconstruction query points are also subsampled from the original field and are chosen independently from both the context and target-encoder point sets. This preserves the decoupled setting and effectively trains the model in a masked fashion, encouraging the latent tokens to encode the underlying continuous field rather than a fixed subset of sampled locations. It outputs the local fluid state variables, such as velocity vector $\mathbf{u}$ and pressure p:

$$[\mathbf{u}(q), p(q)] = f_\phi^{\text{dec}}(\hat{Z}_t, q) \tag{4}$$

By conditioning a multi-layer perceptron (MLP) on the latent vector $\hat{Z}_t$, the INR Decoder acts as a continuous basis function for the aerodynamic field. This architecture is particularly advantageous for complex aircraft geometries, as it completely bypasses the need for volumetric meshing and allows for arbitrary physical resolution purely bounded by the query points provided.

2.6 Training objectives

The primary AeroJEPA configuration is trained jointly to align predicted and flow-derived target tokens, prevent representation collapse, and retain the ability to reconstruct the physical field. Following recent JEPA formulations, we use SIGReg (Balestriero & LeCun, 2025; Maes et al., 2026) as an explicit regularizer on the latent distribution.

For coupled end-to-end training, we optimize

$$\mathcal{L}_{\text{total}} = \lambda_\ell \mathcal{L}_{\text{lat}} + \lambda_r \mathcal{L}_{\text{rec}} + \lambda_s \mathcal{L}_{\text{sig}}. \tag{5}$$

The latent-matching term is

$$\mathcal{L}_{\text{lat}} = \left\| \hat{Z}_t - Z_t \right\|_2^2, \tag{6}$$

which aligns the predictor output with the target-encoder output token by token. The reconstruction term is

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{q \in \mathcal{D}} \left[\left\| f_\phi^{\text{dec}}(\hat{Z}_t, q) - \mathcal{F}(q) \right\|_2^2 \right], \tag{7}$$

with the supervised channels determined by the surface or volumetric prediction task.

The regularization term $\mathcal{L}_{\text{sig}}$ applies SIGReg to the latent representations through random low-dimensional projections toward an isotropic Gaussian reference distribution. Its role is to discourage dimensional collapse without requiring an exponential-moving-average teacher or stop-gradient training heuristic.

The primary checkpoints use the coupled end-to-end objective with $\lambda_\ell = 1.0$, $\lambda_r = 1.0$, and $\lambda_s = 0.001$. These values define the models used in the main experiments and are not claimed to be universally optimal. On SuperWing, Appendix D.4 examines the effect of the SIGReg weight and a

contrastive alternative. On HiLiftAeroML, it compares end-to-end AeroJEPA with reconstruction-only training of the same architecture.

3 COMPUTATIONAL EXPERIMENTS

We evaluate AeroJEPA across two complementary aerodynamic benchmarks, considering full-field accuracy and computational cost, the downstream utility of its geometry and aerodynamic representations, selective decoding, and preliminary latent-space optimization. The selective-decoding protocol, training-objective ablations, and detailed latent-optimization experiment are reported in Appendix D.2, Appendix D.4, and Appendix D.5, respectively.

3.1 DATASETS

We evaluate AeroJEPA on two complementary aerodynamic datasets. The first is the HiLiftAeroML dataset introduced by Ashton et al. (Ashton et al., 2026), which provides high-fidelity CFD data generated with solution-adapted WMLES for high-lift configurations. This dataset is particularly relevant for assessing whether the proposed latent-space prediction strategy can capture complex separated-flow phenomena on realistic aircraft geometries. The second is SuperWing (Yang et al., 2025), a large-scale transonic swept-wing dataset designed for data-driven aerodynamic design. SuperWing contains diverse parameterized wing geometries simulated over a broad range of operating conditions, making it suitable for evaluating generalization across geometry and flow-condition variations. Additional information on the geometry families, flow regimes, and the role of each dataset in our evaluation protocol is provided in Appendix B.

We compare AeroJEPA with three established aerodynamic surrogate architectures. FIGConvUNet provides a factorized convolutional approach for learning on high-resolution 3D fields (Choy et al., 2025), while Transolver is a widely adopted physics-attention model for neural PDE solving (Wu et al., 2024a). GeoTransolver extends this approach with geometry-aware attention and reports state-of-the-art performance on established industrial aerodynamic benchmarks (Adams et al., 2025). Together, these methods provide strong and complementary direct-regression baselines for evaluating field accuracy, computational cost, and the trade-off between reconstruction fidelity and representation utility.

3.2 HILIFTAEROML: REPRESENTATIONS AND HIGH-RESOLUTION FIELD PREDICTION

Full-field surrogate comparison. HiLiftAeroML is the most computationally demanding benchmark in our study because it combines realistic high-lift aircraft geometries with approximately 15 million surface points and 50 million volumetric points per case. We evaluate the velocity and pressure fields on the boundary-layer surface, where the flow exhibits strong spatial variation. All baseline models are evaluated using their official NVIDIA PhysicsNeMo implementations. For Transolver and GeoTransolver, the surface is divided into randomized chunks of $131{,}072$ points, with every evaluation point processed exactly once to reconstruct the complete evaluated field. The PhysicsNeMo implementation of FIGConvUNet does not support this chunked reconstruction procedure. Consequently, FIGConvUNet predicts only one randomly sampled chunk of $131{,}072$ points from each case, and its reported accuracy and computational cost apply only to that subset rather than to the complete field. AeroJEPA encodes a subsample of $131{,}072$ geometry points, matching the sampling resolution used during training, and predicts a compact aerodynamic latent state. Its implicit decoder is then evaluated over the complete boundary-layer field. The decoded value at a coordinate does not depend on the other query points evaluated in the same batch, allowing the decoder queries to be divided into arbitrary batches without changing the predictions. The baselines receive their prescribed geometric features, including surface normals, whereas AeroJEPA uses point coordinates and operating conditions.

Table 1 reports relative L_2 errors together with floating-point cost, inference time, and model size. All inference measurements were obtained on a single NVIDIA H200 GPU. AeroJEPA matches GeoTransolver in pressure error, obtains the lowest errors for u and w, and is second to GeoTransolver for v. Reconstructing the complete evaluated field requires 57.00 TFLOPs and 43 seconds with AeroJEPA, compared with 189.44 TFLOPs and 115 seconds for Transolver and 309.13 TFLOPs and 187 seconds for GeoTransolver. AeroJEPA therefore reduces full-field inference time

Table 1: Field-prediction accuracy and computational cost on HiLiftAeroML, reported as mean ± standard deviation across test cases. Inference times were measured on a single NVIDIA H200 GPU. Transolver, GeoTransolver, and AeroJEPA are evaluated over the complete boundary-layer field. Due to limitations of its official PhysicsNeMo implementation, FIGConvUNet is evaluated on one randomly sampled chunk of 131,072 points per case. Its accuracy, TFLOPs, and inference time are therefore not directly comparable with the full-field results of the other methods. Parameter counts refer to deployed models, except that both training and deployed sizes are reported for AeroJEPA.

Model	Rel. $L_2(p)\downarrow$	Rel. $L_2(u)\downarrow$	Rel. $L_2(v)\downarrow$	Rel. $L_2(w)\downarrow$	TFLOPs ↓	Time (s) ↓	Parameters (M) ↓
FIGConvUNet	0.0066 ± 0.0009	0.2112 ± 0.0012	0.3692 ± 0.0346	0.2538 ± 0.0337	88.34 ± 2.61	44 ± 1.2	6.17
GeoTransolver	$\mathbf{0.0048 \pm 0.0005}$	0.1503 ± 0.0017	$\mathbf{0.2489 \pm 0.0287}$	0.1898 ± 0.0389	309.13 ± 8.97	187 ± 4.4	9.63
Transolver	0.0051 ± 0.0006	0.1650 ± 0.0025	0.3180 ± 0.0289	0.1990 ± 0.0406	189.44 ± 5.50	115 ± 2.5	3.99
AeroJEPA (ours)	$\mathbf{0.0048 \pm 0.0026}$	$\mathbf{0.1445 \pm 0.0423}$	0.2749 ± 0.0388	$\mathbf{0.1880 \pm 0.0455}$	$\mathbf{57.00 \pm 1.65}$	$\mathbf{43 \pm 0.7}$	**3.09** deployed / 3.45 training

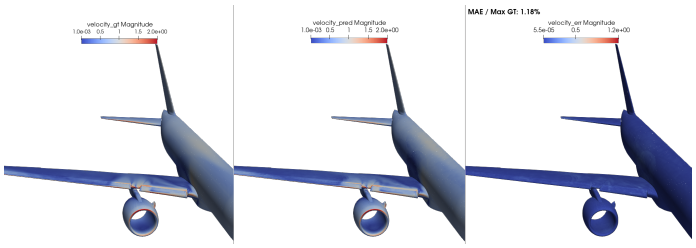

Figure 2: HiLiftAeroML reconstruction for test geometry LHC013 at 18° angle of attack. AeroJEPA captures the main boundary-layer velocity patterns and high-gradient regions on an unseen configuration while decoding the field continuously over the full aircraft surface. The relative error reported in the figure is 1.18% in MAE / max-GT-mean for the velocity-magnitude field. Additional views are provided in Appendix D.1.

by 62.6% relative to Transolver and by 77.0% relative to GeoTransolver. The 88.34 TFLOPs and 44 seconds reported for FIGConvUNet correspond only to one 131,072-point chunk and are not directly comparable with the full-field costs of the other methods. AeroJEPA is also the smallest evaluated model at deployment because the target encoder used during training is discarded at inference.

Beyond aggregate errors, the qualitative reconstructions confirm that the model captures the relevant high-lift flow structures on unseen cases. Figure 2 shows the decoded velocity field for test geometry LHC013 at 18° angle of attack. The prediction remains consistent with the reference CFD solution in the most sensitive surface regions, including areas affected by strong gradients and flow separation. This qualitative agreement is important because the benchmark evaluates the full resolved boundary-layer field rather than a reduced set of probes or integrated outputs; additional pressure views are provided in Appendix D.1.

Latent analysis. We next examine whether the learned latent spaces encode physically meaningful structure. Figure 3.2 compares a PCA projection of the context latents from AeroJEPA against an autoencoder baseline. This baseline preserves a context branch and mixes context and target information through cross-attention before decoding to keep the same information paths as our AeroJEPA, but the training objective is different. AeroJEPA organizes the geometry manifold more coherently, suggesting that the predictive objective induces a more structured representation of the design space. This qualitative separation is supported quantitatively in Appendix D.1: ridge probes recover the four dominant high-lift control-surface deflections from the context latent with R^2 values between 0.965 and 0.988, even though these parameters are never used during training.

The predicted latents after the predictor also encode nontrivial aerodynamic information, despite the fact that the model is trained only on the primitive fields (u, v, w, p) and never on integrated coefficients such as C_L or C_D. As shown in Figure 3.2, the predicted-latent manifold varies smoothly with aerodynamic quantities, and the corresponding linear probes recover C_L and C_D with $R^2 = 0.939$ and 0.986, respectively. This indicates that the predictive objective induces a latent representation aligned not only with geometry, but also with downstream aerodynamic performance.

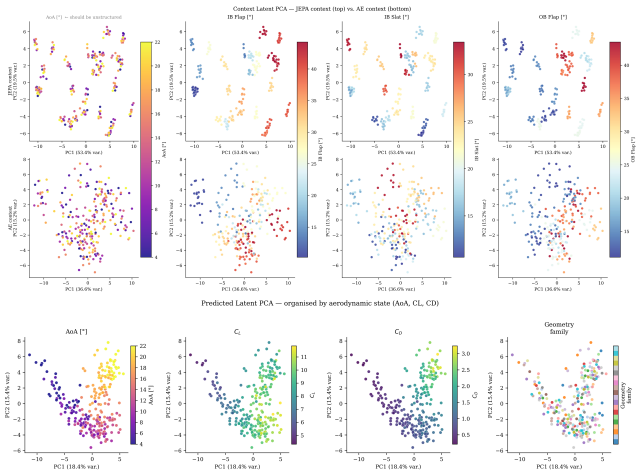

Figure 3: Main HiLiftAeroML latent-space results. *Left:* PCA projection of the context latents comparing AeroJEPA with an autoencoder baseline, showing a more coherent organization by aircraft geometry. *Right:* projection of the predicted latents after conditioning on the operating state, showing smooth variation with aerodynamic quantities even though the model is trained only on the primitive fields (u, v, w, p).

We also find that the context latent is interpretable beyond linear recovery. The concept-vector arithmetic analysis in Appendix D.6 shows that the main flap and slat directions act as nearly disentangled latent controls: flap-slat cross-talk is weak, while the remaining couplings occur primarily between inboard and outboard elements of the same surface, which is physically consistent with how these controls co-vary in the dataset.

Taken together, these experiments show that AeroJEPA combines competitive reconstruction on very large fields with geometry and aerodynamic representations that support probing and controlled latent traversal.

Appendix D.2 evaluates selective decoding on an outboard-wing region of 50 held-out cases. AeroJEPA obtains a pressure RMSE of 1.468 and an R^2 of 0.953, compared with RMSEs of 8.462 and 8.537 and R^2 values of -0.161 and -0.089 for Transolver and GeoTransolver, respectively. AeroJEPA also produces identical values when the same coordinates are divided into different query batches, with a repartition NRMSE of zero, compared with 0.0225 and 0.0134 for the two direct-regression baselines. These results support localized, query-set-independent decoding.

Additional quantitative probe recoveries, interpolation results, and qualitative reconstructions are provided in Appendix D.1.

3.3 SuperWing: broad generalization and latent-space optimization

We use SuperWing to evaluate a complementary regime: generalization across a broad family of transonic swept wings and the usefulness of the learned representations for downstream aerodynamic analysis. Unlike HiLiftAeroML, where the main computational difficulty is the extreme size of each target field, SuperWing contains fixed-resolution surface fields of approximately $32{,}768$ points. The complete field therefore fits in a single forward pass for all evaluated methods, allowing direct comparison with the native one-pass operation of the direct-regression baselines.

Table 2 reports full-field relative L_2 errors and inference TFLOPs. GeoTransolver and Transolver obtain the lowest reconstruction errors, while AeroJEPA remains more accurate than FIGConvUNet. This is the expected reconstruction-representation trade-off: the direct-regression transformers devote their capacity directly to field accuracy, whereas AeroJEPA additionally organizes the problem into reusable geometry and aerodynamic representations. The SuperWing results therefore do not show an across-the-board reconstruction advantage for AeroJEPA; instead, they establish that the learned representations remain available at a moderate cost in full-field accuracy.

Table 2: Full-field relative L_2 errors and inference cost on SuperWing, reported as mean $\pm$ standard deviation across test cases. All methods process the complete $32{,}768$-point field in one pass.

Model	$C_{f,\tau} \downarrow$	$C_{f,z} \downarrow$	$C_p \downarrow$	TFLOPs $\downarrow$
FIGConvUNet	0.1429 ± 0.0525	0.2912 ± 0.1080	0.1946 ± 0.0713	**0.27**
GeoTransolver	$\mathbf{0.0280 \pm 0.0177}$	$\mathbf{0.0529 \pm 0.0335}$	$\mathbf{0.0309 \pm 0.0157}$	0.45
Transolver	0.0324 ± 0.0179	0.0606 ± 0.0341	0.0358 ± 0.0164	0.35
AeroJEPA (ours)	0.0548 ± 0.0258	0.1084 ± 0.0513	0.0644 ± 0.0258	0.32

Additional decoded-field comparisons and force parity plots for SuperWing are reported in Appendix D.3.

Beyond field reconstruction, we evaluate whether the two latent spaces retain information that can be accessed without decoding and integrating the complete surface field. Linear probes applied to the predicted aerodynamic tokens recover C_L and C_D, as well as their within-geometry changes, despite these coefficients not being direct training targets. The geometry tokens similarly retain information about the underlying wing design parameters. Detailed probe results and the loss ablations used to evaluate this representation structure are reported in Appendix D.4.

Finally, we test whether the geometry representation supports preliminary design-space exploration by optimizing the probe-predicted aerodynamic efficiency C_L/C_D at a fixed cruise condition directly in the 128-dimensional context latent space. The differentiable search uses the frozen AeroJEPA predictor and linear probes for C_L and C_D, and is restricted by a latent trust region, bounds on reliably recovered design parameters, aerodynamic constraints, and the observed L/D envelope. The optimization converges toward the high-efficiency region of the dataset without repeatedly decoding the surface field, while the recovered design resembles a high-efficiency wing present in the dataset. Because the synthetic optimum has not been validated with CFD, this result demonstrates constrained exploration of the learned representation rather than improved physical performance. Further details are provided in Appendix D.5.

Taken together, the SuperWing experiments reveal the intended trade-off of the framework. Direct-regression transformers are more accurate full-field predictors on this benchmark, whereas AeroJEPA provides explicit geometry and aerodynamic representations that support lightweight readouts and preliminary latent-space exploration without repeated field decoding. This complements the HiLiftAeroML results, where AeroJEPA remains competitive on multi-million-point fields and permits localized decoding from a reusable full-aircraft geometry representation.

4 Conclusions

AeroJEPA is a predictive representation-learning framework for 3D aerodynamics that separates geometry and aerodynamic information into context and predicted tokens, while a continuous decoder provides field values at requested coordinates. Across HiLiftAeroML and SuperWing, these representations support linear readouts of design variables and aerodynamic quantities, controlled interpolation, concept-vector arithmetic, and preliminary latent-space optimization without requiring full-field reconstruction for each downstream task. AeroJEPA remains competitive on the multi-million-point HiLiftAeroML fields with lower computational cost than the evaluated transformer baselines, while direct-regression models achieve greater full-field accuracy on SuperWing. This exposes the intended trade-off between reconstruction fidelity and reusable representations. The selective-decoding experiment further shows that AeroJEPA can evaluate a localized region independently of how its query coordinates are batched.

The present study establishes these capabilities across two complementary steady-flow benchmarks. Future work will extend the framework to unsteady and multi-fidelity regimes, study additional downstream engineering tasks, and integrate the learned representations with inverse-design and shape-parameterization pipelines, including the validation of latent-space design candidates through new CFD simulations.

AI USE STATEMENT

Generative AI tools were used to improve grammar, polish the writing, and clarify selected expressions in the manuscript. All AI-assisted text was reviewed and revised by the authors. Generative AI was not used to generate experimental data, execute experiments, or determine the scientific conclusions.

REPRODUCIBILITY STATEMENT

We support reproducibility by providing detailed descriptions of the AeroJEPA architecture, training objectives, preprocessing procedure, sampling strategy, optimization settings, and dataset partitions in the main paper and Appendix C. Appendix D.4 reports the training-objective ablations, while Appendix D.2 and Appendix D.5 describe the selective-decoding and latent-optimization protocols, respectively. The evaluated datasets are publicly available, and the baseline experiments use their official NVIDIA PhysicsNeMo implementations. Hardware, training durations, inference protocols, and principal hyperparameters are also reported. Code, configuration files, and evaluation scripts will be made publicly available upon publication.

REFERENCES

Corey Adams, Rishikesh Ranade, Ram Cherukuri, and Sanjay Choudhry. Geotransolver: Learning physics on irregular domains using multi-scale geometry aware physics attention transformer. *arXiv preprint arXiv:2512.20399*, 2025.

Saeed Akbari, Pedram H Dabaghian, and Omer San. Blending machine learning and sequential data assimilation over latent spaces for surrogate modeling of boussinesq systems. *Physica D: Nonlinear Phenomena*, 448:133711, 2023.

Neil Ashton, Adam M Clark, Christopher Ivey, Liam Heidt, Sanjeeb Bose, Rishi Ranade, Rahul Agrawal, and Konrad Goc. High-fidelity cfd data generation for hiliftaeroml using solution-adapted wmles. In *AIAA SCITECH 2026 Forum*, pp. 0042, 2026.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15619–15629, 2023.

Hema Aubeelack, Stéphane Segonds, Christian Bes, Thierry Druot, Joël Brezillon, Adrien Bérard, Marylène Duffau, and Guillaume Gallant. Surrogate model development for optimized blended-wing-body aerodynamics. *Journal of Aircraft*, 60(2):437–448, 2023.

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024.

Randall Balestriero and Yann LeCun. Lejepa: Provable and scalable self-supervised learning without the heuristics. *arXiv preprint arXiv:2511.08544*, 2025.

Jai Bardhan, Radhikesh Agrawal, Abhiram Tilak, Cyrin Neeraj, and Subhadip Mitra. Hep-jepa: A foundation model for collider physics using joint embedding predictive architecture. *arXiv preprint arXiv:2502.03933*, 2025.

Gregory D Bird, Steven E Gorrell, and John L Salmon. Dimensionality-reduction-based surrogate models for real-time design space exploration of a jet engine compressor blade. *Aerospace Science and Technology*, 118:107077, 2021.

Giovanni Catalani, Siddhant Agarwal, Xavier Bertrand, Frederic Tost, Michael Bauerheim, and Joseph Morlier. Neural fields for rapid aircraft aerodynamics simulations. *Scientific Reports*, 14 (1):25496, 2024.

Delong Chen, Mustafa Shukor, Theo Moutakanni, Willy Chung, Jade Yu, Tejaswi Kasarla, Yejin Bang, Allen Bolourchi, Yann LeCun, and Pascale Fung. Vl-jepa: Joint embedding predictive architecture for vision-language. *arXiv preprint arXiv:2512.10942*, 2025.

Sibo Cheng, Jianhua Chen, Charitos Anastasiou, Panagiota Angeli, Omar K Matar, Yi-Ke Guo, Christopher C Pain, and Rossella Arcucci. Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. *Journal of Scientific Computing*, 94(1): 11, 2023.

Chris Choy, Alexey Kamenev, Jean Kossaifi, Max Rietmann, Jan Kautz, and Kamyar Azizzadenesheli. Factorized implicit global convolution for automotive computational fluid dynamics prediction. *arXiv preprint arXiv:2502.04317*, 2025.

Viet Chu, Omar Mashaal, and Hatem Abou-Zeid. Wirelessjepa: A multi-antenna foundation model using spatio-temporal wireless latent predictions. *arXiv preprint arXiv:2601.20190*, 2026.

Matthieu Destrade, Oumayma Bounou, Quentin Le Lidec, Jean Ponce, and Yann LeCun. Value-guided action planning with jepa world models. *arXiv preprint arXiv:2601.00844*, 2025.

Katrina Drozdov, Ravid Shwartz-Ziv, and Yann LeCun. Video representation learning with joint-embedding predictive architectures. *arXiv preprint arXiv:2412.10925*, 2024.

Xiaosong Du, Ping He, and Joaquim RRA Martins. Rapid airfoil design optimization via neural networks-based parameterization and surrogate modeling. *Aerospace Science and Technology*, 113:106701, 2021.

James Duvall and Karthik Duraisamy. Discretization-independent surrogate modeling of physical fields around variable geometries using coordinate-based networks. *Data-Centric Engineering*, 6: e5, 2025.

Hamidreza Eivazi, Soledad Le Clainche, Sergio Hoyas, and Ricardo Vinuesa. Towards extraction of orthogonal and parsimonious non-linear modes from turbulent flows. *Expert Systems with Applications*, 202:117038, 2022.

Ali ElSheikh, Rui-Xi Wang, Weimin Wu, Yibo Wen, Payam Dibaeinia, Jennifer Yuntong Zhang, Jerry Yao-Chieh Hu, Mei Knudson, Sudarshan Babu, Shao-Hua Sun, et al. Cell-jepa: Latent representation learning for single-cell transcriptomics. *arXiv preprint arXiv:2602.02093*, 2026.

Zhengcong Fei, Mingyuan Fan, and Junshi Huang. A-jepa: Joint-embedding predictive architecture can listen. *arXiv preprint arXiv:2311.15830*, 2023.

Alexander Forrester, Andras Sobester, and Andy Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.

Alexander IJ Forrester, Neil W Bressloff, and Andy J Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071):2177–2204, 2006.

Víctor Francés-Belda, Alberto Solera-Rico, Javier Nieto-Centenero, Esther Andrés, Carlos Sanmiguel Vila, and Rodrigo Castellanos. Toward aerodynamic surrogate modeling based on β-variational autoencoders. *Physics of Fluids*, 36(11), 2024.

Quentin Garrido, Mahmoud Assran, Nicolas Ballas, Adrien Bardes, Laurent Najman, and Yann LeCun. Learning and leveraging world models in visual representation learning. *arXiv preprint arXiv:2403.00504*, 2024.

Yousef Ghazi, Nahla Alhazmi, Radek Tezaur, and Charbel Farhat. Training a neural-network-based surrogate model for aerodynamic optimisation using a gaussian process. *International Journal of Computational Fluid Dynamics*, 36(7):538–554, 2022.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. Pmlr, 2017.

C Hu, S Martin, and R Dingreville. Accelerating phase-field predictions via recurrent neural networks learning the microstructure evolution in latent space. *Computer Methods in Applied Mechanics and Engineering*, 397:115128, 2022.

Naiwen Hu, Haozhe Cheng, Yifan Xie, Shiqi Li, and Jihua Zhu. 3d-jepa: A joint embedding predictive architecture for 3d self-supervised representation learning. *arXiv preprint arXiv:2409.15803*, 2024.

Subash Katel, Haoyang Li, Zihan Zhao, Raghav Kansal, Farouk Mokhtar, and Javier Duarte. Learning symmetry-independent jet representations via jet-based joint embedding predictive architecture. *arXiv preprint arXiv:2412.05333*, 2024.

Katiana Kontolati, Somdatta Goswami, George Em Karniadakis, and Michael D Shields. Learning nonlinear operators in latent spaces for real-time predictions of complex dynamics in physical systems. *Nature Communications*, 15(1):5101, 2024.

Xianhang Li, Chen Huang, Chun-Liang Li, Eran Malach, Josh Susskind, Vimal Thilak, and Etai Littwin. Rethinking jepa: Compute-efficient video ssl with frozen teachers. *arXiv preprint arXiv:2509.24317*, 2025.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. Leworld-model: Stable end-to-end joint-embedding predictive architecture from pixels. *arXiv preprint arXiv:2603.19312*, 2026.

Shentong Mo and Shengbang Tong. Connecting joint-embedding predictive architecture with contrastive self-supervised learning. *Advances in neural information processing systems*, 37: 2348–2377, 2024.

Shentong Mo and Sukmin Yun. Dmt-jepa: Discriminative masked targets for joint-embedding predictive architecture. *arXiv preprint arXiv:2405.17995*, 2024.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

Avishka Perera, Kumal Hewagamage, Saeedha Nazar, Kavishka Abeywardana, Hasitha Gallella, Ranga Rodrigo, and Mohamed Afham. Crossjepa: Cross-modal joint-embedding predictive architecture for efficient 3d representation learning from 2d images. *arXiv preprint arXiv:2511.18424*, 2025.

Helen Qu, Rudy Morel, Michael McCabe, Alberto Bietti, François Lanusse, Shirley Ho, and Yann LeCun. Representation learning for spatiotemporal physical systems. *arXiv preprint arXiv:2603.13227*, 2026.

Geri Skenderi, Hang Li, Jiliang Tang, and Marco Cristani. Graph-level representation learning with joint-embedding predictive architectures. *arXiv preprint arXiv:2309.16014*, 2023.

Vlad Sobal, Jyothir SV, Siddhartha Jalagam, Nicolas Carion, Kyunghyun Cho, and Yann LeCun. Joint embedding predictive architectures focus on slow features. *arXiv preprint arXiv:2211.10831*, 2022.

Alberto Solera-Rico, Carlos Sanmiguel Vila, Miguel Gómez-López, Yuning Wang, Abdulrahman Almashjary, Scott TM Dawson, and Ricardo Vinuesa. β-variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nature Communications*, 15(1):1361, 2024.

Hugo Thimonier, José Lucas De Melo Costa, Fabrice Popineau, Arpad Rimmel, and Bich-Liên Doan. T-jepa: Augmentation-free self-supervised learning for tabular data. *arXiv preprint arXiv:2410.05016*, 2024.

Jonas Ulmen, Ganesh Sundaram, and Daniel Görges. Learning state-space models of dynamic systems from arbitrary data using joint embedding predictive architectures. *IFAC-PapersOnLine*, 59(18):19–24, 2025.

Stijn Verdenius, Andrea Zerio, and Roy LM Wang. Lat-pfn: A joint embedding predictive architecture for in-context time-series forecasting. *arXiv preprint arXiv:2405.10093*, 2024.

Ricardo Vinuesa, Steven L Brunton, and Gianmarco Mengaldo. Explainable ai: Learning from the learners. *arXiv preprint arXiv:2601.05525*, 2026.

Abhijeet Vishwasrao, Francisco Giral, Mahmoud Golestanian, Federica Tonti, Andrea Arroyo Ramo, Adrian Lozano-Duran, Steven L Brunton, Sergio Hoyas, Soledad Le Clainche, Hector Gomez, et al. Agentic exploration of pde spaces using latent foundation models for parameterized simulations. *arXiv preprint arXiv:2604.09584*, 2026.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *arXiv preprint arXiv:2402.02366*, 2024a.

Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler faster stronger. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4840–4851, 2024b.

Yunjia Yang, Weishao Tang, Mengxin Liu, Nils Thuerey, Yufei Zhang, and Haixin Chen. Superwing: a comprehensive transonic wing dataset for data-driven aerodynamic design. *arXiv preprint arXiv:2512.14397*, 2025.

Brandon Yee and Pairie Koh. Pi-jepa: Label-free surrogate pretraining for coupled multiphysics simulation via operator-split latent prediction. *arXiv preprint arXiv:2604.01349*, 2026.

Raul Yondo, Esther Andrés, and Eusebio Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in aerospace sciences*, 96:23–61, 2018.

Weiwei Zhang, PENG Xuhao, KOU Jiaqing, and WANG Xu. Heterogeneous data-driven aerodynamic modeling based on physical feature embedding, 2024.

Junhong Zou, Wei Qiu, Zhenxu Sun, Xiaomei Zhang, Zhaoxiang Zhang, and Xiangyu Zhu. Adafield: Generalizable surface pressure modeling with physics-informed pre-training and flow-conditioned adaptation. *arXiv preprint arXiv:2601.07139*, 2026.

APPENDIX

A OTHER RELATED WORKS

Broader surrogate-modeling literature. Beyond the works emphasized in the main paper, there is a broad literature on surrogate modeling for aerodynamic design, variable-fidelity optimization, and engineering-response approximation (Du et al., 2021; Ghazi et al., 2022). There is also extensive work on latent-state and reduced-order surrogate modeling for physical systems more broadly, including latent simulation, data assimilation, and design-space exploration (Hu et al., 2022; Cheng et al., 2023; Akbari et al., 2023; Kontolati et al., 2024; Bird et al., 2021; Aubeelack et al., 2023; Zhang et al., 2024). These papers help situate AeroJEPA within the larger surrogate-modeling landscape, but they are less central to our contribution than methods that explicitly combine aerodynamic field prediction with learned latent structure.

Broader latent-design and JEPA literature. On the JEPA side, predictive embedding methods have already been explored for audio, graphs, time series, tabular data, planning, and several scientific domains (Fei et al., 2023; Skenderi et al., 2023; Verdenius et al., 2024; Thimonier et al., 2024; Drozdov et al., 2024; Mo & Tong, 2024; Mo & Yun, 2024; Garrido et al., 2024; Destrade et al., 2025; Bardhan et al., 2025; Katel et al., 2024; Chu et al., 2026; ElSheikh et al., 2026; Ulmen et al., 2025). We view these works as evidence that predictive latent learning is broadly useful, while AeroJEPA specializes that paradigm to aerodynamic surrogate modeling with a geometry-to-flow latent bridge.

B ADDITIONAL DATASET DETAILS

This section summarizes the two datasets utilized throughout the paper, emphasizing geometry parameterization, observable fluid variables, and the structural point-cloud representation mandated by the AeroJEPA pipeline. Furthermore, we outline the exact dataset partition strategies formulated to rigorously evaluate generalizability.

B.1 HILIFTAEROML

HiLiftAeroML (Ashton et al., 2026) serves as the high-fidelity aerodynamic benchmark in our study. The dataset catalogs realistic high-lift aircraft configurations simulated via solution-adapted wall-modeled large-eddy simulation (WMLES). This physical complexity makes it particularly relevant for evaluating continuous latent representations under highly separated, turbulent flow. Computationally, it represents the most demanding environment from a spatial scale perspective. In our formulation, the surface-oriented boundary-layer representation resolves approximately 12–15 million points per sample, whereas the internal volumetric representation constructs roughly 50 million elements per case.

Geometrically, the configurations are denoted by specific design geometry (e.g., "LHCXX") defined internally by 8 continuous design parameters. Four of these scalars dictate the highest observed flow variation: the inboard and outboard flap deflections alongside the inboard and outboard slat deflections. For every given geometry configuration, the dataset provides 10 discrete operational flow snapshots sweeping across an Angle of Attack (AoA) domain bounding from 4° to 22°. Given the extreme computational burden intrinsic to evaluating millions of points across full WMLES traces, we strictly partition the benchmark topologies into 205 cases reserved for the training manifold, and 50 cases held out for testing.

B.2 SUPERWING

SuperWing (Yang et al., 2025) acts as the large-scale transonic benchmark designed to evaluate continuous generalization across topological and operational variabilities. The corpus contains exactly $4{,}239$ parameterized wing geometries mapped across $28{,}856$ unique Reynolds-averaged Navier–Stokes (RANS) state solutions spanning a broad aerodynamic domain of Mach numbers and freestream angles of attack. Each simulation is originally distributed on a structured surface discretization of roughly $32{,}000$ points. To ensure architectural parity, we treat these grids purely as unstructured point sets so the AeroJEPA pipeline remains invariant across dataset discretizations.

Target fluid fields across SuperWing are iteratively mapped as continuous surface vectors, evaluating the local pressure coefficient (C_p) and skin-friction coefficient (C_f), contrary to volumetric arrays. Geometrically, each surface is prescribed by 54 latent morphological degrees of freedom regulating localized shape, wing twist, and spanwise dihedral.

Crucially, to rigorously enforce zero-shot geometric generalizability, the dataset partitioning is structured strictly across the geometrical dimension rather than by random simulation sweeps. The parameterized wings are isolated into an $80\%/10\%/10\%$ split corresponding to training, validation, and testing regimes, respectively. Because the stratification dictates that all unique operating parameters for a specific held-out wing topology remain completely unseen during optimization, the evaluation cleanly measures fundamental physical extrapolation over unseen boundary shapes.

C Additional model details

This appendix provides a comprehensive, step-by-step description of the AeroJEPA pipeline, ranging from raw computational fluid dynamics (CFD) representation to latent prediction and subsequent continuous field decoding.

C.1 Point-cloud preprocessing and subsampling

Standard unstructured CFD meshes exhibit highly variable node densities. To standardize the input representation, connectivity operations are discarded, allowing both the physical boundary and the surrounding flow to be formally treated as continuous point sets. Let the complete dense surface geometry be denoted as $\mathcal{P}$, and the comprehensive ground-truth fluid states over the spatial domain Ω as $\mathcal{F}$. For surface-only fluid problems, the geometry branch processes purely spatial coordinates $\mathbf{x} \in \mathbb{R}^3$. For volumetric domains, this input is strictly augmented with the local Signed-Distance Function (SDF).

Because the complete CFD point sets exceed the memory required for direct processing, we uniformly sample bounded observation sets for each case. The geometry branch receives a context set $\mathcal{P}_c \subset \mathcal{P}$ of cardinality N_c, while the target branch receives an independently sampled flow set $\mathcal{F}_t \subset \mathcal{F}$ of cardinality N_t. A third independently sampled coordinate set $\mathcal{Q} \subset \Omega$, with cardinality N_q, provides the reconstruction queries. Farthest point sampling is applied only after this initial sampling step to select the tokenizer centroids within each bounded set.

The geometry, target, and reconstruction sets are sampled independently, so no pointwise correspondence is imposed between their elements. Incidental overlap is not prohibited, but the decoder queries are not aligned with the context or target tokens. This separation prevents the model from relying on a fixed pointwise mapping and trains the predicted tokens to support field evaluation at independently selected coordinates.

C.2 Tokenization with sampled centroids

Following subsampling, the irregular target and context point sets are projected into fixed-length sequences of latent spatial tokens. Specifically, the architecture leverages FPS once more over the subsampled sets $\mathcal{P}_c$ and $\mathcal{F}_t$ to designate foundational centroid coordinates. A localized message-passing neighborhood aggregation layer compiles point-wise variables around each centroid into preliminary latent tokens. This projection efficiently transitions the variable-scale physical point cloud into a structured, bounded topology compatible with global transformer operations. Based on the respective domain complexities, the token context sequence lengths (M) and feature dimensions (d) are scaled; the high-dimensional spatial topologies resident in HiLiftAeroML necessitate sequences of $M = 3072$ tokens wherein $d = 64$, whereas the lower-resolution surface fields intrinsic to SuperWing are sufficiently parameterized by $M = 512$ tokens of dimensionality $d = 128$.

C.3 Encoder backbones

Both the context encoder $\mathcal{E}_c$ and the target encoder $\mathcal{E}_t$ utilize computational graphs built upon localized Point Transformer attention mechanisms (Wu et al., 2024b). Crucially, the massive dimensionality reduction separating the raw N_c and N_t inputs from the computational bounds is handled entirely by

the preceding centroid-clustering and neighborhood aggregation layers. Consequently, the encoder backbones do not inherently alter the topological scale; they operate purely as flat processing stacks over the highly compressed, M-sized token distributions. By employing localized self-attention rather than global quadratic computations, the framework strictly bounds memory consumption. Through progressively stacked blocks, the encoder incrementally routes physical information across neighboring tokens. Shallow layers capture highly localized boundary layer interactions and surface curvature semantics, whereas deeper transformer blocks contextualize macro-aerodynamic coupling and wake propagation over multiple token hops. The resulting latent interface inherently preserves the geometric topology, formulating an optimizable blueprint for generative design.

C.4 Predictor architecture and conditioning

Sharing structural topology with the encoders, the core latent map is represented by a non-hierarchical predictor network operating seamlessly over the fixed output token resolution. The objective of the predictor is to synthesize target flow tokens purely from the geometry context tokens, and the driving physical constraints, without requiring spatial pooling operations.

Architecturally, the predictor instantiates learnable latent queries corresponding to the spatial coordinates of a fixed set of centroids. These spatial query coordinates are first embedded via high-frequency Fourier positional encodings. The network processes these queries through a stacked, flat sequence that tightly intertwines localized point-transformer self-attention and token-based cross-attention mechanisms. The self-attention pathway facilitates continuous spatial refinement and consistency among the target queries themselves, whereas the interleaving cross-attention layers allow these queries to explicitly fetch aligned geometric features from the upstream context tokens acting as keys and values.

Furthermore, recognizing that minor perturbations in freestream operating variables (e.g., Angle of Attack α, Reynolds Number Re, Mach Number Ma) fundamentally bifurcate the boundary layer dynamics independent of structural modification, we inject these physics conditions natively into the network. A lightweight multi-layer perceptron processes the scalar variables, projecting them deeply into the hidden transformer layers through adaptive feature modulation (Peebles & Xie, 2023). Consequently, the predictor framework robustly maps an invariant geometric parameterization into a variable continuous flow topology, establishing a smooth, differentiable property leveraged subsequently by gradient-based latent optimization probes.

C.5 INR decoder implementation details

To remap the predicted token distribution to an arbitrarily resolved continuous space, the predicted tokens $\hat{Z}_t$ condition the INR decoder. For any arbitrary element $\mathbf{q}_i \in \mathcal{Q}$, the coordinate is first transformed via Fourier feature encodings to enhance high-frequency spatial sensitivity before being processed by the decoding multi-layer perceptron (MLP). In volumetric settings, query elements are augmented by the respective SDF evaluations, explicitly distinguishing the interior solid boundary, the near-wall viscous boundary layer, and the external flow.

C.6 Training configurations and hyperparameters

The primary models are trained end to end with latent-matching and reconstruction weights $\lambda_\ell = 1.0$ and $\lambda_r = 1.0$, and SIGReg weight $\lambda_s = 0.001$. This configuration provides the strongest overall reconstruction and representation results in the SuperWing ablation reported in Appendix D.4. Network weights are optimized using AdamW with gradient clipping. All models, including the baselines, were trained on a single NVIDIA H200 GPU. On HiLiftAeroML, AeroJEPA required approximately 48 hours for 300 epochs, while the baseline models required approximately 96 hours for the same number of epochs. On SuperWing, AeroJEPA and the baseline models required approximately 24 hours for 200 epochs. Table 3 summarizes the sampling cardinalities, architectural dimensions, and optimization settings.

Table 3: Primary architectural parameterization and optimization bounds for the AeroJEPA framework, delineated by the underlying aerodynamic environment.

Configuration Variable	**HiLiftAeroML**	**SuperWing**
Sampling Capacities		
Geometry Context Subsample (N_c)	$131,072$	$8,192$
Fluid Target Subsample (N_t)	$131,072$	$8,192$
Reconstruction Query Set (N_q)	$131,072$	$8,192$
Latent Space Structure		
Context Tokens (M)	$3,072$	512
Token Dimensionality (d)	64	128
Transformer Backbone Depth	6 Layers	6 Layers
Optimization Setup		
Optimizer Formulation	AdamW	AdamW
Learning Rate	1×10^{-3}	1×10^{-3}
LR Annealing Schedule	Cosine Warmup	Cosine Warmup
Weight Decay	1×10^{-3}	1×10^{-3}
Gradient Clip Threshold	10.0	10.0
Total Training Epochs	300	200

D Additional results and experiments

D.1 Additional HiLiftAeroML results

Qualitative reconstruction. Figure 4 gathers additional qualitative views of the decoded fields for test geometry LHC013 at 18° angle of attack. Together with the velocity view retained in the main text, these complementary pressure and velocity plots confirm the same pattern: AeroJEPA preserves the large-scale structure of the boundary-layer solution and remains consistent with the reference CFD field in challenging regions.

Latent organization and interpolation. Figure 5 reports the ridge-probe recovery of aerodynamic coefficients from the predicted latent, complementing the predicted-manifold visualization shown in the main text. Figure 6 then examines interpolation in the latent space. We introduce a scalar interpolation parameter α and move between latent states corresponding to different angles of attack and geometry configurations. The decoded fields remain physically meaningful along the trajectory, and the resulting aerodynamic coefficients stay close to the ground-truth values, indicating that the learned manifold supports smooth transitions across both operating conditions and design changes.

Decoded-field consistency. We also assess whether the decoded fields yield accurate aerodynamic quantities when forces are estimated directly from the reconstructed surface solution, rather than inferred through a linear probe in latent space. Figure 7 shows the pressure-coefficient distribution on a wing section for case LHC019 at 12° angle of attack, comparing the coefficient extracted from the decoded field against the reference CFD profile. Figure 8 further reports parity plots for C_L and C_D computed from the decoded predicted fields. These results are complementary to the latent probing analysis in the main paper: they show that AeroJEPA not only stores aerodynamically meaningful information in latent space, but also recovers integrated force quantities from the actual decoded flow field with good agreement.

Linear recovery and concept arithmetic. Figure 9 reports the ridge-regression recovery of design variables from the context latents. These quantities are never used during training, so this experiment directly tests whether the context encoder captures geometry information in a form that is linearly accessible. The four dominant control-surface parameters are all recovered with high fidelity ($R^2 = 0.965$-0.988), which supports the interpretation of the context manifold in the main text. Figure 15 complements this result through latent arithmetic. We define concept vectors from the rows of the ridge-regression weight matrix, traverse the context latent space along those directions, and decode the resulting states back into standardized physical design parameters. The targeted variable

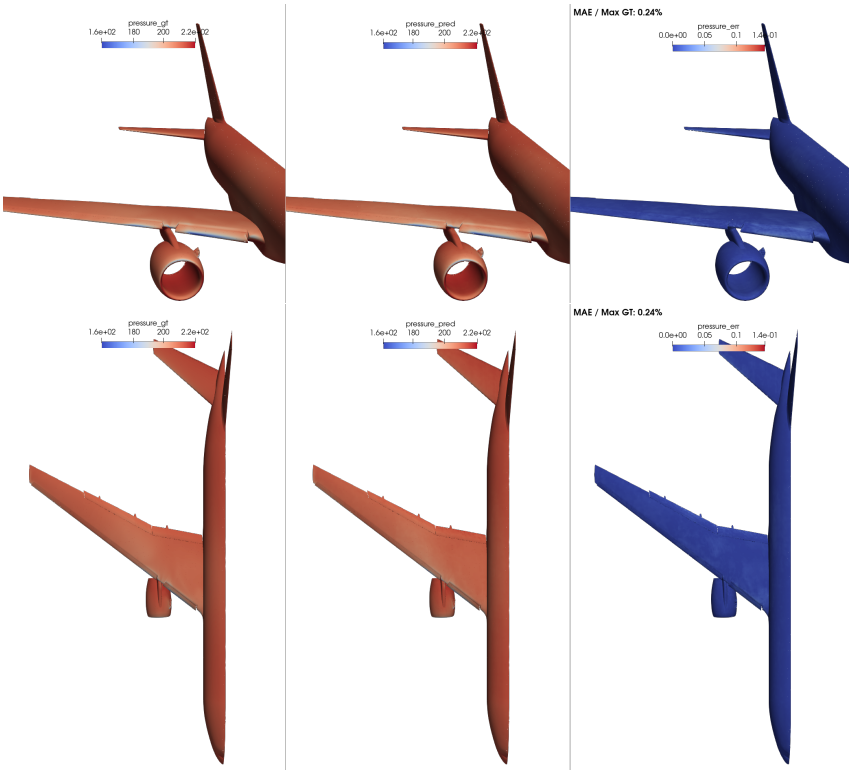

Figure 4: Additional qualitative pressure views for test geometry LHC013 at $18°$ angle of attack. These complementary views support the velocity-based qualitative comparison shown in the main text and confirm that AeroJEPA preserves the main surface-pressure structure on the unseen case. The relative error reported in the figure is 0.24% in MAE / max-GT-mean for the pressure field.

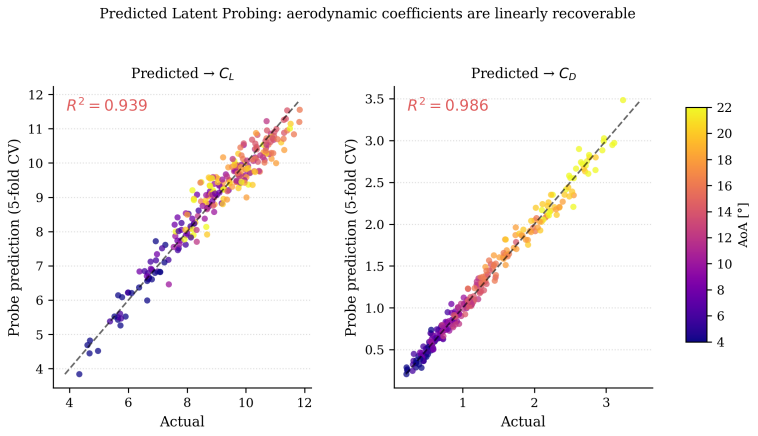

Figure 5: Ridge linear probing from predicted latents to aerodynamic coefficients on HiLiftAeroML. Even though the model is trained only on the primitive fields (u, v, w, p), the predicted latent supports accurate recovery of C_L and C_D, consistent with the smooth organization of the predicted manifold shown in the main text.

dominates the response along its corresponding trajectory, while the remaining parameters stay comparatively stable. The residual couplings mainly appear between inboard and outboard flaps and between inboard and outboard slats, which is consistent with the strong geometric dependencies of the high-lift configuration.

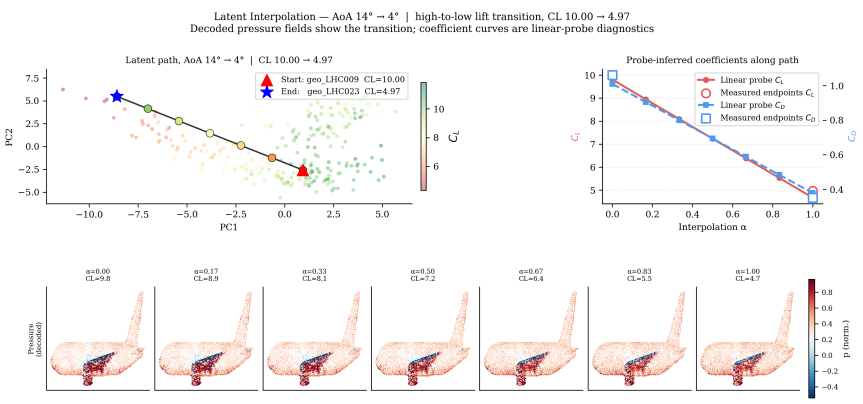

Figure 6: Latent interpolation on HiLiftAeroML between different operating conditions and geometry states. AeroJEPA yields decoded fields that remain physically meaningful across the interpolation path, with integrated coefficients that stay close to the corresponding ground-truth values.

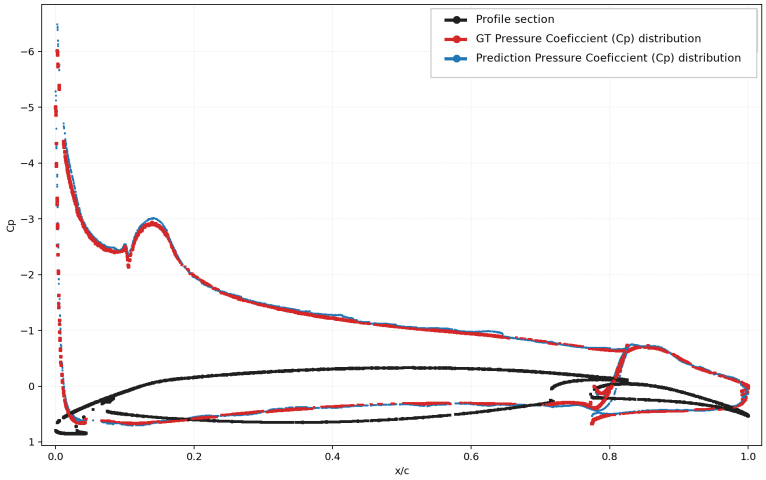

Figure 7: Pressure-coefficient profile extracted from the decoded field on a wing section for case LHC019 at 12° angle of attack. The comparison shows that AeroJEPA preserves the local pressure distribution needed to recover section-level aerodynamic behavior from the decoded solution.

D.2 Selective decoding on HiLiftAeroML

We evaluate whether a trained AeroJEPA can predict the aerodynamic field on a selected subset of surface coordinates without decoding the remainder of the aircraft. No model is retrained or fine-tuned for this experiment. The evaluation contains 50 held-out cases from five aircraft geometries.

Because the source files do not contain semantic component identifiers, we define a single fixed and visually verified outboard main-wing region using the surface flag and normalized coordinate bounds $-0.55 \leq y \leq -0.05$, $-0.55 \leq x \leq -0.28$, and $0.44 \leq z \leq 0.68$. This region excludes the wing root, engine, fuselage, and tail. For every case, we sample $65{,}536$ query points uniformly without replacement from this region. All models are evaluated on exactly the same coordinates and reference CFD values.

Figure 10 visualizes the absolute pressure error on a representative held-out case. AeroJEPA encodes $131{,}072$ uniformly sampled geometry points from the aircraft surface, as in training, but sends only the selected wing coordinates to its decoder. Predictions are evaluated using query batches of $8{,}192$ points and then repeated with batches of $4{,}096$ points. Transolver and GeoTransolver receive the same $65{,}536$ wing coordinates as two randomly partitioned sets of $32{,}768$ points. A second independent

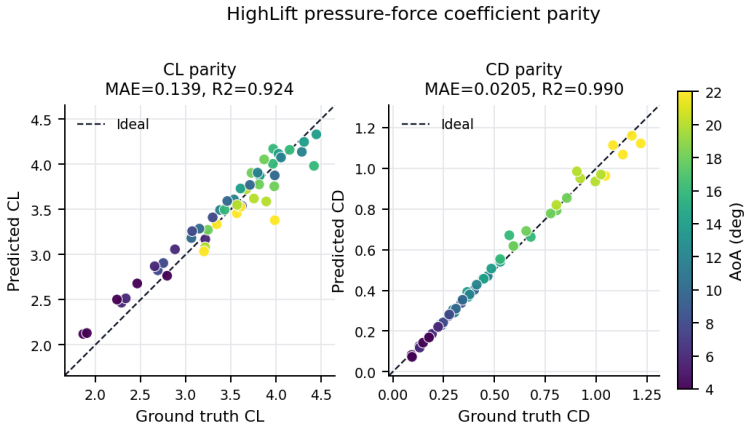

Figure 8: Parity plots for C_L and C_D computed from the decoded predicted fields. Unlike the latent probing experiment, these coefficients are obtained by estimating aerodynamic forces directly from the reconstructed surface solution, showing that AeroJEPA yields accurate integrated force predictions from the decoded field itself.

Table 4: Selective decoding on the outboard-wing region of 50 held-out HiLiftAeroML cases. Pressure RMSE is reported in the dataset pressure units. Repartition NRMSE measures the change in the four normalized output channels when the same coordinates are grouped into different query sets. Bracketed values are 95% geometry-cluster bootstrap intervals.

Model	Pressure RMSE ↓	Pressure R^2 ↑	u R^2 ↑	v R^2 ↑	w R^2 ↑	Repartition NRMSE ↓
AeroJEPA	**1.468** [1.080, 1.761]	**0.953**	**0.932**	**0.933**	**0.962**	**0.0000**
Transolver	8.462 [7.999, 8.925]	−0.161	−0.130	−0.188	−0.064	0.0225
GeoTransolver	8.537 [8.017, 9.037]	−0.089	−0.094	−0.189	−0.113	0.0134

partition is used to measure whether the prediction at a coordinate changes with the other coordinates processed alongside it.

Table 4 summarizes the quantitative selective-decoding results. AeroJEPA obtains a pressure RMSE of 1.468 and pressure R^2 of 0.953, with similarly high R^2 values across the three velocity components. Its predictions are identical under the two query batch sizes, producing a repartition NRMSE of zero. Transolver and GeoTransolver obtain pressure RMSEs of 8.462 and 8.537, respectively, and their predictions change when the same coordinates are repartitioned. The experiment therefore demonstrates that AeroJEPA can reuse a geometry encoding to decode a selected spatial region and that its output at a coordinate is independent of the other decoder queries evaluated in the same batch.

D.3 Additional SuperWing surrogate results

Decoded fields. Figure 11 shows representative decoded-field comparisons for $C_{f,z}$, $C_{f,\tau}$, and C_p on a SuperWing test case. The predicted surface fields remain consistent with the reference solution across the three quantities most relevant to transonic wing analysis.

Integrated aerodynamic response from decoded fields. Figure 12 reports parity plots for aerodynamic forces estimated directly from the decoded SuperWing fields. This comparison is performed on C_L and C_D recovered from the reconstructed surface solution itself. The decoded fields therefore preserve not only local surface quantities, but also the integrated aerodynamic response needed for downstream evaluation.

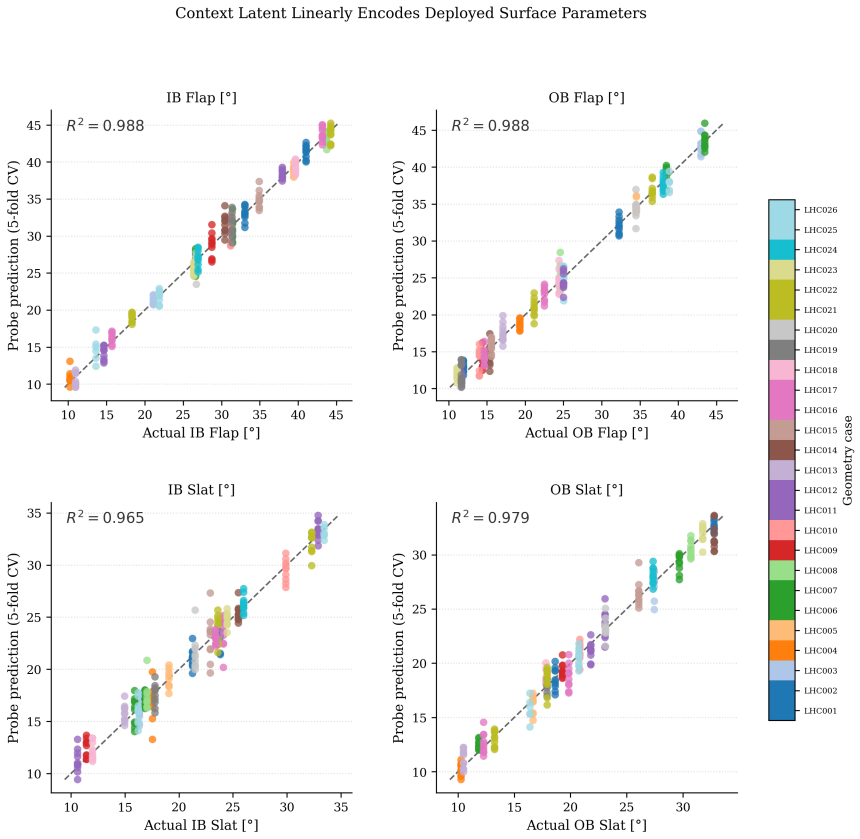

Figure 9: Recovery of HiLift design variables from the context latents using ridge regression. Although the model is trained only from input mesh points and target flow fields, the context representation remains strongly aligned with the underlying geometric parameters.

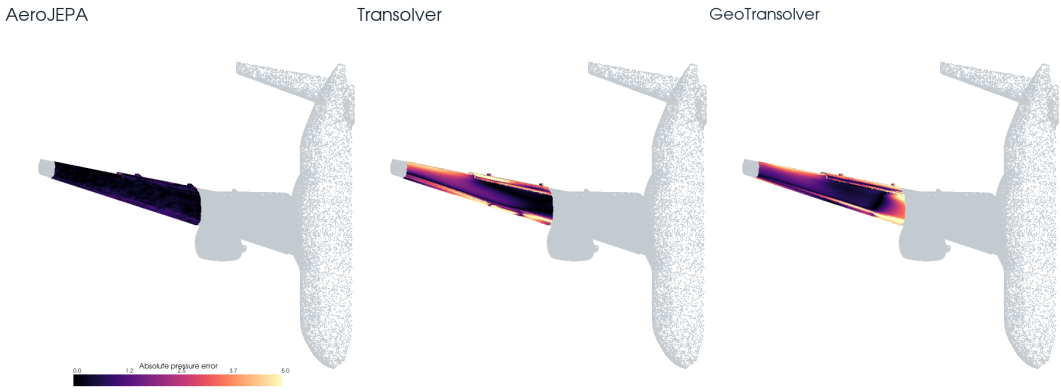

Figure 10: Absolute pressure error on the selected outboard-wing region for a representative held-out HiLiftAeroML case. All methods are evaluated on the same wing coordinates. The uncolored remainder of the aircraft shows the geometry and is not included in the displayed field predictions. AeroJEPA maintains substantially lower error over the selected region than the evaluated direct-regression models.

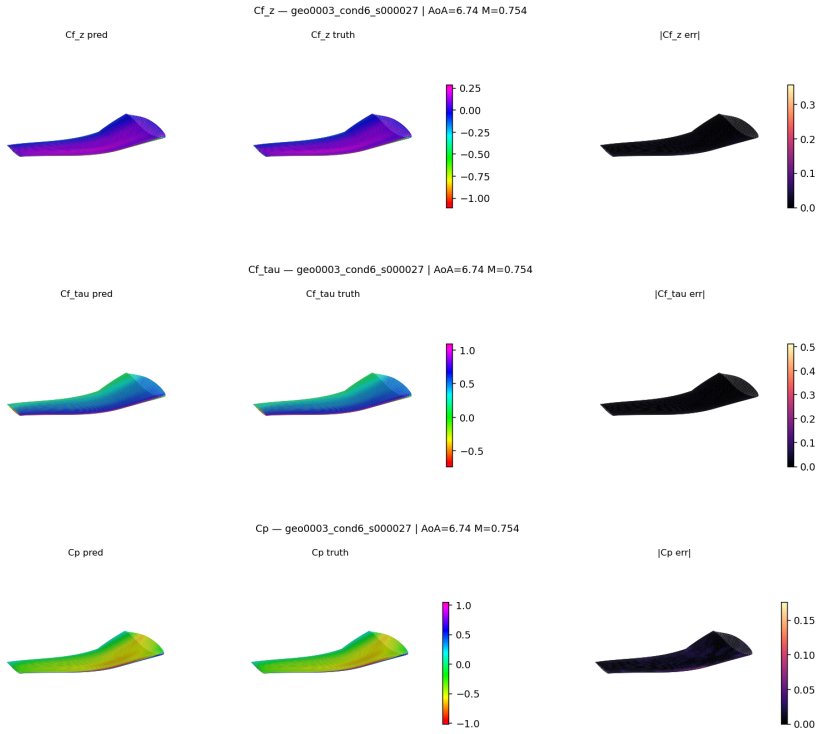

Figure 11: Representative decoded-field comparisons on SuperWing for $C_{f,z}$, $C_{f,\tau}$, and C_p. The predicted surface fields remain consistent with the reference solution across the three quantities most relevant to transonic wing analysis.

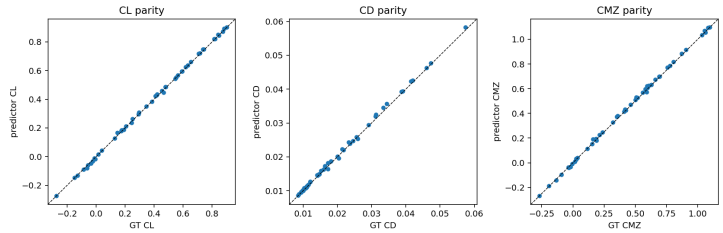

Figure 12: Parity plots for aerodynamic forces estimated from the decoded SuperWing fields. The comparison is performed on C_L and C_D recovered from the reconstructed surface solution, showing that the decoded fields preserve the integrated aerodynamic response with good fidelity.

D.4 Training-objective and representation ablations

We first study how the representation and reconstruction behavior changes with the training objective. Table 5 compares three SIGReg weights with a contrastive alternative on SuperWing. Representation quality is evaluated using linear probes trained on the training geometries and evaluated on held-out geometries. Design R^2 measures the mean recovery of eight selected design parameters from the context tokens, Aero R^2 measures the mean recovery of C_L and C_D, and ΔAero R^2 measures recovery of within-geometry changes in these coefficients. The context diagnostics use mean-pooled geometry tokens from held-out cases. Context cosine gap measures the separation between same-geometry, different-condition pairs and condition-matched, different-geometry pairs. Context within variation is the fraction of standardized context-feature variation occurring within rather than between geometries. Effective rank summarizes the spectrum of the centered context covariance.

Table 5: Training-objective ablation on SuperWing. Probe metrics are evaluated on held-out geometries. Field MSE is reported in training-normalized units on 12 held-out cases.

Objective	Design $R^2 \uparrow$	Aero $R^2 \uparrow$	ΔAero $R^2 \uparrow$	Context cosine gap $\uparrow$	Context within (%) $\downarrow$	Context eff. rank	Field MSE $\downarrow$
SIGReg $\lambda_s = 10^{-3}$	**0.886**	**0.993**	**0.987**	**0.983**	0.94	4.78	**0.0085**
SIGReg $\lambda_s = 10^{-2}$	0.859	0.978	0.964	**0.983**	0.92	4.72	0.0125
SIGReg $\lambda_s = 10^{-1}$	0.763	0.915	0.839	0.980	**0.91**	5.11	0.0158
Contrastive	0.460	0.783	0.704	0.738	17.74	3.76	0.5072

Table 6: Matched 30-epoch comparison between the complete AeroJEPA objective and reconstruction-only training on HiLiftAeroML. Design and aerodynamic quantities are recovered using linear probes.

Objective	IB flap $R^2 \uparrow$	OB flap $R^2 \uparrow$	IB slat $R^2 \uparrow$	OB slat $R^2 \uparrow$	ΔC_L $R^2 \uparrow$	ΔC_D $R^2 \uparrow$	Context gap $\uparrow$	Context R@1 $\uparrow$	Predicted R@1 $\uparrow$	Eff. rank
Full AeroJEPA	**0.948**	**0.962**	**0.839**	0.596	**0.822**	**0.661**	**1.341**	**100%**	**92%**	**4.68**
Reconstruction only	0.886	0.873	0.621	**0.739**	0.736	0.621	0.800	88%	4%	1.76

The $\lambda_s = 10^{-3}$ configuration provides the strongest overall combination of design recovery, aerodynamic recovery, and reconstruction accuracy. Increasing the regularization weight progressively reduces the probe and reconstruction results, while the contrastive alternative substantially degrades both field prediction and the organization of the geometry representation. We therefore use $\lambda_s = 10^{-3}$ for the primary AeroJEPA models.

We next compare the complete AeroJEPA objective with reconstruction-only training on HiLiftAeroML in Table 6. Both models use the same architecture and are evaluated after 30 epochs, so this experiment isolates the effect of latent matching at an equal training budget. These matched early checkpoints are distinct from the fully trained model used for the primary HiLiftAeroML results.

Reconstruction-only training retains substantial linearly accessible information, confirming that the architecture itself provides a strong inductive bias. The complete objective nevertheless improves three of the four control-surface probes, both aerodynamic-change probes, and the organization of both latent spaces. The largest difference appears in predicted-token retrieval, which increases from 4% to 92%, while context retrieval increases from 88% to 100%. These results show that latent matching provides a substantially clearer organization of the predicted aerodynamic representation while preserving reconstruction capability.

D.5 Latent-space aerodynamic optimization

We evaluate whether the learned geometry representation can serve as a differentiable search space for preliminary aerodynamic design exploration. On SuperWing, we maximize aerodynamic efficiency C_L/C_D at a fixed cruise condition $(\alpha_{\text{cruise}}, M_{\text{cruise}})$ by treating the 128-dimensional mean-pooled context representation z_{ctx} as the optimization variable. The search uses neither CAD modification nor field decoding inside the optimization loop.

Differentiable objective. The frozen AeroJEPA predictor maps the context representation and operating condition to a predicted aerodynamic representation,

$$z_{\text{pred}} = \Phi_{\text{pred}}\left(z_{\text{ctx}}, \alpha_{\text{cruise}}, M_{\text{cruise}}\right). \tag{8}$$

Linear ridge probes, defined in Appendix D.7, then recover C_L and C_D from z_{pred}. Their composition produces an end-to-end differentiable map from geometry representation to aerodynamic coefficients. We optimize

$$\min_{z_{\text{ctx}} \in \mathbb{R}^{128}} J(z_{\text{ctx}}) = -\frac{C_L(z_{\text{ctx}})}{C_D(z_{\text{ctx}})}, \tag{9}$$

with gradients obtained by backpropagation through the frozen predictor and probes.

Search constraints. Unconstrained optimization could move the representation outside the region supported by the training data. We therefore solve Equation 9 using SLSQP subject to four families of constraints derived from training-set statistics. First, a Mahalanobis trust region restricts the search around the training distribution,

$$\left(z_{\text{ctx}} - \mu_{\text{train}}\right)^{\top} \Sigma_{\text{train}}^{-1} \left(z_{\text{ctx}} - \mu_{\text{train}}\right) \leq \tau. \tag{10}$$

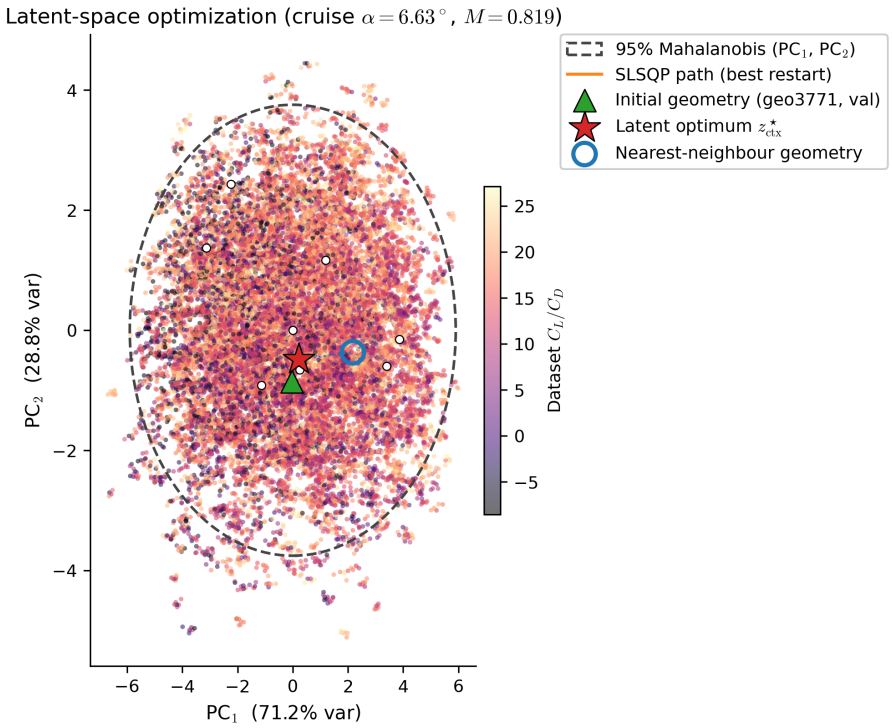

Figure 13: Latent-space optimization trajectories. The training context representations are projected onto two principal components and colored by dataset C_L/C_D. Gray curves show the eight SLSQP restarts, and the orange curve highlights the restart producing the selected optimum. White circles indicate initial points, the red star indicates $z^\star_{\mathrm{ctx}}$, and the blue circle indicates the dataset geometry retrieved as its nearest neighbor in design space. The dashed contour shows the projected Mahalanobis trust region.

Second, we constrain the nine SuperWing design parameters that are reliably recovered from the context representation, selected using five-fold cross-validated $R^2 \geq 0.85$,

$$x_k^{\min} \leq W_k^\top \widetilde{z}_{\mathrm{ctx}} + b_k \leq x_k^{\max}. \tag{11}$$

Third, we impose a drag floor $C_D \geq 0.9 \min_{\mathrm{train}} C_D$ and lift and drag ceilings equal to 1.05 times their respective training-set maxima. Finally, the predicted L/D is bounded by the maximum observed in the dataset. Together, these constraints restrict the search to a calibrated region of the geometry and aerodynamic representations.

We perform eight random restarts. Initial context representations are drawn from the training distribution and projected inside the Mahalanobis trust region. Jacobians for the affine design constraints and Mahalanobis constraint are supplied analytically, while the aerodynamic gradients are obtained using automatic differentiation. Once SLSQP converges to $z^\star_{\mathrm{ctx}}$, the corresponding design vector is recovered using the linear design probe,

$$x^\star = W_{\mathrm{design}} \widetilde{z}^\star_{\mathrm{ctx}} + b_{\mathrm{design}}. \tag{12}$$

We then retrieve the dataset geometry whose standardized design vector is closest to $x^\star$ in Euclidean distance.

Optimization behavior. All eight restarts converge to the same neighborhood while satisfying the imposed constraints. Figure 13 shows the optimization trajectories in a PCA projection of the training context representations. The SLSQP iterations move toward larger probe-predicted C_L/C_D values while remaining inside the trust region. The final latent solution lies close to its retrieved dataset geometry, indicating that the optimization terminates near a region represented by observed wings.

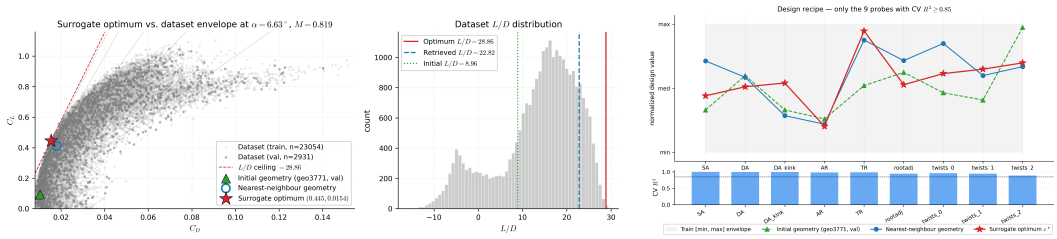

Figure 14: Constrained latent-space optimization on SuperWing. *Left:* the probe-predicted optimum approaches the high-efficiency region of the observed (C_D, C_L) envelope and lies near the right tail of the dataset L/D distribution. *Right:* parallel-coordinates comparison of the initial design, the recovered latent optimum, and its nearest observed wing over the nine reliably decoded design parameters. The nearest observed wing follows a similar parameter profile to the optimized latent point.

Figure 14 relates the optimized representation to the aerodynamic envelope and design parameters. The predicted optimum moves from its initialization toward the high-efficiency region of the observed (C_D, C_L) distribution while satisfying the imposed lift, drag, and L/D constraints. The recovered design combines a large reference area and aspect ratio, high sweep, aggressive taper, and root-biased washout. Its nearest observed neighbor follows a similar profile over the nine reliably recovered design parameters, connecting the optimized representation to a wing configuration present in the dataset.

The experiment shows that the context representation provides a smooth and differentiable search space, that its linear design and aerodynamic readouts can support constrained optimization, and that the resulting point can be translated into an interpretable design parameterization. The optimization requires only the frozen predictor and lightweight linear probes, avoiding repeated field decoding during the search. Because the synthetic optimum has not been evaluated through a new CFD simulation, this experiment establishes the feasibility of representation-based design exploration rather than verified aerodynamic improvement.

D.6 Concept-vector latent arithmetic on HiLiftAeroML

A semantic latent space should not only support gradient-based optimization (App. D.5) but also expose interpretable *directions* along which design intent can be expressed by simple vector arithmetic. We test this on HiLiftAeroML by asking whether the AeroJEPA's 64-dimensional (mean) context latent z_{ctx} encodes high-lift control surfaces – inboard / outboard flap and slat deflections – along separable linear axes, despite never having been trained on those labels.

Concept directions from linear probes. We fit a single ridge probe per design parameter x_k on the train split, $x_k = w_k^\top \tilde{z}_{\text{ctx}} + b_k$ with $\tilde{z}_{\text{ctx}} = (z_{\text{ctx}} - \mu_{\text{ctx}})/\sigma_{\text{ctx}}$, where the design parameters are the four main HiLiftAeroML deflections (IB Flap, OB Flap, IB Slat, OB Slat) plus the two flap-gap multipliers. The unit-norm probe direction $v_k = w_k/\|w_k\|$ is interpreted as the latent "concept vector" for parameter x_k. To probe the disentanglement of the representation, we walk the train-mean latent along one direction at a time,

$$z_{\text{ctx}}(\gamma) \;=\; \mu_{\text{ctx}} + \gamma\, v_k, \tag{13}$$

and at each step γ predict *all* design parameters from the shifted latent through the same probe matrix. The slope of the predicted parameter x_j along the walk on direction v_k has a closed form,

$$\frac{\mathrm{d}\tilde{x}_j}{\mathrm{d}\gamma} = \frac{\big(W_{\text{ctx}}\,(v_k/\sigma_{\text{ctx}})\big)_j}{\sigma_{x_j}}, \tag{14}$$

expressed in train-set standard deviations of x_j per unit step in latent space; we report this number as the panel-level "σ/γ" sensitivity.

The disentanglement matrix. Figure 15 shows the resulting 4×4 matrix for the four main deflections. Rows index the latent direction walked; columns index the design parameter read out.

The diagonal panels (highlighted in red) measure the *intended* response of each concept direction: walking the IB-Flap concept vector should change predicted IB-Flap deflection. The off-diagonal panels measure *cross-talk*: how much an unrelated parameter moves as a side-effect.

The structure in Figure 15 shows that each main deflection has a strongly aligned latent direction. The diagonal responses are substantially larger than most off-diagonal responses, while flap-to-slat and slat-to-flap interactions remain small. The main residual couplings occur between the inboard and outboard elements of the same control surface. Thus, the representation separates flap and slat variations while retaining the correlations between related inboard and outboard elements observed in the dataset.

The probes used to define every concept direction are linear and trained *only* on the train split, with no design-parameter supervision ever seen by the AeroJEPA itself. Yet in the resulting representation, design intent maps onto linear directions, and the residual entanglement is not arbitrary noise but the physically meaningful coupling that engineers already build into multi-element high-lift systems. Three points follow:

1. The geometry-only context encoder *discovers* the underlying parametric basis of the design space without being told what it is. This is the practical signal that the latent space is genuinely semantic rather than merely high-capacity.
2. Concept-vector arithmetic provides a cheap interface for exploring configurations: a designer can push along the OB-Slat direction and read the predicted flap and slat values, getting an immediate, calibrated preview of what changes co-occur in real wings.
3. The *block-diagonal disentanglement structure* between flap and slat axes is a constructive ingredient for the optimization in App. D.5: when a constrained search needs to move design parameters independently, the latent space provides near-orthogonal directions to do so along the parameters that are geometrically separable, while preserving the natural coupling between IB and OB segments where the dataset itself shows it.

Together with the gradient-based optimization of App. D.5, this experiment characterizes the AeroJEPA latent both as a *search space* (smooth, gradient-friendly, linearly decodable) and as an *interpretation space* (separable concept directions reflecting physical actuation modes).

D.7 Linear-probe analysis of the AeroJEPA latent space

The aerodynamic optimization in App. D.5 and the concept-vector arithmetic in App. D.6 both rely on a single, simple primitive: a *linear* readout from the AeroJEPA representations. This section defines that readout, lists the three latent quantities it is applied to, and reports the goodness-of-fit numbers that justify treating those readouts as semantic.

Probe definition. For a latent vector $z \in \mathbb{R}^D$ and a scalar target y, we define a *linear probe* as the composition of train-set standardisation and ridge regression with cross-validated regularisation,

$$\hat{y}(z) \;=\; w^\top \tilde{z} + b, \qquad \tilde{z} \;=\; \frac{z - \mu_{\text{train}}}{\sigma_{\text{train}}}, \tag{15}$$

where (w, b) are obtained by minimising $\|\tilde{Z}w + b\,\mathbf{1} - y\|_2^2 + \lambda\,\|w\|_2^2$ on the train split, with λ selected by an inner cross-validation over $\lambda \in \{10^{-4}, 10^{-3}, \ldots, 10^{4}\}$.

Three latent quantities, three probe families. We extract three mean-pooled latent representations per case, with dimension $D = 64$ on HiLiftAeroML and $D = 128$ on SuperWing:

- *Context representation* $z_{\text{ctx}} \in \mathbb{R}^D$, obtained by mean-pooling the context tokens produced from the geometry, before conditioning on the operating variables.
- *Predicted representation* $z_{\text{pred}} \in \mathbb{R}^D$, obtained by mean-pooling the predictor output after conditioning the geometry representation on the operating variables.
- *Target representation* $z_{\text{tgt}} \in \mathbb{R}^D$, obtained by mean-pooling the target-encoder tokens produced from the reference flow field. It provides the latent prediction target during training and is discarded at inference.

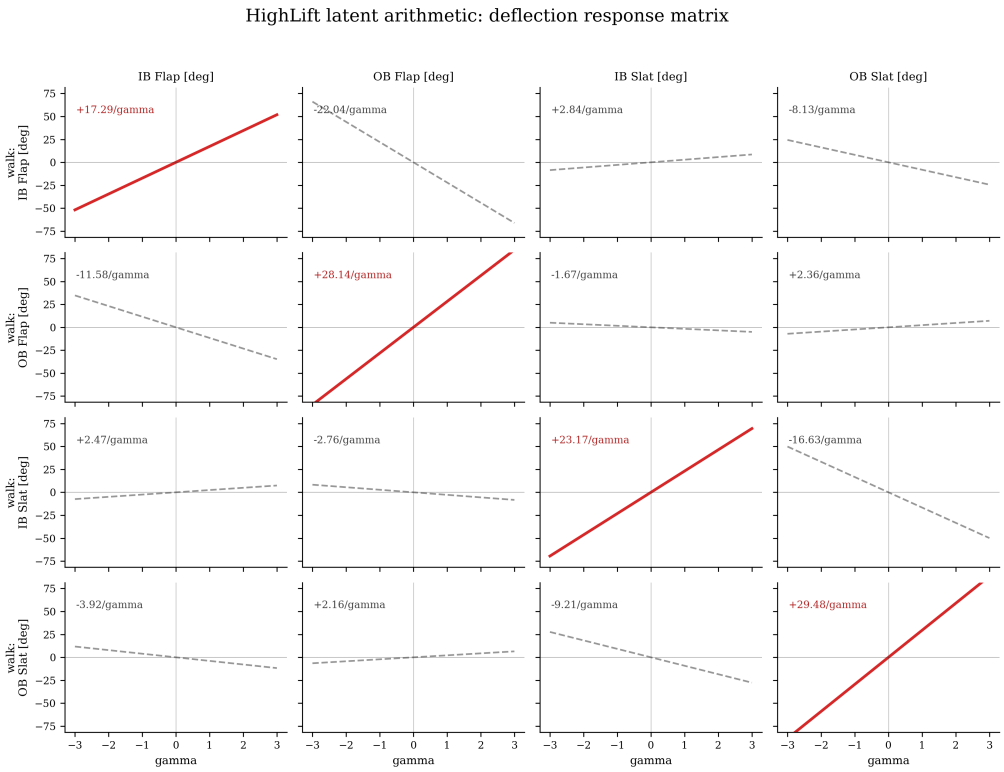

Figure 15: **Concept-vector arithmetic in the HiLiftAeroML context latent.** 4×4 response matrix obtained by walking the train-mean latent μ_{ctx} along the unit-norm linear-probe direction of one design parameter at a time. *Rows:* latent direction walked (IB Flap, OB Flap, IB Slat, OB Slat). *Columns:* design parameter read out by the corresponding linear probe. Each panel reports the sensitivity in train standard deviations per unit walk (σ/γ). Diagonal panels (red) show the intended concept response and reach near-unit slope. Off-diagonal panels (grey, dashed) quantify cross-talk between concept directions: the flap–slat off-diagonal block is essentially flat, indicating limited cross-talk between the flap and slat directions, while inboard / outboard pairs of the same control surface partially co-activate – the same correlation engineers impose by deflecting IB and OB segments in tandem. The disentanglement is obtained *without any design-parameter supervision* during AeroJEPA training; the linear probes are fit only on the train split.

The following three probe families evaluate which quantities are linearly accessible from the context and predicted representations:

1. **Context $\rightarrow$ design parameters.** For each design parameter x_k (e.g. each of the 54 SuperWing morphological coefficients or the 8 HiLiftAeroML control-surface deflections), a separate probe with $z = z_{\text{ctx}}$ and $y = x_k$. The full set of K probes assembles into a single matrix readout $x \approx W_{\text{ctx}}\, \tilde{z}_{\text{ctx}} + b_{\text{ctx}}$ because every probe shares the same standardisation $(\mu_{\text{ctx}}, \sigma_{\text{ctx}})$.
2. **Predicted $\rightarrow$ aerodynamic coefficients.** Two probes, $z = z_{\text{pred}}$, $y \in \{C_L, C_D\}$.
3. **Predicted $\rightarrow$ AoA (control).** A single probe with $z = z_{\text{pred}}$ and $y = \alpha$, used to verify that the predictor output – but not the geometry-only context – carries flow-condition information.

What the probes report on each dataset. On SuperWing, the primary configuration achieves a mean predicted-representation R^2 of 0.993 for C_L and C_D, and a mean R^2 of 0.987 for their within-geometry changes, as reported in Appendix D.4. Eight selected design parameters are recovered

from the predicted representation with a mean R^2 of 0.886. Separately, nine of the 54 SuperWing design parameters satisfy the cross-validated threshold $R^2 \geq 0.85$ when decoded from the context representation, and these parameters define the design constraints used in Appendix D.5. On the fully trained HiLiftAeroML model, the four dominant control-surface deflections are recovered from the context representation with R^2 values between 0.965 and 0.988, while the predicted representation supports linear recovery of C_L and C_D with $R^2 = 0.930$ and 0.996, respectively.

From probes to a downstream tool. Equation equation 15 has two properties that the rest of the paper exploits. First, it is *closed-form differentiable* in z – the gradient $\partial\hat{y}/\partial z = w/\sigma_{\text{train}}$ is just a rescaled probe vector. This is what allows the SLSQP optimization of App. D.5 to back-propagate through the predictor and the probes in a single autograd graph, with the design-bound constraints expressed as affine inequalities with closed-form Jacobians. Second, the unit-norm probe direction $v_k = w_k/\|w_k\|$ is interpreted as the latent *concept vector* for parameter x_k; walking z_{ctx} along v_k and reading out every other parameter gives the disentanglement matrix of App. D.6.

The probes are the bridge between the learned representation and the design-space quantities a domain expert cares about. They turn the AeroJEPA latents into a *usable* object – one that can be searched (App. D.5) and queried by analogy (App. D.6) - without ever fine-tuning the underlying network.