LEARNING TRANSFERABLE AERODYNAMIC LAWS ACROSS PHYSICAL DOMAINS FOR HETEROGENEOUS FLAPPING-WING SYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Aerodynamic force prediction across flapping-wing platforms must contend with systems that share governing physics yet differ in geometry, aerodynamic coefficients, actuation dynamics, and observable state spaces. Direct force measurement on a target platform is costly and intrusive: sensors, wiring and mounting structures perturb the very dynamics being measured,so experimental data is scarce precisely where simulation is abundant. We study data-efficient learning across such heterogeneous physical domains. We first establish, independently of any learned model, that our domains obey a common non-dimensional law: after absorbing a single scalar factor their binned normal-force coefficients coincide over the overlapping angle-of-attack range ($\rho = 0.98$, shape $R^2 = 0.95$). We then introduce a predictor that represents kinematics in stroke-phase rather than absolute time and factorises into a domain-private input map, a shared law operator, and a domain-private output scale. Without target force labels it attains $R^2 = 0.76$; with 128 target trajectories, $R^2 = 0.96$, against 0.87 for matched target-only training under an identical data budget. Our method outperforms five published domain-adaptation and fine-tuning methods, each tuned over a four-decade sweep of its alignment strength, while using 12k parameters and training 16–20$\times$ faster than the strongest of them. Controls with shuffled labels, shuffled conditions and relearned law operators confirm that the gain comes from transferred physical structure rather than generic regularisation: sharing the law operator is the one component whose removal costs accuracy at every data budget, by up to 0.12 R^2.

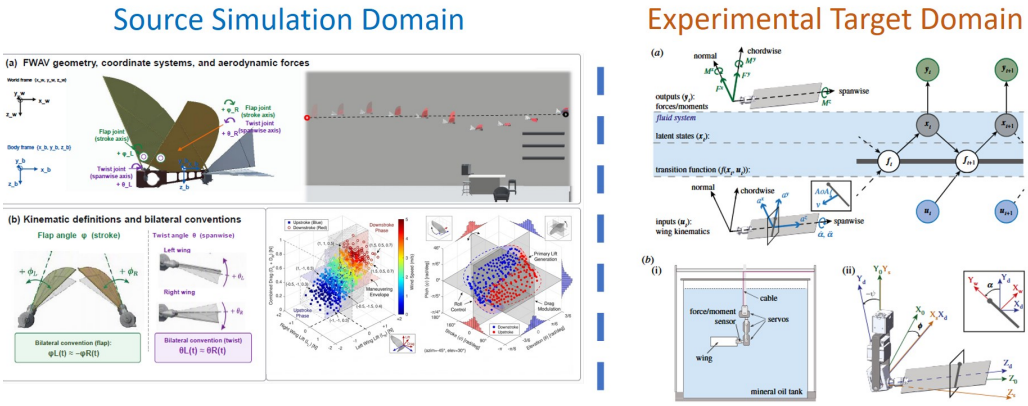


Figure 1: Overview of the heterogeneous aerodynamic-transfer setting. The source simulation domain provides synchronized bilateral wing kinematics and aerodynamic forces across randomized operating conditions. The experimental target domain comprises a dynamically scaled robotic wing with wing kinematics as inputs, aerodynamic forces and moments as outputs, and latent states capturing the unsteady fluid dynamics. CLAW transfers the shared nonlinear aerodynamic response while retaining domain-private input and force calibration. The experimental-system schematics on the right are adapted from Bayiz & Cheng (2021b).

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.

1 INTRODUCTION

Flapping-wing flight generates aerodynamic forces through unsteady, three-dimensional vortex flows that resist closed-form description (Sane, 2003b). Control of insect-scale fliers nevertheless requires an instantaneous force model, and fitting one to a new platform is expensive. Force measurement on a physical flapping wing requires a six-axis sensor, a controlled fluid environment, and post-processing to remove inertial, buoyancy and gravitational components (Bayiz & Cheng, 2021b); worse, the sensor, its wiring and its mounting perturb the wing dynamics being measured. Simulation is nearly free and non-intrusive.

This asymmetry motivates learning across heterogeneous physical domains: use abundant simulation to learn the aerodynamic law, and a small experimental budget to identify what is specific to the target platform. The difficulty is that flapping-wing platforms share governing physics while differing in geometry, kinematic range, fluid medium and observable state space. In the pair we study, a rigid-wing simulator and the dynamically scaled robotic wing of Bayiz & Cheng (2021b), Reynolds number matches ($\mathrm{Re} \approx 10^3$) but flapping frequency differs by $50\times$, pitching rate by two orders of magnitude, and fluid density and wing area by factors that are never observed directly. A predictor trained on one domain and applied to the other is therefore reading inputs whose numerical scale is entirely unfamiliar, even though the underlying relation is unchanged.

Our approach CLAW (Cross-domain Law-Aware Wing-force Transfer) separates what transfers from what does not. We represent kinematics in stroke phase rather than absolute time, which removes the frequency gap by construction, and factorise the predictor into a domain-private input map, a shared law operator, and a domain-private output scale. Crucially, we do *not* align the two domains' feature distributions: the distributional gap here encodes the operating regime that the law consumes, not acquisition nuisance. Appendix 9 provides a complete visual summary of the source–target systems, phase-domain representation, and factorized architecture. Contributions are:

1. **A model-independent test that transfer is physically possible.** Before any learning, we show the two domains' normal-force coefficients coincide once a single scalar factor is absorbed ($\rho = 0.98$, shape $R^2 = 0.95$; §6.1). This turns "does it transfer?" from a question about the learner into a question about the method.
2. **A predictor that transfers the law and not the units** (§4), reaching $R^2 = 0.958$ with 128 target trajectories and $R^2 = 0.761$ with no target force labels, against 0.865 for target-only training under the same budget.
3. **A margin over five published methods** (§6.4), including two designed specifically for regression, each tuned over a four-decade sweep of its alignment strength, while using 12k parameters and training 16–20$\times$ faster than the strongest of them.

2 RELATED WORK

Flapping-wing aerodynamic modelling. Flapping loads arise from coupled translational, rotational, added-mass and wake-history effects, producing phase-dependent responses that steady or memoryless formulations represent poorly (Dickinson et al., 1999; Sane & Dickinson, 2002; Sane, 2003a). Quasi-steady formulations calibrated against revolving-wing measurements (Sane & Dickinson, 2002) or against CFD (Nakata et al., 2015) predict instantaneous forces well within the platform they were fitted to, and data-driven state-space models extend this to learned dynamics (Bayiz & Cheng, 2021b). All are identified within a single platform and flow regime. **Domain adaptation for regression.** Most adaptation methods align feature distributions by moment matching (Sun & Saenko, 2016; Long et al., 2015) or adversarial training (Ganin et al., 2016), and their strongest variants derive objectives from classifier outputs that do not extend to scalar regression. Two methods target regression directly: RSD aligns principal angles between feature subspaces, motivated by the observation that regression, unlike classification, is sensitive to feature scaling (Chen et al., 2021); DARE-GRAM aligns the inverse Gram matrices that determine the least-squares solution (Nejjar et al., 2023). Both retain the premise that source and target representations should coincide. We show that in this setting the premise itself is the obstacle. **Transfer by fine-tuning.** Full fine-tuning can distort pretrained features under distribution shift, motivating linear probing first (Kumar et al., 2022), adapting only the layers matched to the shift (Lee et al., 2023), interpolating pretrained and

fine-tuned weights (Wortsman et al., 2022), or restricting updates to biases (Ben Zaken et al., 2022). We evaluate all of these on our architecture, so that the comparison isolates the adaptation strategy.

Learning across physical domains. Adaptation in flight systems has largely meant online adaptation on real data: domain-adversarially invariant meta-learning with online least-squares (O'Connell et al., 2022), or a blade-element model with a learned residual (Bauersfeld et al., 2021). Operator learning transfers between simulated domains under conditional shift (Goswami et al., 2022).

3 Physical Background

For a rigid flat-plate wing, we model the aerodynamic force as the sum of quasi-steady and added-mass contributions:

$$F_d = \tfrac{1}{2}\rho_d S_d |v|^2 C_N(\alpha) + \rho_d c_d^2 b_d \lambda_d \left[\tfrac{\pi}{4}\left(\dot{v}\sin\alpha + v\dot{\alpha}\cos\alpha\right) - \tfrac{\pi}{16} c_d \ddot{\alpha}\right], \tag{1}$$

where d indexes the physical platform. The normal-force coefficient $C_N(\alpha) = C_L(\alpha)\cos\alpha + C_D(\alpha)\sin\alpha$ is monotone over the operating range and odd in α. The added-mass contribution introduces explicit dependence on recent kinematic history. Appendix 11 gives the complete coefficient model, calibration procedure, and operating-regime analysis.

Equation equation 1 separates platform-specific fluid and geometric factors from a common nonlinear dependence on kinematics. This structure motivates CLAW's three components: a domain-private input map accommodates platform-specific kinematic ranges, a shared nonlinear operator transfers the aerodynamic response, and a domain-private affine output map recovers the force scale and offset of each platform. Because the two systems evaluate the shared response over different regions of kinematic space, their distributional difference encodes the physical operating regime rather than an acquisition artifact. CLAW therefore transfers the aerodynamic operator without forcing the source and target feature distributions to coincide.

The systems operate at comparable Reynolds numbers but substantially different frequencies:

$$\frac{\mathrm{Re}_{\mathrm{src}}}{\mathrm{Re}_{\mathrm{tgt}}} \approx 1.3, \qquad \frac{k_{\mathrm{src}}}{k_{\mathrm{tgt}}} \approx 38. \tag{2}$$

CLAW consequently transfers across both disjoint kinematic ranges and different temporal regimes.

4 Method

4.1 Problem Formalization

Let $\mathcal{U}$ denote the kinematic information represented by a local phase window. The window explicitly contains angle of attack, flow speed, and wing rotation, while temporal changes such as $\dot{v}$, $\dot{\alpha}$, and $\ddot{\alpha}$ are represented through variation across adjacent phase samples. A domain d is characterized by a kinematic distribution P_d and a force response $F_d : \mathcal{U} \to \mathbb{R}$. Motivated by Eq. equation 1, we model

$$F_d(u) = g_d C(T_d(u)) + \eta_d + \varepsilon_d(u), \tag{3}$$

where T_d is a domain-private transformation, C is a shared nonlinear response, (g_d, η_d) are private output-calibration parameters, and ε_d contains platform-specific effects not represented by the factorization. The transfer hypothesis is that C can be estimated from abundant source data and reused in the target domain after adapting T_d, g_d, and η_d. Appendix 12 illustrates the private encoders, shared aerodynamic operator, and private output maps.

Given n_{src} labeled source samples and $n_{\mathrm{tgt}} \ll n_{\mathrm{src}}$ labeled target samples, the objective is to minimize the target risk

$$\mathcal{R}_{\mathrm{tgt}}(\widehat{F}_{\mathrm{tgt}}) = \mathbb{E}_{u\sim P_{\mathrm{tgt}}}\left[\ell\Big(\widehat{F}_{\mathrm{tgt}}(u), F_{\mathrm{tgt}}(u)\Big)\right]. \tag{4}$$

The design addresses three properties of the source-target pair: unknown relative force scale, limited kinematic overlap, and noncommensurate sampling and stroke frequencies. These motivate private output calibration, private input adapters, and phase-domain resampling, respectively.

4.2 PHASE-DOMAIN REPRESENTATION

For a trajectory with dominant stroke frequency f_0, define the normalized phase and phase-resampling operator jointly as

$$\varphi(t) = (f_0 t) \bmod 1, \qquad \Pi_{n_\varphi} j = u\left(t\left(\frac{j}{n_\varphi}\right)\right), \quad j = 0, \ldots, n_\varphi - 1. \tag{5}$$

For odd window width w, the cyclic local window is

$$W_j = \left[\Pi_{n_\varphi}[u]((j+r) \bmod n_\varphi)\right]_{r=-\lfloor w/2 \rfloor}^{\lfloor w/2 \rfloor} \in \mathbb{R}^{w \times d_u}, \tag{6}$$

where d_u denotes the number of kinematic channels.

Lemma 1 (Frequency Invariance)**.** *For trajectories with stroke frequencies $f_0^{(1)} \neq f_0^{(2)}$, Eq. equation 6 spans the same fraction w/n_φ of a stroke, and an offset r represents the same phase increment r/n_φ.*

Proof. The operator Π_{n_φ} is defined over the dimensionless phase $\varphi \in [0, 1)$. Although the physical duration $w/(n_\varphi f_0)$ varies across trajectories, the corresponding phase support does not. □

Phase resampling therefore provides a fixed-width representation across different physical time scales. The discarded frequency is retained through the cycle condition and continuous phase encoding

$$c = [f_0, A_\phi, A_\alpha, \Delta\psi] \in \mathbb{R}^4, \qquad p(j) = \left[\cos\frac{2\pi j}{n_\varphi}, \sin\frac{2\pi j}{n_\varphi}, \cos\frac{4\pi j}{n_\varphi}, \sin\frac{4\pi j}{n_\varphi}\right] \in \mathbb{R}^4. \tag{7}$$

The first vector contains the stroke frequency, stroke amplitude, pitch amplitude, and stroke-to-pitch phase lead. The second remains continuous at the phase boundary.

Signed angle convention. The angle of attack remains signed because $C_N(-\alpha) = -C_N(\alpha)$. Replacing α by $|\alpha|$ would require the predictor to infer force sign from the remaining variables. Phase provides a possible surrogate, but the relation between phase and angle sign depends on the platform-specific pitch schedule. Section 10 evaluates the loss of cross-domain consistency.

4.3 FACTORIZED PREDICTOR

For domain d, the private adapter and complete predictor are

$$h_d = \tanh(\mathbf{A}_d \operatorname{vec}(W) + \beta_d), \qquad \widehat{F}_d(W, c, p) = g_d \mathcal{L}(h_d, c, p) + \eta_d, \tag{8}$$

where $(\mathbf{A}_d, \beta_d)$ and (g_d, η_d) are domain-private. The shared nonlinear operator is

$$\mathcal{L}(h, c, p) = \mathbf{w}^\top \sigma(\mathbf{W}_3 \sigma(\mathbf{W}_2 \sigma(\mathbf{W}_1 [h; c; p] + b_1) + b_2) + b_3) + b_4, \tag{9}$$

where σ is the SiLU activation, $\mathbf{W}_1 \in \mathbb{R}^{m \times (d_h + 8)}$, and $\mathbf{W}_2, \mathbf{W}_3 \in \mathbb{R}^{m \times m}$.

The private adapter accommodates domain-dependent kinematic ranges, the shared operator represents the transferable nonlinear response, and the private affine output map accounts for residual amplitude and offset. The adapter is intentionally capacity-limited: a single affine projection followed by a fixed elementwise nonlinearity is less able than a deep private network to represent the complete aerodynamic response independently. This encourages the shared operator to capture higher-order interactions among kinematics, cycle condition, and phase. Section 6.6 evaluates this design choice.

Proposition 1 (Exact Recovery Under Channel-Wise Affine Output Shift)**.** *Let*

$$\mathbf{F}_{\mathrm{tgt}}(u) = \mathbf{g}_{\mathrm{tgt}} \odot \mathbf{z}(u) + \boldsymbol{\eta}_{\mathrm{tgt}}, \quad \mathbf{z}(u) = C\big(T_{\mathrm{tgt}}(u)\big), \odot \quad \mathbf{F}_{\mathrm{tgt}}(u), \mathbf{z}(u), \mathbf{g}_{\mathrm{tgt}}, \boldsymbol{\eta}_{\mathrm{tgt}} \in \mathbb{R}^d,$$

and $\odot$ denotes element-wise multiplication. Suppose two labeled target samples u_1 and u_2 satisfy

$$z_j(u_1) \neq z_j(u_2) \qquad \textit{for every } j \in \{1, \ldots, d\}.$$

Then the channel-wise gain vector $\mathbf{g}_{\mathrm{tgt}}$ and offset vector $\boldsymbol{\eta}_{\mathrm{tgt}}$ are uniquely identifiable in the noise-free setting. If $\boldsymbol{\eta}_{\mathrm{tgt}} = \mathbf{0}$, then one labeled sample u satisfying

$$z_j(u) \neq 0 \qquad \textit{for every } j \in \{1, \ldots, d\}$$

uniquely identifies $\mathbf{g}_{\mathrm{tgt}}$.

Proof. For each output channel $j \in \{1, \dots, d\}$, the two labeled samples satisfy

$$\begin{bmatrix} z_j(u_1) & 1 \\ z_j(u_2) & 1 \end{bmatrix} \begin{bmatrix} g_{\mathrm{tgt},j} \\ \eta_{\mathrm{tgt},j} \end{bmatrix} = \begin{bmatrix} F_{\mathrm{tgt},j}(u_1) \\ F_{\mathrm{tgt},j}(u_2) \end{bmatrix}.$$

The determinant of the coefficient matrix is

$$z_j(u_1) - z_j(u_2),$$

which is nonzero by assumption. Therefore, the system has a unique solution for $\left(g_{\mathrm{tgt},j}, \eta_{\mathrm{tgt},j}\right)$ in every output channel j. Applying this argument independently to all d channels uniquely determines $\mathbf{g}_{\mathrm{tgt}}$ and $\boldsymbol{\eta}_{\mathrm{tgt}}$. If $\boldsymbol{\eta}_{\mathrm{tgt}} = \mathbf{0}$ and $z_j(u) \neq 0$ for every j, then

$$g_{\mathrm{tgt},j} = \frac{F_{\mathrm{tgt},j}(u)}{z_j(u)} \qquad \text{for every } j \in \{1, \dots, d\}.$$

Hence, $\mathbf{g}_{\mathrm{tgt}}$ is uniquely determined. □

4.4 Conflict between Distribution Alignment and Law Sharing

A common adaptation objective augments supervised source prediction with latent distribution alignment:

$$\begin{aligned} \min_{\mathcal{A}_{\mathrm{src}},\, \mathcal{A}_{\mathrm{tgt}},\, \mathcal{L}} \quad & \mathbb{E}_{u \sim P_{\mathrm{src}}} \left[\ell \left(\widehat{F}_{\mathrm{src}}(u), F_{\mathrm{src}}(u) \right) \right] \\ & + \lambda\, \mathcal{D}\left((\mathcal{A}_{\mathrm{src}})_{\#} P_{\mathrm{src}} \parallel (\mathcal{A}_{\mathrm{tgt}})_{\#} P_{\mathrm{tgt}} \right). \end{aligned} \tag{10}$$

where $\mathcal{D}$ may be an MMD (Long et al., 2015), moment discrepancy (Sun & Saenko, 2016), adversarial divergence (Ganin et al., 2016), or subspace discrepancy (Chen et al., 2021; Nejjar et al., 2023). Alignment is compatible with law sharing only if matching latent distributions does not erase differences in response distribution.

Proposition 2 (Alignment–Law Incompatibility)**.** *Suppose that $\mathcal{L}(\mathcal{A}_d(u)) = C(u)$ for P_d-almost every u in both domains and that $(\mathcal{A}_{\mathrm{src}})_{\#} P_{\mathrm{src}} = (\mathcal{A}_{\mathrm{tgt}})_{\#} P_{\mathrm{tgt}}$. Then $C_{\#} P_{\mathrm{src}} = C_{\#} P_{\mathrm{tgt}}$.*

Proof. Since $C = \mathcal{L} \circ \mathcal{A}_d$ almost everywhere,

$$C_{\#} P_d = \mathcal{L}_{\#} (\mathcal{A}_d)_{\#} P_d. \tag{11}$$

Applying $\mathcal{L}_{\#}$ to the common latent distribution proves the result. □

Corollary 1. *If $C_{\#} P_{\mathrm{src}} \neq C_{\#} P_{\mathrm{tgt}}$, exact law consistency and exact latent-distribution alignment cannot hold simultaneously.*

The premise is empirically testable because the common response depends on angle of attack through Eq. equation 16, while the domains realize different angle and derivative distributions. Section 6.2 evaluates this condition, and Sec. 6.4 tests the predicted trade-off.

4.5 Training and Target Calibration

Let $\theta = (\mathbf{W}_{1:3}, \mathbf{w}, b_{1:4})$ denote the shared and $\psi_d = (\mathbf{A}_d, \beta_d, g_d, \eta_d)$ the domain-private parameters.

Stage 1: source training. We estimate

$$(\widehat{\theta}, \widehat{\psi}_{\mathrm{src}}) = \arg\min_{\theta, \psi_{\mathrm{src}}} \left[\frac{1}{n_{\mathrm{src}}} \sum_{i=1}^{n_{\mathrm{src}}} H_\delta \left(\widehat{F}_{\mathrm{src}}(W_i, c_i, p_i) - y_i \right) + \omega \|\theta\|_2^2 \right]. \tag{12}$$

where H_δ is the Huber loss. No source-target distribution-matching term is included.

Stage 2: target adaptation. For labeled target trajectories, we initialize $\psi_{\mathrm{tgt}}^{(0)} \leftarrow \widehat{\psi}_{\mathrm{src}}$ and $\theta^{(0)} \leftarrow \widehat{\theta}$, then optimize the target objective under the matched adaptation protocol.

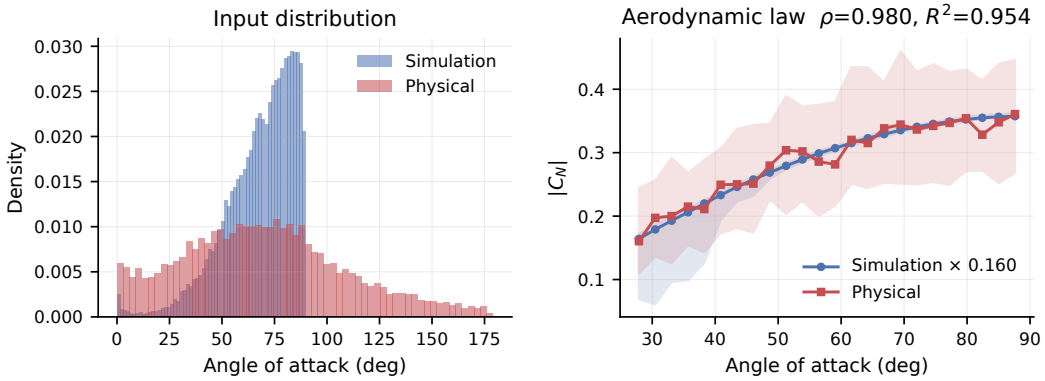


Figure 2: Cross-domain aerodynamic-law agreement. Despite different input distributions, the scaled $|C_N|$ responses coincide over the shared angle-of-attack range.

Scale-only target calibration. When target labels are used only for output calibration, the shared operator and target adapter are frozen and

$$(\widehat{g}_{\mathrm{tgt}}, \widehat{\eta}_{\mathrm{tgt}}) = \arg\min_{g,\eta} \sum_{i=1}^{b} H_\delta(g z_i + \eta - y_i)\,, \qquad z_i = \mathcal{L}\Big(\mathcal{A}_{\mathrm{tgt}}(W_i), c_i, p_i; \widehat{\theta}\Big)\,. \tag{13}$$

This is a two-parameter convex calibration problem and is distinct from strict zero-shot transfer.

Normalization. Scale statistics are estimated from the source partition and applied unchanged. Inputs and targets are scaled but not centered, preserving the physical zero implied by equation 14 and equation 18. Appendix 15.1 reports the estimators and target-domain robust-scale convention.

5 Experimental Setup

Source domain. The source system comprises a rigid pair of flat-plate wings flapping about a fixed base in air. The simulator runs at 500 Hz and evaluates aerodynamic forces using Eqs. equation 14 and equation 18. Each episode independently samples five kinematic parameters, while all six body degrees of freedom remain constrained so that the recorded loads are purely aerodynamic. The complete configuration, calibration procedure, and randomization ranges are reported in Appendix 13.

Target domain. The target domain uses the public dataset of Bayiz & Cheng (2021a). Each trajectory contains 470 samples, with inertial bias and buoyancy removed by the original authors (Bayiz & Cheng, 2021b). Because the released channel ordering differs from the published specification, Appendix 14 documents the channel-recovery procedure and validation checks.

Phase-domain assembly. Both domains are processed using the phase representation in Sec. 4.2, with $n_\varphi = 32$ samples per stroke and window width $w = 9$. This produces 1,329,920 source windows and 12,896 target windows. All partitions are constructed at the trajectory level rather than the window level, preventing overlapping windows from the same trajectory from appearing in different splits. Appendix 15 reports the complete split and window counts.

Evaluation protocol. We report R^2 on held-out target test trajectories, averaged over five random seeds and eight seeds for the ablation studies. All adaptation baselines use the same architecture, source pretraining, data partitions, and optimization budget, differing only in the adaptation mechanism. For methods with an alignment coefficient, the weight is selected from $\{0.01, 0.1, 1, 10\}$ using the target validation split. The detailed settings are reported in Appendix 16.

Table 2: Zero-shot target-test R^2. Values are mean $\pm$ sample standard deviation over five seeds; "Uncal." omits output calibration.

Method	Venue	R^2	Uncal.
MK-MMD (Long et al., 2015)	ICML'15	0.448 ± 0.031	-1.61
DANN (Ganin et al., 2016)	JMLR'16	0.517 ± 0.114	—
CORAL (Sun & Saenko, 2016)	ECCVW'16	0.624 ± 0.046	-0.07
DARE-GRAM (Nejjar et al., 2023)	CVPR'23	0.719 ± 0.027	-3.46
Importance weighting	—	0.753 ± 0.015	—
RSD (Chen et al., 2021)	ICML'21	$\mathbf{0.768 \pm 0.009}$	-2.83
CLAW, no alignment	Ours	0.761 ± 0.017	-2.79

Table 3: Target-validation alignment-weight sweep. Bold values indicate the selected setting for each adaptation baseline.

Method	.01	.1	1	10
CLAW	0.744; no alignment			
CORAL	0.732	0.688	**0.744**	0.734
MK-MMD	**0.631**	0.519	0.389	0.272
DANN	**0.702**	0.669	0.447	0.516
RSD	0.714	**0.718**	0.718	0.718
DARE-GRAM	**0.745**	0.734	0.733	0.715

Alignment multiplier λ

6 Results

6.1 Cross-Domain Aerodynamic-Law Agreement

Before learning, we verify that the two systems share the same aerodynamic response. We compute $C_N \propto F/|v|^2$, bin by $|\alpha|$, and fit one multiplicative scale over their overlapping range $\alpha \in [26.6°, 88.9°]$. The resulting curves agree with Pearson correlation $\rho = 0.980$ and shape $R^2 = 0.954$ (Fig. 2). The agreement remains strong when the two half-strokes are evaluated separately, with correlations of $0.966/0.953$ and shape R^2 values of $0.923/0.903$. The two domains therefore share the aerodynamic response required for transfer.

6.2 The Alignment-Incompatibility Condition Holds

The domains nevertheless operate over different angle-of-attack and induced-response distributions (Table 1). A two-sample Kolmogorov–Smirnov test on $|C_N|$ gives $D = 0.146$ and $p < 10^{-300}$, establishing $C_{\#}P_{\text{src}} \neq C_{\#}P_{\text{tgt}}$. Corollary 1 therefore applies: exact latent distribution alignment conflicts with preservation of the shared aerodynamic response.

Table 1: Source and target angle-of-attack and induced $|C_N|$ distributions.

Distribution	p_5	Median	p_{95}
Source α	-1.507	-0.408	1.498
Target α	-2.164	0.056	2.153
Source $\lvert C_N \rvert$	0.995	1.930	2.908
Target $\lvert C_N \rvert$	1.241	2.484	3.340

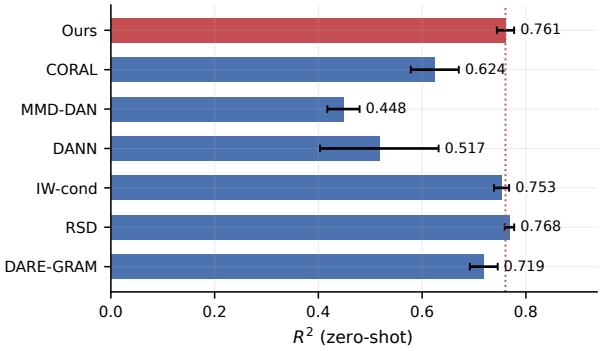


Figure 3: Zero-shot target-test R^2 across adaptation methods. Error bars show variation over five seeds.

6.3 Zero-Shot Transfer

With the source-trained aerodynamic law frozen and only the domain-private output calibration fitted, CLAW reaches $R^2 = 0.761$ (Fig. 3 and Table 2). CLAW substantially outperforms CORAL, MK-MMD, DANN, and DARE-GRAM, matches importance weighting, and is statistically indistinguishable from RSD under the paired five-seed comparison ($p = 0.365$). The result establishes that direct transfer of the aerodynamic response is more effective than feature-distribution alignment or source-sample reweighting. Without fitting $(g_{\text{tgt}}, \eta_{\text{tgt}})$, CLAW obtains $R^2 = -2.79$ because the transferred response remains expressed in source-domain force units. The two-parameter output calibration recovers the target force scale and raises performance to $R^2 = 0.761$, confirming the affine recovery property in Proposition 1.

Table 4: Few-shot target-test R^2. Values are means over five seeds, with sample standard deviations in parentheses. All transfer methods use the CLAW architecture and the same settings.

	Target trajectories b					
Method	4	8	16	32	64	128
From scratch	.708 (.093)	.759 (.016)	.785 (.027)	.797 (.032)	.846 (.002)	.865 (.002)
Head-only	.769 (.016)	.769 (.013)	.769 (.015)	.771 (.015)	.772 (.015)	.771 (.015)
BitFit	.767 (.016)	.767 (.013)	.771 (.014)	.771 (.018)	.779 (.015)	.784 (.014)
LP-FT	.812 (.036)	.858 (.019)	.877 (.016)	.889 (.007)	.910 (.006)	.930 (.005)
C-Mixup	.824 (.061)	.853 (.017)	.883 (.010)	.894 (.006)	.909 (.004)	.925 (.007)
Adapter-only	.825 (.047)	.854 (.028)	.886 (.011)	.891 (.017)	.921 (.005)	.945 (.004)
WiSE-FT	**.833** (.044)	.857 (.025)	.890 (.011)	.895 (.007)	.917 (.007)	.943 (.003)
LoRA	.826 (.048)	.859 (.029)	.876 (.012)	.894 (.018)	.924 (.004)	.953 (.002)
L2-SP	.823 (.063)	**.860** (.026)	**.888** (.010)	**.903** (.008)	**.929** (.008)	.956 (.003)
CLAW, full FT	.821 (.062)	**.860** (.026)	**.888** (.010)	.899 (.008)	.926 (.009)	**.958** (.002)
Gain over scratch	+.113	+.101	+.103	+.102	+.080	+.093
Paired seeds improved	5/5	5/5	5/5	5/5	5/5	5/5

6.4 Effect of Alignment Strength

Table 3 evaluates each alignment objective over four orders of magnitude. The overall pattern supports the incompatibility predicted in Sec. 4.4. MK-MMD degrades consistently from 0.631 to 0.272 as its alignment weight increases. DANN also deteriorates substantially overall, although its strongest weight partially recovers from the result at $\lambda = 1$. DARE-GRAM decreases from 0.745 to 0.715, whereas RSD remains nearly unchanged for $\lambda \geq 0.1$. CORAL is non-monotonic and reaches its best value at an intermediate weight. The competitive methods therefore either favor weak alignment or become largely insensitive to further increases in alignment strength. No alignment setting consistently improves over the unaligned source-trained reference. This result supports the more specific conclusion that aggressive distribution matching is detrimental when the domains induce different response distributions, rather than the broader claim that every form of cross-domain regularization must be harmful.

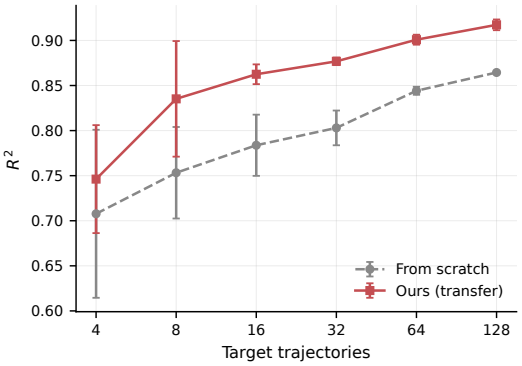


Figure 4: Few-shot target performance of CLAW and matched target-only training across target-data.

6.5 Few-Shot Transfer

Table 4 reports the main few-shot comparison, while Figure 4 highlights the consistent gain over matched target-only training. CLAW with full fine-tuning reaches $R^2 = 0.958$ at $b = 128$, exceeding the from-scratch reference by 0.093. Across all six target-data budgets, the gain over matched from-scratch training ranges from 0.080 to 0.113 and is positive for all five paired seeds. The transfer benefit is therefore consistent across budgets, with its largest absolute advantage occurring in the lowest-data regime.

The variance decreases rapidly with target-data availability. At $b = 4$, CLAW and from-scratch training have seed standard deviations of 0.062 and 0.093, respectively, reflecting sensitivity to the small set of sampled trajectories. At $b = 128$, both standard deviations fall to 0.002, and the method ordering becomes substantially more stable.

The results also distinguish three adaptation behaviors. Head-only tuning and BitFit remain near $R^2 = 0.77$, indicating that target labels cannot be fully exploited while the shared response is held fixed. Methods that adapt the shared response, including LP-FT, C-Mixup, adapter-only tuning, WiSE-FT, LoRA, L2-SP, and full fine-tuning, perform similarly through the intermediate budgets. L2-SP gives the highest rounded mean at several intermediate budgets, whereas CLAW achieves the strongest result at $b = 128$.

Table 5: Architecture ablation. Entries are CLAW's paired R^2 advantage over each variant; parentheses give the number of seeds favoring CLAW. Differences and significance tests are computed from the unrounded per-seed values. $^{*}p < 0.05$ and $^{**}p < 0.01$ under paired tests.

	Target trajectories b					
Ablation	4	8	16	32	64	128
No private adapter	.017 (5)	−.001 (3)	.075* (8)	.032 (5)	.004 (6)	.004 (5)
Shared output scale	.019 (8)	.005* (8)	.018 (5)	.019 (6)	.001 (4)	.000 (6)
Unfactorized MLP	.029 (7)	.059* (7)	.084** (8)	.099** (8)	.008 (6)	.001 (4)
Share adapter, relearn response	.064* (6)	.053* (6)	.121** (8)	.075* (6)	.051** (8)	.043** (8)

6.6 Architectural Ablations

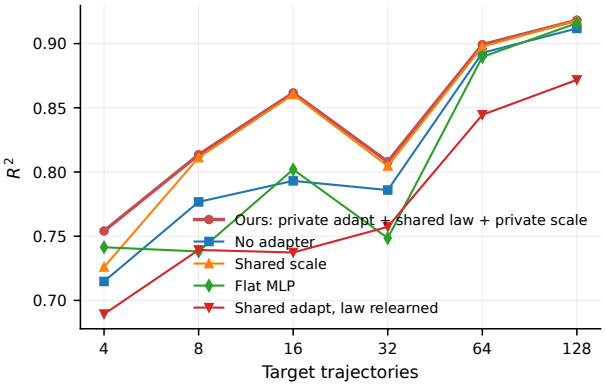


Figure 5: Architecture ablations across target-data budgets.

Figure 5 isolates the contribution of each component in CLAW's physical factorization. The complete architecture combines a domain-private input adapter, a shared nonlinear aerodynamic operator, and domain-private output calibration. The most pronounced degradation occurs when the adapter is shared and the aerodynamic operator is relearned, establishing the shared nonlinear response as the central transferable component.

The unfactorized MLP also underperforms CLAW throughout the intermediate-data regime despite containing more parameters. This result demonstrates that the improvement arises from the physical transfer structure rather than model capacity. Removing the private adapter produces its largest degradation at $b = 16$, while sharing the output scale has a smaller effect under full fine-tuning. Together, the ablations establish that CLAW's principal advantage comes from preserving the shared aerodynamic operator while retaining platform-specific input and output transformations.

6.7 Computational Cost

CLAW introduces no cross-domain alignment term and therefore adds no alignment-specific training overhead. Table 6 shows that CORAL and MK-MMD require $2.7\times$ and $7.0\times$ the source-epoch time, while the SVD-based RSD and DARE-GRAM objectives require $16.4\times$ and $20.5\times$, respectively. CLAW contains 12,195 parameters, requires 5.7 ms per target fine-tuning epoch at $b = 128$, and scores the complete test split in 0.46 ms, approximately 4.2×10^6 windows per second.

7 Conclusion

We introduced CLAW, a physically factorized framework that transfers a shared nonlinear aerodynamic operator while preserving platform-specific input and output calibration. Stroke-phase representation enables transfer across a $52\times$ frequency gap without forcing the source and target distributions to coincide. With 128 target trajectories, CLAW reaches $R^2 = 0.958$, compared with 0.865 for matched target-only training, and improves over target-only learning for every seed at every evaluated budget. Zero-shot transfer with two-parameter output calibration reaches $R^2 = 0.761$. Theory and controlled experiments identify the shared operator as the central transferable component. Replacing shared-law transfer with a shared input adapter reduces R^2 at every budget by up to 0.121 ($p < 0.01$), while preserving signed angle of attack converts negative transfer into a target-domain gain. The alignment sweep further confirms that distribution matching degrades or fails to improve transfer when systems operate over different regions of the shared law.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript organization, language editing, LaTeX formatting, and the presentation and interpretation of experimental results. All AI-assisted text, and technical claims were reviewed and verified by the authors against the underlying data, source code, experiment logs, and cited literature. Generative AI tools were not treated as authors, did not independently conduct experiments, and did not determine the final scientific claims. The authors take full responsibility for the content of the paper and all associated artifacts.

REFERENCES

Leonard Bauersfeld, Elia Kaufmann, Philipp Foehn, Sihao Sun, and Davide Scaramuzza. NeuroBEM: Hybrid Aerodynamic Quadrotor Model. In *Proceedings of Robotics: Science and Systems*, Virtual, July 2021. doi: 10.15607/RSS.2021.XVII.042.

Yagiz Bayiz and Bo Cheng. Flapping wing aerodynamics with PRSSM. Dryad, 2021a. URL https://doi.org/10.5061/dryad.zgmsbccbs. Dataset.

Yagiz E. Bayiz and Bo Cheng. State-space aerodynamic model reveals high force control authority and predictability in flapping flight. *Journal of The Royal Society Interface*, 18(181):20210222, 08 2021b. ISSN 1742-5689. doi: 10.1098/rsif.2021.0222. URL https://doi.org/10.1098/rsif.2021.0222.

Elad Ben Zaken, Yoav Goldberg, and Shauli Ravfogel. BitFit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 1–9, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-short.1. URL https://aclanthology.org/2022.acl-short.1/.

Xinyang Chen, Sinan Wang, Jianmin Wang, and Mingsheng Long. Representation subspace distance for domain adaptation regression. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1749–1759. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/chen21u.html.

Michael H. Dickinson, Fritz-Olaf Lehmann, and Sanjay P. Sane. Wing rotation and the aerodynamic basis of insect flight. *Science*, 284(5422):1954–1960, 1999. doi: 10.1126/science.284.5422.1954. URL https://www.science.org/doi/abs/10.1126/science.284.5422.1954.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016. URL http://jmlr.org/papers/v17/15-239.html.

Somdatta Goswami, Katiana Kontolati, Michael D Shields, and George Em Karniadakis. Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*, pp. 1–10, 2022.

Ananya Kumar, Aditi Raghunathan, Robbie Jones, Tengyu Ma, and Percy Liang. Fine-tuning can distort pretrained features and underperform out-of-distribution, 2022. URL https://arxiv.org/abs/2202.10054.

Yoonho Lee, Annie S. Chen, Fahim Tajwar, Ananya Kumar, Huaxiu Yao, Percy Liang, and Chelsea Finn. Surgical fine-tuning improves adaptation to distribution shifts, 2023. URL https://arxiv.org/abs/2210.11466.

Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 97–105, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/long15.html.

Toshiyuki Nakata, Hao Liu, and Richard J. Bomphrey. A cfd-informed quasi-steady model of flapping-wing aerodynamics. *Journal of Fluid Mechanics*, 783:323–343, 2015. doi: 10.1017/jfm.2015.537.

Ismail Nejjar, Qin Wang, and Olga Fink. Dare-gram : Unsupervised domain adaptation regression by aligning inverse gram matrices. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11744–11754, 2023. doi: 10.1109/CVPR52729.2023.01130.

Michael O'Connell, Guanya Shi, Xichen Shi, Kamyar Azizzadenesheli, Anima Anandkumar, Yisong Yue, and Soon-Jo Chung. Neural-fly enables rapid learning for agile flight in strong winds. *Science Robotics*, 7(66):eabm6597, 2022. doi: 10.1126/scirobotics.abm6597. URL https://www.science.org/doi/abs/10.1126/scirobotics.abm6597.

Sanjay P Sane. The aerodynamics of insect flight. *The Journal of experimental biology*, 206(Pt 23):4191—4208, December 2003a. ISSN 0022-0949. doi: 10.1242/jeb.00663. URL https://doi.org/10.1242/jeb.00663.

Sanjay P. Sane. The aerodynamics of insect flight. *Journal of Experimental Biology*, 206(23):4191–4208, 12 2003b. ISSN 0022-0949. doi: 10.1242/jeb.00663. URL https://doi.org/10.1242/jeb.00663.

Sanjay P Sane and Michael H Dickinson. The aerodynamic effects of wing rotation and a revised quasi-steady model of flapping flight. *The Journal of experimental biology*, 205(Pt 8):1087—1096, April 2002. ISSN 0022-0949. doi: 10.1242/jeb.205.8.1087. URL https://authors.library.caltech.edu/13751/1/SANjeb02.pdf.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In Gang Hua and Hervé Jégou (eds.), *Computer Vision – ECCV 2016 Workshops*, pp. 443–450, Cham, 2016. Springer International Publishing. ISBN 978-3-319-49409-8.

Mitchell Wortsman, Gabriel Ilharco, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo Lopes, Hannaneh Hajishirzi, Ali Farhadi, Hongseok Namkoong, and Ludwig Schmidt. Robust fine-tuning of zero-shot models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7949–7961, 2022. doi: 10.1109/CVPR52688.2022.00780.

8 APPENDIX

9 SYSTEM AND METHOD OVERVIEW

Figure 6 summarizes the complete heterogeneous transfer setting. The source and target systems operate at comparable Reynolds numbers but substantially different reduced frequencies and physical time scales. CLAW maps both systems into stroke phase, preserves domain-private input and output calibration, and transfers the shared nonlinear aerodynamic operator.

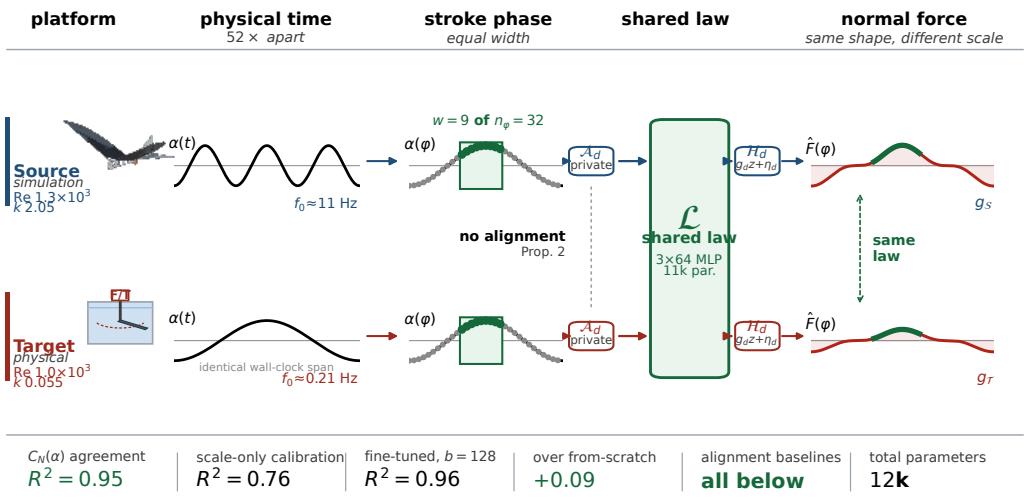


Figure 6: Overview of CLAW across heterogeneous flapping-wing systems. The source simulation and experimental target operate at comparable Reynolds numbers but differ by approximately $52\times$ in flapping frequency and $38\times$ in reduced frequency. Stroke-phase representation places equal fractions of a wing stroke on a common temporal coordinate. Domain-private input maps calibrate platform-specific kinematics, the shared nonlinear operator transfers the aerodynamic response, and domain-private affine output maps recover the force scale and offset. The transferred response achieves shape $R^2 = 0.95$ across domains, zero-shot $R^2 = 0.76$ after output calibration, and $R^2 = 0.96$ after fine-tuning with 128 target trajectories.

Table 6: Training cost on one AMD MI250X GPU. A source epoch processes 1.33×10^6 windows.

Method	Source epoch	Relative cost
CLAW	0.39 s	1.0×
CORAL	1.06 s	2.7×
MK-MMD	2.75 s	7.0×
RSD	6.45 s	16.4×
DARE-GRAM	8.05 s	20.5×

10 SIGNED ANGLE OF ATTACK

Representing the angle of attack by its magnitude reverses the transfer benefit. With $|\alpha|$, source pretraining reaches $R^2 = 0.705$ at $b = 128$, compared with 0.809 from random initialization. With signed α, the corresponding results are 0.916 and 0.860. When sign is removed, the model must reconstruct force direction from stroke phase, but the corresponding phase-to-sign relationship is platform-specific and does not transfer reliably.

11 Detailed Aerodynamic Formulation

11.1 Quasi-Steady Aerodynamic Force

For a rigid flat-plate wing section at angle of attack α relative to an incident flow of magnitude $|v|$, the quasi-steady normal force is

$$F_{\mathrm{qs}} = \tfrac{1}{2}\rho S|v|^2 C_N(\alpha), \qquad C_N(\alpha) = C_L(\alpha)\cos\alpha + C_D(\alpha)\sin\alpha, \tag{14}$$

where ρ is the fluid density and S is the wing area. Following the leading-edge suction analogy and its application to revolving-wing measurements (Dickinson et al., 1999), we use

$$C_L(\alpha) = A_1 \sin 2\alpha, \qquad C_D(\alpha) = C_{D0} + A_2(1-\cos 2\alpha). \tag{15}$$

Substitution gives

$$C_N(\alpha) = A_1 \sin 2\alpha \cos\alpha + [C_{D0} + A_2(1-\cos 2\alpha)]\sin\alpha. \tag{16}$$

For the adopted coefficients, C_N increases monotonically on $[0, \pi/2]$, with $C_N(\alpha) \sim 2A_1\alpha$ as $\alpha \to 0$ and $C_N(\pi/2) = C_{D0} + 2A_2$. Moreover, $C_L(-\alpha) = -C_L(\alpha)$ and $C_D(-\alpha) = C_D(\alpha)$ imply

$$C_N(-\alpha) = -C_N(\alpha). \tag{17}$$

This signed dependence determines the angle representation evaluated in Sec. 10.

Monotonicity is central to cross-domain compatibility. A thin-airfoil model with the hard stall cap $C_L(\alpha) = \min(a_0\alpha, C_{L,\max})$ produces a decreasing normal-force response after the cap activates. The corresponding rejected source configuration is documented in Appendix 13.4.

11.2 Added-Mass Contribution

Fluid acceleration contributes an additional force that depends on the temporal derivatives of relative flow and wing rotation. For a two-dimensional flat plate of chord c and span b (Sane & Dickinson, 2002), the added-mass contribution is

$$F_{\mathrm{am}} = \rho\tfrac{\pi}{4}c^2 b\,(\dot{v}\sin\alpha + v\dot{\alpha}\cos\alpha) - \rho\tfrac{\pi}{16}c^3 b\,\ddot{\alpha}. \tag{18}$$

The complete source force is

$$F = F_{\mathrm{qs}} + \lambda F_{\mathrm{am}}, \tag{19}$$

where λ calibrates the added-mass magnitude as described in Appendix 13.2. The quasi-steady term depends on instantaneous angle and flow speed, whereas the added-mass term depends on temporal derivatives and therefore introduces explicit history dependence.

11.3 Platform Dependence and Transfer Structure

Making the platform index d explicit yields

$$F_d = \tfrac{1}{2}\rho_d S_d|v|^2 C_N(\alpha) + \rho_d c_d^2 b_d \lambda_d \left[\tfrac{\pi}{4}\left(\dot{v}\sin\alpha + v\dot{\alpha}\cos\alpha\right) - \tfrac{\pi}{16}c_d\ddot{\alpha}\right]. \tag{1}$$

Fluid density, wing geometry, and added-mass calibration determine the platform-specific force magnitude. The nonlinear state dependence is expressed through the common physical variables α, $|v|$, $\dot{v}$, $\dot{\alpha}$, and $\ddot{\alpha}$. The rotational added-mass term retains an explicit chord dependence, which is absorbed together with the remaining platform-specific effects through private input and output calibration.

This decomposition directly defines CLAW's architecture. The domain-private input map resolves differences in kinematic range and numerical scale. The shared nonlinear operator captures the transferable aerodynamic response. The domain-private affine output map recovers platform-specific force scale and offset. CLAW therefore shares the aerodynamic relationship while preserving the distinct coordinate and force representations of the two physical systems.

The same decomposition explains the conflict with indiscriminate distribution alignment. The platforms occupy different regions of the law's physical argument space. Matching their latent distributions would consequently erase information about where the shared response is evaluated. The formal incompatibility result appears in Sec. 4.4.

Table 7: Dimensionless operating parameters of the source and target systems. The target chord is inferred from the reported Reynolds number, measured tip speed, and assumed mineral-oil viscosity.

Quantity	Source	Target	Source/target
Medium	Air	Mineral oil	—
ν ($m^2\,s^{-1}$)	1.5×10^{-5}	$\approx 1.15 \times 10^{-4}$	0.13
Chord c (m)	0.034	≈ 0.097	0.35
Frequency f (Hz)	5.2–14.7	0.16–0.27	~ 52
Tip speed U ($m\,s^{-1}$)	0.58	1.19	0.49
$\mathrm{Re} = Uc/\nu$	1.3×10^3	1.0×10^3	1.3
$k = \pi fc/U$	2.05	0.055	**38**

11.4 Dimensionless Operating Regimes

The Reynolds number and reduced frequency are

$$\mathrm{Re} = \frac{Uc}{\nu}, \qquad k = \frac{\pi fc}{U}, \tag{20}$$

where U is the characteristic tip speed, c is chord length, ν is kinematic viscosity, and f is flapping frequency. Table 7 compares the source and target operating regimes.

The Reynolds numbers are closely matched, while the representative reduced frequencies differ by approximately a factor of 38. The source therefore operates in a strongly unsteady regime near $k \approx 2$, whereas the target operates much closer to the quasi-steady regime near $k \approx 0.05$. The transfer task consequently tests whether the shared aerodynamic response generalizes across strongly different temporal operating regimes.

12 CLAW Architecture

Figure 7 presents the complete CLAW architecture. Each platform uses a private input encoder that maps a phase-domain window $W_d \in \mathbb{R}^{9\times4}$ to a 32-dimensional latent representation. The source and target representations then pass through the same three-layer nonlinear aerodynamic operator, conditioned on the cycle-level and phase variables $(c, p) \in \mathbb{R}^8$. A private affine output map converts the shared response into the force scale and offset of each platform. The architecture contains $11{,}009$ shared parameters and $1{,}186$ private parameters per domain.

13 Source-Domain Simulation and Calibration

This appendix documents the source-domain simulator, aerodynamic calibration, kinematic randomization, and simulator-selection criteria. It distinguishes quantities prescribed by the aerodynamic model from quantities determined empirically during simulator calibration.

13.1 Simulation Configuration

The source domain simulates a rigid pair of flat-plate wings at 500 Hz, with all six body degrees of freedom constrained. At each integration step, the relative-flow velocity is evaluated at the aerodynamic reference point of each wing, yielding the local flow magnitude and angle of attack. The quasi-steady and added-mass contributions are then evaluated using Eqs. equation 14 and equation 18. Table 8 summarizes the formal source configuration.

13.2 Added-Mass Calibration

The added-mass model requires estimates of $\dot{v}$, $\dot{\alpha}$, and $\ddot{\alpha}$. Direct finite differences at 500 Hz strongly amplify high-frequency variation, particularly in the second derivative of the angle of attack. Without filtering, the 95th percentile of $|\ddot{\alpha}|$ reaches 1.2×10^5 rad s^{-2}, approximately one order of magnitude above the characteristic scale $\omega|\dot{\alpha}|$.

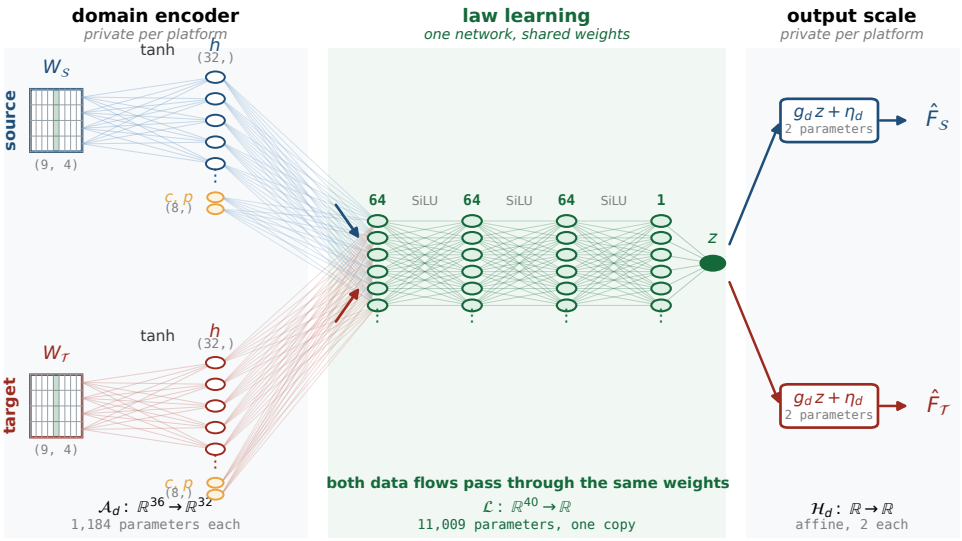


Figure 7: CLAW architecture. Domain-private encoders map source and target phase windows into calibrated latent coordinates. Both data streams then pass through the same three-layer nonlinear aerodynamic operator with shared weights. Domain-private affine output maps recover the force scale and offset of each platform. The model contains 11,009 shared parameters and 1,186 private parameters per domain.

Table 8: Formal source-domain simulation parameters. The added-mass scale, derivative smoothing coefficient, and rotational acceleration clamp are determined by the calibration procedure in Sec. 13.2.

Group	Parameter	Value
Geometry	Wing area S	3.4×10^{-3} m^2
	Chord c	0.034 m
	Span b	0.10 m
	Aerodynamic reference radius	0.02 m
Fluid	Density ρ	1.225 kg m^{-3}
	Kinematic viscosity ν	1.5×10^{-5} m^2 s^{-1}
Coefficients	A_1	1.8
	A_2	1.8
	C_{D0}	0.04
Added mass	Scale λ	0.02
	Derivative smoothing κ	0.85
	Acceleration clamp γ	2.0
Integration	Time step	2 ms
	Episode duration	8 s; first 0.5 s removed

We therefore apply first-order exponential smoothing and constrain the rotational acceleration according to

$$\widetilde{x}_n = \kappa \widetilde{x}_{n-1} + (1-\kappa)x_n, \quad f_{\text{cut}} \approx \frac{f_s(1-\kappa)}{2\pi\kappa}, \quad |\ddot{\alpha}| \le \gamma\omega|\dot{\alpha}|, \quad \omega = 2\pi f_0. \tag{21}$$

For $\kappa = 0.85$, the approximate cutoff frequency is 14 Hz, comparable to the upper end of the simulated flapping-frequency range. The filtered 95th percentile of $|\ddot{\alpha}|$ is 1.9×10^4 rad s^{-2}.

Table 9: Added-mass calibration over three pilot episodes per setting. The reported ratio is $\mathrm{median}(|F_{\mathrm{am}}|/|F_{\mathrm{qs}}|)$.

λ	κ	$\lvert F_{\mathrm{am}}\rvert/\lvert F_{\mathrm{qs}}\rvert$	$p_{95}\lvert\ddot{\alpha}\rvert$ (rad s^{-2})
0.80	0.30	0.694	1.2×10^5
0.08	0.30	0.750	1.2×10^5
1.00	0.85	7.361	1.1×10^4
0.02	0.85	**0.139**	1.9×10^4

Table 10: Per-episode source-domain kinematic randomization.

Parameter	Distribution	Sampling range	Realized p_5–p_{95}
Frequency (Hz)	Log-uniform	$[4, 15]$	4.7–14.5
Stroke amplitude (°)	Uniform	$[32, 72]$	32.8–71.5
Pitch-amplitude scale	Log-uniform	$[2.5, 9]$	2.7–8.3
Pitch center (°)	Uniform	$[-20, 20]$	-18.0–17.9
Pitch phase lead (cycles)	Uniform	$[-0.12, 0.12]$	-0.11–0.11

The added-mass scale is selected using the median magnitude ratio

$$r_{\mathrm{am}} = \mathrm{median}\left(\frac{|F_{\mathrm{am}}|}{|F_{\mathrm{qs}}|}\right). \tag{22}$$

The selected value $\lambda = 0.02$ gives $r_{\mathrm{am}} = 0.139$. This corresponds to the combined-magnitude fraction

$$\frac{r_{\mathrm{am}}}{1 + r_{\mathrm{am}}} = 0.122, \tag{23}$$

which lies within the prespecified range of 10–25%. The unscaled analytic expression causes the rotational added-mass term to dominate the simulated force.

In the selected configuration, $\mathrm{corr}(F_{\mathrm{am}}, \ddot{\alpha}) = -0.937$, confirming that the rotational contribution remains active and that the source response is not purely instantaneous.

13.3 Kinematic Randomization

Each source episode independently samples the five kinematic parameters listed in Table 10. The largest pairwise Spearman correlation among the realized parameters is $|\rho| = 0.19$, and every sampled parameter combination is unique across the 260 episodes. The resulting design varies frequency, stroke amplitude, pitch amplitude, pitch center, and stroke-to-pitch phase without introducing strong pairwise dependence among the controls.

Large phase offsets improve stroke-to-pitch coverage but also increase reversed flow, for which Eq. equation 15 extrapolates beyond its forward-flow coefficient branch. Table 11 quantifies this trade-off.

Restricting the phase lead to ± 0.12 cycles reduces the reversed-flow fraction from 47% in the wide-range pilot to 0.3% in the adopted configuration while narrowing phase-offset coverage from approximately ± 2.6 to ± 0.70 radians. The adopted range therefore favors coefficient-model validity over complete coverage of the target phase-offset distribution.

13.4 Simulator Configuration Selection

Three pilot configurations were excluded before fixing the formal simulator. The comparisons clarify which physical and numerical properties are required for a meaningful cross-system transfer experiment.

Stall-capped thin-airfoil coefficients. For $C_L(\alpha) = \min(a_0\alpha, C_{L,\max})$, with $a_0 = 4.2$ and $C_{L,\max} = 1.15$, the cap activates near 15.7°. Because the median operating angle is approximately

Table 11: Relationship between pitch phase lead and reversed-flow prevalence. The wide-range pilot contains 235 episodes.

Absolute phase lead (cycles)	Episodes	Reversed-flow fraction
$[0.00, 0.10)$	53	0.304
$[0.10, 0.20)$	41	0.404
$[0.20, 0.30)$	60	0.512
$[0.30, 0.45)$	81	0.596
Adopted range, ± 0.12	260	0.003

Table 12: Diagnostic channel associations for `ds_u_raw`. The final model interpretation uses column 0 as signed angle of attack, column 1 as flow-speed magnitude, and columns 5–6 as pitch rate and pitch acceleration. Columns 2–4 are not used by the proposed model.

Column	Range	Diagnostic association	$\lvert r\rvert$	$\lvert r\rvert$ with F_x
0	$[-3.14, 3.14]$	Signed angular coordinate	0.697	**0.712**
1	$[0, 1.17]$	$\sqrt{\dot{\phi}^2 + \dot{\theta}^2}$	0.994	0.003
2	$[-2.96, 3.34]$	$\ddot{\phi}$	0.864	0.100
3	$[-2.37, 2.88]$	Deviation-related component	0.439	—
4	$[-1.36, 0]$	Unused acceleration-related component	0.964	—
5	$[-4.45, 4.41]$	$\dot{\psi}$	**1.000**	0.148
6	$[-12.1, 12.3]$	$\ddot{\psi}$	**0.999**	0.224

71°, the capped response decreases over much of the operating range. The binned cross-domain response correlation is -0.80 under this model and 0.95 under Eq. equation 15.

Unscaled source geometry. A pilot configuration with $S = 0.34\,\mathrm{m}^2$ and $c = 0.34\,\mathrm{m}$ operates near $\mathrm{Re} \approx 10^5$, approximately two orders of magnitude above the target. Its force scale is also approximately two orders of magnitude larger. This configuration was excluded because geometric and Reynolds-number mismatch dominated the intended difference in operating kinematics.

Quasi-steady-only dynamics. Removing Eq. equation 18 makes the source response instantaneous in angle of attack and flow speed. Shuffling the cycle-level condition changes source-domain R^2 only from 0.991 to 0.988, indicating negligible cycle-level dependence. This configuration was excluded because it does not test transfer of history-dependent aerodynamic structure.

14 Target-Dataset Channel Mapping

The public target release stores kinematics in `ds_u_raw` with shape $(548, 470, 7)$. The published specification lists

$$u = [|v|, \mathrm{AoA}, a_x, a_y, a_z, \dot{\alpha}, \ddot{\alpha}]^\top, \tag{24}$$

but direct application of this ordering is inconsistent with the released values. In particular, treating column 1 as angle of attack gives correlation 0.003 with the normal force, rather than the strong angle-force association reported by the original study.

We audited the released channels against quantities reconstructed from `ds_pos`. Table 12 reports the diagnostic associations used to recover the model inputs.

Columns 5 and 6 match pitch rate and pitch acceleration to three decimal places, identifying them as $\dot{\alpha}$ and $\ddot{\alpha}$. Column 1 matches flow-speed magnitude with correlation 0.994. Column 0 spans $[-\pi, \pi]$, has the strongest association with the normal force, and is the only released channel consistent with a signed angular coordinate over the complete cycle. We therefore use column 0 as the signed angle of attack and column 1 as $|v|$.

Preserving the sign of column 0 is essential. As shown in Sec. 10, replacing signed α with $|\alpha|$ reverses the transfer advantage because the model must infer force direction from a platform-specific phase-to-sign relationship.

Table 13: Assembled source and target datasets. Source episodes are per wing, whereas target episodes correspond to individual experimental trajectories.

Quantity	Source	Target
Raw trajectories	260	548
Episodes after cycle detection	590	183
Detected stroke cycles	41,560	403
Phase points per cycle n_φ	32	
Window length w	9	
Windows	1,329,920	12,896
Train/validation/test episodes	413/88/89	128/27/28

15 Data Assembly, Splits, and Normalization

Table 13 summarizes the assembled datasets. The counts satisfy $41{,}560 \times 32 = 1{,}329{,}920$ source windows and $403 \times 32 = 12{,}896$ target windows. The episode partitions also sum to the retained totals: $413 + 88 + 89 = 590$ for the source and $128 + 27 + 28 = 183$ for the target.

Cycle detection retains only complete cycles. Of the 548 target trajectories, 365 yield fewer than two retained cycles and are excluded. The released trajectories contain 470 samples at 25 Hz, corresponding to 18.8 s and approximately three to five wing strokes. The leading and trailing partial strokes are not retained.

All partitions are constructed at the trajectory level. Because adjacent windows within a cycle overlap by $8/9$, a window-level split would produce direct train-test leakage.

15.1 Normalization

Input scales are estimated from the source training partition and applied unchanged to all remaining source and target inputs. Each input channel uses

$$s_j = p_{95}(|x_j|), \tag{25}$$

whereas the response uses the Gaussian-consistent robust scale

$$s_y = \frac{\mathrm{IQR}(y)}{1.349}. \tag{26}$$

No channel is centered, preserving the physical origin of the force model.

Using $p_{95}(|y|)$ instead of the IQR-based response scale places 48.9% of the source samples within $|y| < 0.1$, where the Huber gradient is too small for efficient optimization. The IQR-based scale reduces this fraction to 17.2%.

16 Training Configuration

The shared and domain-private parameter counts sum to the reported model size:

$$11{,}009 + 1{,}186 = 12{,}195. \tag{27}$$

Table 14 reports the complete architecture, optimization, and adaptation settings.

16.1 Alignment-Weight Selection

Each alignment weight is selected by minimizing target-validation MSE, without inspecting target-test performance. The selected settings are $\lambda = 1$ for CORAL, $\lambda = 0.01$ for MK-MMD, $\lambda = 0.01$ for DANN, $\lambda = 0.1$ for RSD, and $\lambda = 0.01$ for DARE-GRAM.

The raw alignment losses differ by several orders of magnitude. CORAL's raw discrepancy is approximately 10^{-6}, whereas the RSD loss is of order one. Each alignment objective is therefore multiplied by a fixed method-specific normalization constant such that its weighted contribution is approximately $O(10^{-2})$ at grid multiplier $\lambda = 1$. The reported grid value is applied on top of this fixed normalization.

Table 14: Training configuration shared across methods except where a baseline requires a method-specific setting.

Group	Setting	Value
Architecture	Adapter width d_h	32
	Shared-operator width/depth	$64/3$
	Activation	SiLU; adapter uses $\tanh$
	Shared parameters	11,009
	Private parameters per domain	1,186
	Total parameters	12,195
Optimization	Optimizer	AdamW
	Learning rate	2×10^{-3}
	Weight decay ω	10^{-4}
	Schedule	Cosine annealing
	Loss	Huber, $\delta = 1$
Schedule	Source epochs	30
	Target epochs	40
	Source/target batch	$4096/\min(4096, \max(256, N))$
	Early-stopping patience	10
Alignment	Standard alignment batch	384
	RSD/DARE-GRAM batch	128
	Weight grid λ	$\{0.01, 0.1, 1, 10\}$
Method-specific	LoRA rank	4
	L2-SP coefficient	10^{-3}

17 Numerical Stability of Subspace Baselines

RSD and DARE-GRAM differentiate through a singular-value decomposition, whose reverse-mode gradient becomes ill-conditioned when squared singular values are close. For $U\Sigma V^\top = A$, the gradient contains terms of the form

$$F_{ij} = \frac{1}{\sigma_j^2 - \sigma_i^2}, \qquad i \neq j. \tag{28}$$

A denominator of order 2×10^{-7} therefore produces a gradient factor of order 5×10^6. RSD introduces an additional instability through $\sin\theta = \sqrt{1 - \cos^2\theta}$, whose derivative

$$-\frac{\cos\theta}{\sqrt{1 - \cos^2\theta}} \tag{29}$$

becomes unbounded as $\cos\theta \to 1$, precisely the regime encouraged by subspace alignment.

For the 64-dimensional, partially saturated features used in this work, the ROCm SVD occasionally failed to converge and propagated non-finite values into subsequent optimization steps. We stabilized the implementation by clamping $\cos\theta$ to $[-1 + 10^{-6}, 1 - 10^{-5}]$, clipping the gradient entering the SVD to unit norm through a forward-identity operator, replacing non-finite feature and gradient values, and limiting the alignment batch to 128. After two consecutive GPU-SVD failures, the decomposition was retried using double-precision CPU arithmetic. These safeguards preserve the intended subspace objective while preventing undefined numerical operations.

With the stabilized implementation, the RSD source-epoch time decreases from 710 s, dominated by repeated CPU fallback, to 6.45 s. The corresponding target result changes from non-finite to $R^2 = 0.768$. All reported RSD and DARE-GRAM results use the stabilized implementations described above.

18 Complete Per-Seed Ablation Results

The main text reports paired mean differences across eight random seeds. Table 15 provides the corresponding absolute per-seed R^2 values for CLAW and the four ablated variants. Each variant

Table 15: Per-seed R^2 for the architecture ablation in Table 5. All variants use full target fine-tuning and differ only in architecture. Mean and s.d. denote the arithmetic mean and sample standard deviation over eight seeds.

Method	Seed	$b = 4$	$b = 8$	$b = 16$	$b = 32$	$b = 64$	$b = 128$
CLAW	0	.782	.812	.875	.879	.897	.917
	1	.809	.867	.855	.740	.909	.924
	2	.704	.817	.866	.862	.894	.912
	3	.845	.859	.862	.869	.899	.920
	4	.630	.713	.850	.690	.897	.918
	5	.830	.760	.874	.879	.895	.910
	6	.691	.670	.854	.880	.895	.909
	7	.729	.735	.849	.873	.885	.908
	Mean	**.752**	**.779**	**.861**	**.834**	**.896**	**.915**
	S.d.	.076	.071	.010	.075	.007	.006
No adapter	0	.743	.818	.839	.746	.896	.917
	1	.744	.841	.847	.858	.903	.914
	2	.752	.775	.766	.770	.894	.913
	3	.776	.724	.829	.848	.884	.912
	4	.560	.726	.685	.707	.888	.904
	5	.844	.768	.873	.884	.903	.911
	6	.812	.797	.759	.761	.897	.917
	7	.656	.794	.684	.840	.873	.898
	Mean	**.736**	**.780**	**.785**	**.802**	**.892**	**.911**
	S.d.	.090	.041	.073	.064	.010	.007
Shared scale	0	.717	.812	.875	.878	.897	.916
	1	.803	.862	.854	.730	.899	.924
	2	.693	.815	.862	.861	.895	.912
	3	.792	.859	.863	.869	.900	.921
	4	.626	.709	.848	.685	.897	.917
	5	.828	.746	.876	.881	.895	.908
	6	.691	.656	.715	.743	.894	.908
	7	.723	.734	.848	.874	.886	.908
	Mean	**.734**	**.774**	**.843**	**.815**	**.895**	**.914**
	S.d.	.068	.074	.053	.081	.004	.006
Unfactorized MLP	0	.771	.763	.754	.787	.881	.915
	1	.789	.652	.822	.647	.879	.902
	2	.737	.763	.768	.760	.898	.921
	3	.804	.815	.852	.868	.898	.924
	4	.606	.698	.814	.681	.892	.917
	5	.720	.772	.778	.761	.899	.916
	6	.685	.641	.734	.745	.892	.913
	7	.676	.655	.692	.634	.872	.906
	Mean	**.724**	**.720**	**.777**	**.735**	**.889**	**.914**
	S.d.	.066	.067	.052	.078	.010	.007
Share adapter	0	.659	.694	.730	.780	.846	.874
	1	.700	.774	.730	.756	.851	.872
	2	.644	.732	.760	.738	.838	.871
	3	.733	.751	.725	.778	.843	.868
	4	.711	.746	.742	.735	.845	.873
	5	.716	.691	.755	.765	.850	.870
	6	.693	.702	.733	.759	.852	.865
	7	.653	.719	.743	.762	.842	.879
	Mean	**.689**	**.726**	**.740**	**.759**	**.846**	**.871**
	S.d.	.033	.030	.013	.016	.005	.004

is independently source-pretrained and subsequently adapted using full target fine-tuning. The table therefore supports paired comparisons within each seed and target-data budget rather than direct comparison with the separately pretrained few-shot baselines in Table 4.