ML LAB RESEARCH PROPOSAL

FlutterForm

An Eigen-Structured Modal-Coupling Transformer for Interpretable, Differentiable Flutter Prediction & Design

Team	Abhinav Garg — solo
Date	July 3, 2026
Project	FlutterForm: eigen-structured modal-coupling transformer for flutter
Target Compute	8× AMD Instinct MI300X (single node), approx. 2 weeks

1. Problem Statement

Flutter — the dynamic aeroelastic instability in which airflow feeds energy into a structure until two vibration modes coalesce and oscillations grow without bound — is the single most safety-critical constraint in aircraft, rotor, and turbomachinery design, and clearing it drives real structural mass and cost. Yet flutter enters the design loop as an opaque, non-differentiable gate: an expensive p–k or CFD–CSD eigen-analysis returns a single flutter speed, with no gradient to *redesign* against it and no reusable account of *which* modes coalesce or *why*. Recent neural surrogates (LSTM / DNN reduced-order models, black-box flutter-speed regressors) speed up the forward evaluation but inherit that opacity — they need large datasets, extrapolate poorly, and still cannot explain the mechanism. I want to test whether hard-wiring the **eigenstructure of flutter itself** into a tiny transformer yields a fully differentiable glass-box that matches black-box accuracy at a fraction of the data and parameters, recovers the coalescence mechanism with no mechanism labels, and — being differentiable end-to-end — turns flutter from a gate into an optimizable design target.

2. Proposed Approach

FlutterForm treats a lifting surface as a short sequence of **structural mode tokens** (the first N natural mode shapes, generalized masses, and frequencies from a cheap assumed-modes / FE model) and is trained under one physically-grounded objective: reconstruct the aeroelastic frequency–damping (V–g / V–f) trajectories and let the flutter point *emerge*. Its core is a primitive I call **eigen-structured coupling attention**: a pairwise outer-product between mode tokens forms the generalized aerodynamic influence-coefficient matrix $Q(k, M_\infty)$ — the bilinear modal coupling that Theodorsen / lifting-surface aerodynamics induce and that standard self-attention cannot represent efficiently — which feeds a **differentiable p–k eigen-solver head** that assembles $[\, p^2 M + K - q_\infty Q(k, M_\infty) \,]\, q = 0$ (generalized mass M, stiffness K, dynamic pressure $q_\infty = \frac{1}{2}\rho V^2$, complex eigenvalue $p = (\gamma + i)\omega$) and marches airspeed V until a mode's damping γ crosses zero, reading off flutter speed V_F, frequency ω_F, and the coalescing mode pair. This generalizes the lab's outer-product block from predicting *coefficients* (FoilForm) to predicting an *operator whose spectrum is the answer* — a reusable pattern for any modal-instability problem (panel & whirl flutter, divergence, thermoacoustics). I chose it over a plain LSTM/DNN surrogate because the eigen-solve gives an interpretable, identifiability-friendly target: the same Q the network learns is the object in a small consistency theorem (under the linear typical-section model the outer-product block recovers the true aerodynamic coupling), so method and theory share one target — the property that lets the mechanism fall out zero-shot. And because the whole map is tiny and differentiable, flutter speed becomes a backprop-able objective, composable with the lab's steady-aero models (FoilCORE / FoilForm) into a differentiable aeroelastic design stack.

3. Dataset

Everything is **generated from public, standard aeroelastic methods** — no proprietary or licensed data, no PHI, no DUA (mirrors the synthetic-identifiability tier of the CLD-Trans proposal). Two generated tiers plus one external, real-world benchmark held out for validation only.

Tier	Physics & generator	Sweep / size
A — Typical section (2–3 DOF)	Pitch–plunge (optional trailing-edge control surface) with Theodorsen unsteady aerodynamics; ground-truth V_F, ω_F, and V–g/V–f trajectories via the standard `p-k` method (`data/generate_pk.py`). Each solve is milliseconds.	mass ratio $\mu \in [5,100]$; frequency ratio $\omega_h/\omega_\alpha \in [0.2,1.0]$; static unbalance $x_\alpha \in [0,0.4]$; elastic-axis $a \in [-0.5,0.2]$; radius of gyration r_α; subsonic Mach w/ Prandtl–Glauert-corrected Theodorsen. **~50k–200k configs**.
B — Simple 3-D wings	Rectangular / swept cantilever plates via assumed-modes (first ~6 bending–torsion modes) + strip-theory or a small doublet-lattice (DLM) unsteady aero; `p-k` ground truth. Harder, closer to real flutter.	planform (aspect ratio, sweep, taper), bending/torsion stiffness ratio, spanwise mass distribution. **~10k–30k wings.**
External — AGARD 445.6	Public NASA weakened-wing experimental flutter data (NTRS reports); the classic transonic double flutter-dip.	Mach 0.499–1.141. **Held-out external validation only — never trained on.**

Preprocessing: nondimensionalize (flutter-speed index $V_F/(b\omega_\alpha\sqrt{\mu})$, reduced frequency $k = \omega b/V$); cache configs + p–k trajectories as memory-mapped `.npy` shards. On-disk footprint < 20 GB total — tiny by CFD/biosignal standards.

4. Evaluation Metrics

Experiment	Metric
Flutter-speed prediction (Tier A/B, **headline**)	relative error $\lvert V_{F,pred}-V_{F,true}\rvert/V_{F,true}$; MAE on flutter-speed index; % within 5% / 10%
Flutter-frequency prediction	relative error on ω_F
Zero-shot coalescing-mode identification (no mechanism labels)	top-1 / top-2 accuracy of predicted coalescing mode pair vs. p–k ground truth
V–g / V–f trajectory reconstruction	RMSE on damping γ(V) and frequency ω(V) curves
Data efficiency vs. capacity-matched black-box LSTM/DNN	error vs. #training configs at 10^2 / 10^3 / 10^4
Extrapolation (out-of-envelope)	error on held-out mass-ratio / sweep ranges outside the training envelope
AGARD 445.6 external (**headline stress test**)	flutter-speed-index & frequency error vs. real wind-tunnel experiment across Mach; capture of the transonic-dip trend
Inverse-design efficacy (differentiability payoff)	flutter-speed gain from gradient-based redesign *through* FlutterForm vs. a p–k-in-the-loop finite-difference optimizer, at matched forward-eval budget
Operator consistency (synthetic)	recovered-Q vs. analytic Theodorsen AIC (Frobenius error), with / without the eigen-solve head
Ablations	no-coupling (diagonal Q) · dense self-attention vs. outer-product · no-eigensolve (regress V_F directly) · assumed-modes count N — 5-seed mean ± 95% CI

Seeds: 42, 123, 7, 0, 256.

5. Compute Requirements

Estimates scaled from a measured single-GPU smoke run (`python train.py mode=tierA train.max_steps=1`), padded 25% for MI300X. The model is **tiny (target < 10k parameters)**; the real workload is dataset generation (CPU-bound p–k solves) plus many small training runs (parameter sweeps, ablations, 5 seeds).

Phase	Hardware	Est. wall-clock
1 · Dataset generation (Tier A+B p–k solves)	CPU / 1 GPU	a few hours
2 · FlutterForm training across the parametric family (bf16, data-parallel sweep across 8 GPUs)	8× MI300X	~0.5–1 day
3 · Black-box baselines (LSTM/DNN) + data-efficiency + extrapolation grids × seeds	1–8× MI300X	~1 day
4 · Ablations + operator-consistency sweeps + AGARD validation	1–8× MI300X	~1 day
Total	**8× MI300X node**	**~3–4 GPU-days inside a 2-week window**

GPU memory: trivial (tiny model, small batches). The only nonstandard op is the differentiable eigen-solve, done via `torch.linalg.eig` on small N×N matrices (N ≤ ~12) with implicit-function-theorem gradients (no backprop through the p–k iteration).
Software / dependencies: ROCm 6.x + official PyTorch ROCm wheel; `pip install -r requirements.txt` (torch, numpy, scipy, matplotlib; DLM in pure numpy). **No CUDA-only kernels** (no flash-attn, no xformers, no custom Triton); eigen-solve has a CPU fallback. No Docker; pure Python venv.

6. Expected Deliverables

- Trained FlutterForm checkpoints (Tier A and Tier B) plus the reproducible p–k data generator (`data/generate_pk.py`).
- **Headline table:** flutter-speed & frequency error and zero-shot coalescing-mode accuracy vs. a capacity-matched LSTM/DNN, with data-efficiency and extrapolation curves.
- Operator-consistency table: recovered aerodynamic coupling Q vs. analytic Theodorsen AIC.
- AGARD 445.6 external-validation figure (flutter index vs. Mach, model vs. experiment) with an honest transonic-fidelity discussion.
- **Differentiable-design demo:** gradient-based redesign of a typical section to raise its flutter speed by backprop *through* FlutterForm, vs. a p–k-in-the-loop baseline — concretely showing flutter as an optimizable constraint (the payoff a forward-only surrogate cannot give).
- A packaged `eigen_coupling_attention` primitive (the reusable attention+eigensolve block).
- A NeurIPS-format draft with plots/tables generated end-to-end by `scripts/make_figures.sh`, per-seed JSON metrics in `results/`, logs in `logs/`.
- MIT-licensed public repo with an MI300X/ROCm README section and a reproducibility statement.

I am intentionally not promising "beats CFD on transonic flutter." The goal is a clean, reproducible test of the interpretable-eigenstructure hypothesis — accurate + tiny + mechanism-recovering + differentiable-for-design on the attached-flow family — with AGARD 445.6 as an external stress test.

7. Risk & Mitigation

Risk	Likelihood	Mitigation
Differentiable eigen-solve is unstable at coalescence (near-defective / repeated eigenvalues) on ROCm	Medium	Use implicit-function-theorem gradients around a standard p–k fixed point (no backprop through the iteration); regularize near defective points; `torch.linalg.eig` CPU fallback; unit-tested against the analytic 2-DOF solution.
Transonic fidelity gap: Theodorsen/DLM cannot produce AGARD's shock-driven double dip	Medium–High	Restrict headline claims to the subsonic/attached-flow family where p–k is exact ground truth; present AGARD strictly as external validation with compressibility-corrected aero. If the dip is not captured, report it honestly as a known limitation motivating a CFD-fidelity follow-up — the empirical paper still stands.
Operator-consistency theorem doesn't close cleanly in the time window	Medium–High	Fallback: empirical operator-recovery (recovered-Q vs. analytic AIC) + zero-shot mode-ID results stand on their own; the theory becomes "supporting" rather than headline (same posture as CLD-Trans).
Black-box LSTM/DNN baseline matches in-distribution accuracy	Medium	Interpretability (mechanism recovery), data-efficiency, and out-of-envelope extrapolation still differentiate FlutterForm even if peak in-distribution error ties — those are the claims the black box structurally cannot make.
Tier-B DLM / strip-theory implementation eats time	Low–Medium	Tier A (typical section) alone supports every core claim; Tier B is additive. Ship Tier A first; DLM behind a flag.
Modal truncation too aggressive (true coalescing pair excluded)	Low	Sweep the assumed-modes count N as an ablation; keep enough modes that the p–k flutter mechanism is converged.