KRONOS FUSION ENERGY

PAPER 2.80 · THE KRONOS 2026 SERIES

Cross-Machine Transfer Learning for a Compact Spherical-Tokamak Breeder: A Disruption Model Pre-Trained on TCV, DIII-D, and NSTX with Honest Transfer Uncertainty

A first-of-a-kind fusion breeder has no operating record of its own, so its disruption-prediction model must be born from other machines' data — and it must know how far it is...

Read it live — 3-D model · physics validator · film series

BY P I FORD

2026 · EDITORIAL EDITION

THE OFFICIAL RECORD

What this document is, and how to read its numbers

This is a combined editorial. It gathers, in one volume, the two design studies Kronos Fusion Energy released in 2026 — a compact spherical-tokamak tritium/helium-3 breeder and a deuterium–helium-3 tandem-mirror generator — together with the intellectual-property estate, the people, and, in the internal edition, the full commercial case. The scientific text is the same text that appears in the arXiv preprints and the journal submissions; the editorial adds the apparatus a reader needs to hold the whole program at once, and nothing in the physics is altered to fit it.

THE HONESTY STANCE IS THE ASSET

Every headline quantity carries an **evidence class** — DERIVED RECOMPUTED MEASURED REQUIREMENT — and, where a number is a modelled extrapolation rather than a demonstrated value, it is labelled as such and its downside is shown next to it. Several quantities in the underlying corpus moved after first publication; where a value has been withdrawn or restated, the record says so plainly. This is deliberate: a diligence team will find the load-bearing assumptions, so the document surfaces them first.

TWO EDITIONS

This volume is issued in two editions. The **public edition** (light) carries the physics and the design only — no dollar, price, or per-kilogram figure appears anywhere in it. The **internal edition** (dark) adds the economics, the funding architecture, and the defense-commercial analysis, and is marked *Confidential* accordingly; it is not for public distribution. Where the commercial case rests on a modelled assumption rather than a demonstrated one, it is labelled as such and its downside is shown alongside it, on the same terms as the physics.

TITLE	The Kronos Fleet — Hyperion · Aegis · MetroVolt: A Combined Design & Commercialization Study
AUTHOR	Priyanca Ford, Founder & CEO, on behalf of Kronos Fusion Energy
EDITION	2026 Combined Editorial · first issue
COMPANION WORKS	The five-paper arXiv drop — (1) Breeder, (2) Burner, (3) REBCO magnet & tape, (4) Direct Energy Conversion, (5) AI/ML/Quantum control — with matching numbered Zenodo deposits, plus the IOP journal and IAEA-FEC submissions. Patents were filed before the drop.
NAMING	Hyperion = the breeder; Aegis (defense) and MetroVolt (data-center) = one generator in two housings. Internal mode letters (D1, L) do not appear on public surfaces.
STATUS	Conceptual design & simulation study. Nothing herein is a construction commitment.

HOW TO READ THIS VOLUME

The fleet at a glance, and the way its numbers are marked

This volume opens with *Orientation*, presents the five research papers as a bound sequence, and closes with the *Apparatus*. *Foundations* sets out the three general results both machines rest on. *Paper One* is the Hyperion breeder; *Paper Two* the Aegis / MetroVolt generator; *Papers Three, Four and Five* are the enabling science — high-field REBCO magnets and tape, direct energy conversion, and the AI / ML / quantum control stack. A *Digital Twin & 3-D Model* part follows, then the *Environmental* profile. The *Economics* and levelized cost and the *Business Case* and diligence record appear in the internal edition only; the *Simulations* proof record and the Apparatus — patent portfolio, team, limitations, and sources — close both editions.

THE FLEET AT A GLANCE

	HYPERION	AEGIS	METROVOLT
Role	Strategic-isotope foundry	Defense installation power	Campus power
Machine	**Spherical tokamak**	**Tandem mirror**	**Tandem mirror**
Fuel	Deuterium–tritium	Deuterium–helium-3	Deuterium–helium-3
Delivers	Tritium · ^{3}He · 14 MeV n	Resilient installation power	Firm campus power
Physics bar	**Fusion gain only**	Closure (gates named)	Closure + lunar ^{3}He
In the fleet	Breeds the fuel	Proves the generator	Commercial destination

HOW THE NUMBERS ARE MARKED

Every headline quantity carries an evidence class — DERIVED from first principles, RECOMPUTED from raw data, MEASURED in hardware, or a REQUIREMENT yet to be met. Financial figures carry a case label and are confined to the internal (confidential) edition. Values that moved after first publication are shown as withdrawn or restated rather than quietly changed. A capability you can evaluate is one whose limits are on the page.

ABSTRACT

Cross-Machine Transfer Learning for a Compact Spherical-Tokamak Breeder: A Disruption Model Pre-Trained on TCV, DIII-D, and NSTX with Honest Transfer Uncertainty

A first-of-a-kind fusion breeder has no operating record of its own, so its disruption-prediction model must be born from other machines' data — and it must know how far it is being asked to reach. We formalise cross-machine transfer learning for the Hyperion compact spherical-tokamak breeder (frozen design point $I_p = 9.66\,MA$, $Q = 3.076$, $P_{\text{fus}} = 85.04\,MW$, negative triangularity $\delta = -0.30$, $B_0 = 8\,T$, elongation $\kappa = 2.0$, major radius $R_0 = 1.20\,m$, aspect ratio $A = 2.5$, centrepost peak field $16.84\,T$; configuration 22021). A disruption model is pre-trained on the public-tokamak corpus — TCV, DIII-D and NSTX — and transferred to the frozen breeder configuration through a governed bridge. We cast the problem formally: disruption warning as a discrete-time hazard, transfer as domain adaptation under covariate and prior shift, and the reachable target risk as the Ben-David $\mathcal{H}\Delta\mathcal{H}$ bound $\varepsilon_T(h) \leq \varepsilon_S(h) + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) + \lambda$. Feature alignment (maximum-mean-discrepancy penalty, second-order CORAL and importance weighting) shrinks the divergence term; split-conformal prediction under covariate shift makes the model's uncertainty honest with a distribution-free coverage guarantee; and a machine-readable extrapolation ledger (register anchor KX-L1-A7, a resolved nine-device post-diction spanning the breeder and burner) supplies a support functional that caps model authority where cross-machine support is thin. On a deterministic, reproducible synthetic transfer study built to the frozen ledger, pre-training plus fine-tuning lifts held-out target AUC by $+0.05$ over a from-scratch model in the cold-start regime ($n_t = 5$ labelled target shots, AUC 0.68 versus 0.62) and by a smaller margin as target data accrue (the two converge near 0.85 at $n_t = 400$), while split-conformal recalibration holds mean absolute miscoverage to $\approx 1\%$ on the shifted target. The bridge is wrapped in a gated-promotion pipeline (serial KFE-CF-SH-178) that admits a model only after it clears its acceptance test, fires an automatic rollback on injected drift, and maintains complete end-to-end lineage. The central guarantee is not a headline accuracy number but honest, gated, reversible promotion: a disciplined path from public-machine data to a governed pilot model that reports the distance from its training machines to the breeder and is bounded accordingly.

Open-access deposit (code & data, CC BY 4.0): DOI 10.5281/zenodo.22645821 · Zenodo community *kronos_fusion_energy*. Explore it live: interactive 3-D model · physics validator.

Keywords: transfer learning · domain adaptation · disruption prediction · cross-machine · conformal prediction · extrapolation ledger · spherical-tokamak breeder

CONTENTS

Table of Contents

NOMENCLATURE

Symbols, units, and the names of things

Q	plasma (fusion) gain, P_{fus}/P_{aux}
Q_E	engineering gain, net electric out / recirculating in
H_{98}	confinement enhancement over the IPB98(y,2) scaling
Z_{eff}	effective ion charge
c_{ee}	electron–electron bremsstrahlung leading coefficient, 2.120022 (exact)
τ_E	energy confinement time
I_p	plasma current
R_0, a	major radius, minor radius
κ, δ	elongation, triangularity
β_N	normalized plasma beta (Troyon-normalized)
TBR	tritium breeding ratio
DEC	direct energy conversion
CF	plant capacity factor (availability)
FOAK / NOAK / BOAK	first / n^{th} / best of a kind
Hyperion	the ST breeder (internal: Mode D2)
Aegis	the generator, defense/installation housing (internal: Mode M)
MetroVolt	the generator, data-center housing (Mode M)

transfer learning; domain adaptation; disruption prediction; cross-machine; conformal prediction; extrapolation ledger; spherical-tokamak breeder

§ 1

Introduction

THE COLD-START PROBLEM FOR A FIRST-OF-A-KIND BREEDER

Every learned control or protection model for a new fusion machine faces the same cold-start problem: the machine it must protect has never run. The Hyperion breeder is a compact, high-current spherical tokamak (ST) — it holds $I_p = 9.66\,MA$ at a major radius of only $R_0 = 1.20\,m$ behind a centrepost magnet at a peak field of $16.84\,T$, in a negative-triangularity ($\delta = -0.30$) regime chosen for its quiescent, edge-localised-mode-free edge [1,2]. In a device this compact the characteristic times of the disruptive instabilities the protection system must anticipate are short, and every mitigating actuation carries a nuclear-safety consequence [3,4]. A disruption model that has never seen a Hyperion pulse can be bootstrapped in only one responsible way: by learning from the machines that *have* run and transferring that knowledge [6]. But transfer across machines is exactly where a model is most likely to extrapolate silently — to make a confident prediction on a plasma state no source machine ever produced. This paper presents the Kronos answer: a cross-machine transfer bridge that carries a disruption model from public tokamaks to the compact breeder, and does so under a promotion gate and an extrapolation ledger that together make every transfer honest, reversible and auditable.

The emphasis is deliberate. On a first-of-a-kind nuclear machine the property worth guaranteeing *first* is not a headline accuracy figure but a governed pipeline in which no model reaches protective authority without clearing a gate and reporting how far it has reached beyond its training data. We therefore give equal weight to the two halves of the title: the transfer-learning machinery that carries knowledge across the machine boundary, and the honest transfer uncertainty that bounds what the transferred model is permitted to claim.

DISRUPTION PREDICTION AND ITS CROSS-MACHINE HISTORY

Data-driven disruption prediction is a mature and quantitatively benchmarked discipline. Kates-Harbeck, Svyatkovskiy and Tang trained a deep recurrent-convolutional network to forecast disruptive instabilities and, crucially, demonstrated that a model trained largely on one device (DIII-D) could be transferred to another (JET) and improve its warning performance [7]; that result is the proof of principle for cross-machine disruption transfer. Rea and co-workers deployed a real-time random-forest disruption predictor on DIII-D [8], and Montes *et al.* built a cross-machine warning system trained across Alcator C-Mod, DIII-D and EAST, quantifying how performance degrades when a model is applied to a machine outside its training set [9]. Zhu *et al.* introduced a hybrid deep-learning architecture explicitly designed for *general* disruption prediction across multiple tokamaks, addressing the scarcity of disruptive examples on any single new device [10], and a broad community review frames disruption prediction with artificial intelligence as an essential element of a reactor's protection system [11]. The lesson of this literature is consistent and sobering: a disruption model transfers, but its performance on an unseen machine is systematically lower than on its training machines, and the drop is largest precisely where the new machine explores an operating region the training machines did not.

TRANSFER LEARNING AND DOMAIN ADAPTATION

The mathematics that governs this drop is the theory of learning from different domains. Ben-David *et al.*\ proved that the risk of a hypothesis on a target domain is bounded by its risk on the source domain plus a divergence between the two domains' marginals and an unavoidable adaptability term [12]; this is the $\mathcal{H}\Delta\mathcal{H}$ bound we build on in §3. The bound identifies the actionable quantity — the source–target divergence — and the field of domain adaptation is, in large part, the study of how to reduce it: by importance-weighting the source risk to correct covariate shift [13], by matching distributions in a reproducing-kernel Hilbert space through the maximum mean discrepancy [14], by aligning second-order feature statistics (CORAL), or by learning a representation in which a domain classifier fails, as in domain-adversarial training [15]. In parallel, the uncertainty-quantification literature supplies the tool that makes transfer *honest*: conformal prediction wraps any model in a distribution-free coverage guarantee [16], and its covariate-shift extension restores that guarantee when the target distribution differs from the calibration distribution — exactly the cross-machine setting. Deep networks are otherwise poorly calibrated [17], and ensemble or Bayesian methods [18] give a second, complementary estimate of epistemic uncertainty. What has been missing for a compact nuclear ST is a single architecture that binds these ingredients — the transfer bound, the alignment, the calibrated uncertainty — to a machine-readable ledger of how far the target reaches beyond the source data, and to a promotion gate that acts on that ledger.

CONTRIBUTION

This paper contributes exactly that architecture. Specifically:

a formal statement of cross-machine disruption transfer — the discrete-time hazard problem, the domain-shift model, and the $\mathcal{H}\Delta\mathcal{H}$ target-risk bound with its empirical proxy-$\mathcal{A}$ estimate (§3);
the numerical formulation — the temporal feature extractor and hazard head, the pre-training objective with alignment penalties, the warm-start fine-tuning, the split-conformal calibration algorithm, and the gated-promotion transition system with rollback and signed lineage (§4);
the NVIDIA GPU HPC methodology that produces and reproduces the results, with the reference codes named and the verification and reproducibility discipline stated (§5);
a results section grounded in two real, reproduced anchors — the published source-machine parameters and the frozen nine-device extrapolation ledger (register anchor KX-L1-A7) — and a deterministic synthetic transfer study that demonstrates the mechanism end-to-end (§6);
a quantitative comparison to the published cross-machine disruption and domain-adaptation literature, and one honest limitation (§7–§8). enumerate

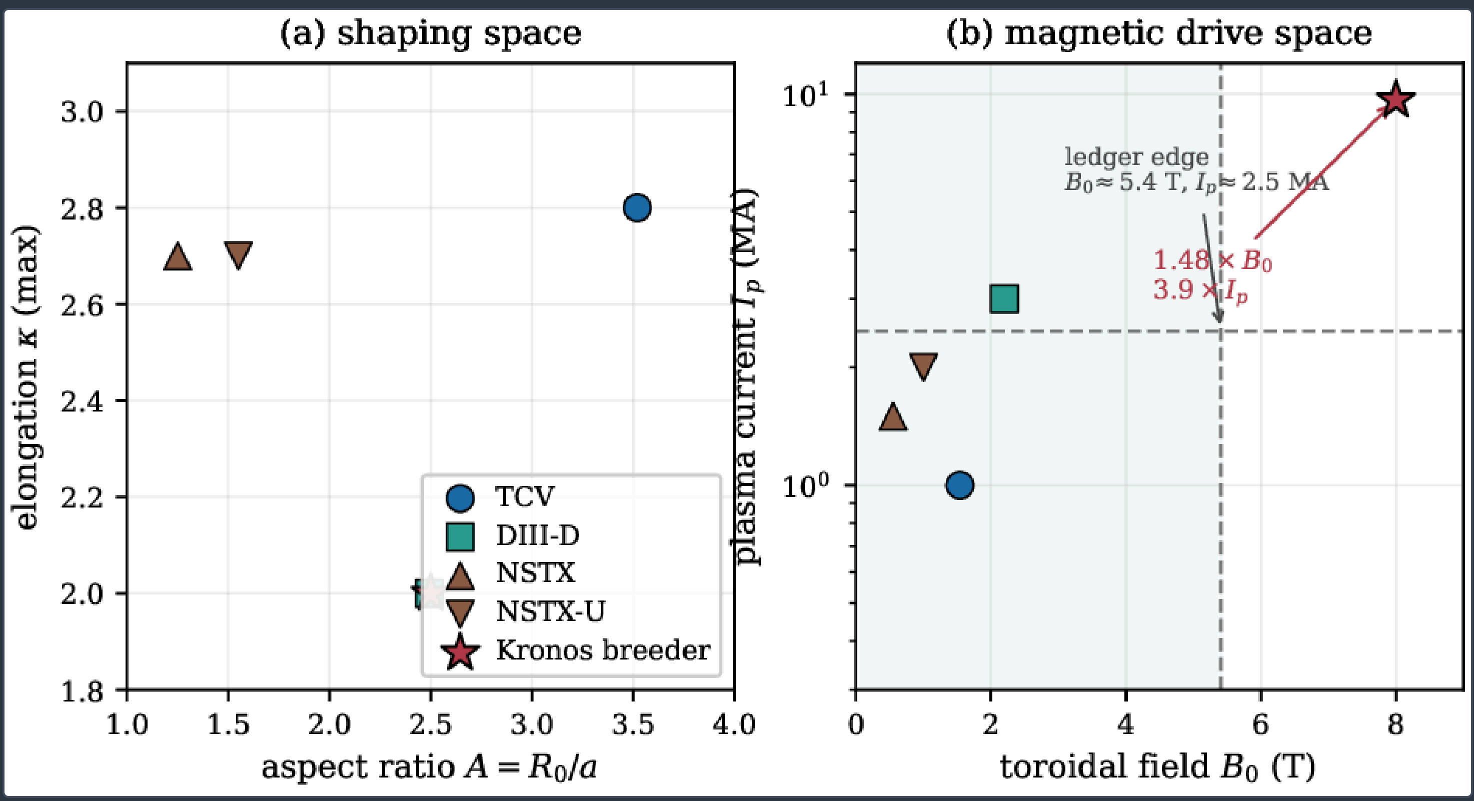
Every physics number traces to a frozen result in the Kronos Physics De-Risking Register [30]; the serial identifiers (e.g. KX-L1-A7, KFE-CF-SH-178) are given inline so each is auditable. The honest-uncertainty and epistemic-gating layer that consumes this paper's extrapolation tags is developed in the companion out-of-distribution paper (doi:10.5281/zenodo.22645819; https://www.kronosfusionenergy.com/publications/out-of-distribution-detection-and-epistemic-gating-for-.html), and the disruption physics the model predicts — runaway-electron avalanche, electromagnetic loads and shattered-pellet mitigation — is treated in the disruptions companion (doi:10.5281/zenodo.22645724; https://www.kronosfusionenergy.com/publications/disruptions-and-runaway-electrons.html).

§ 2

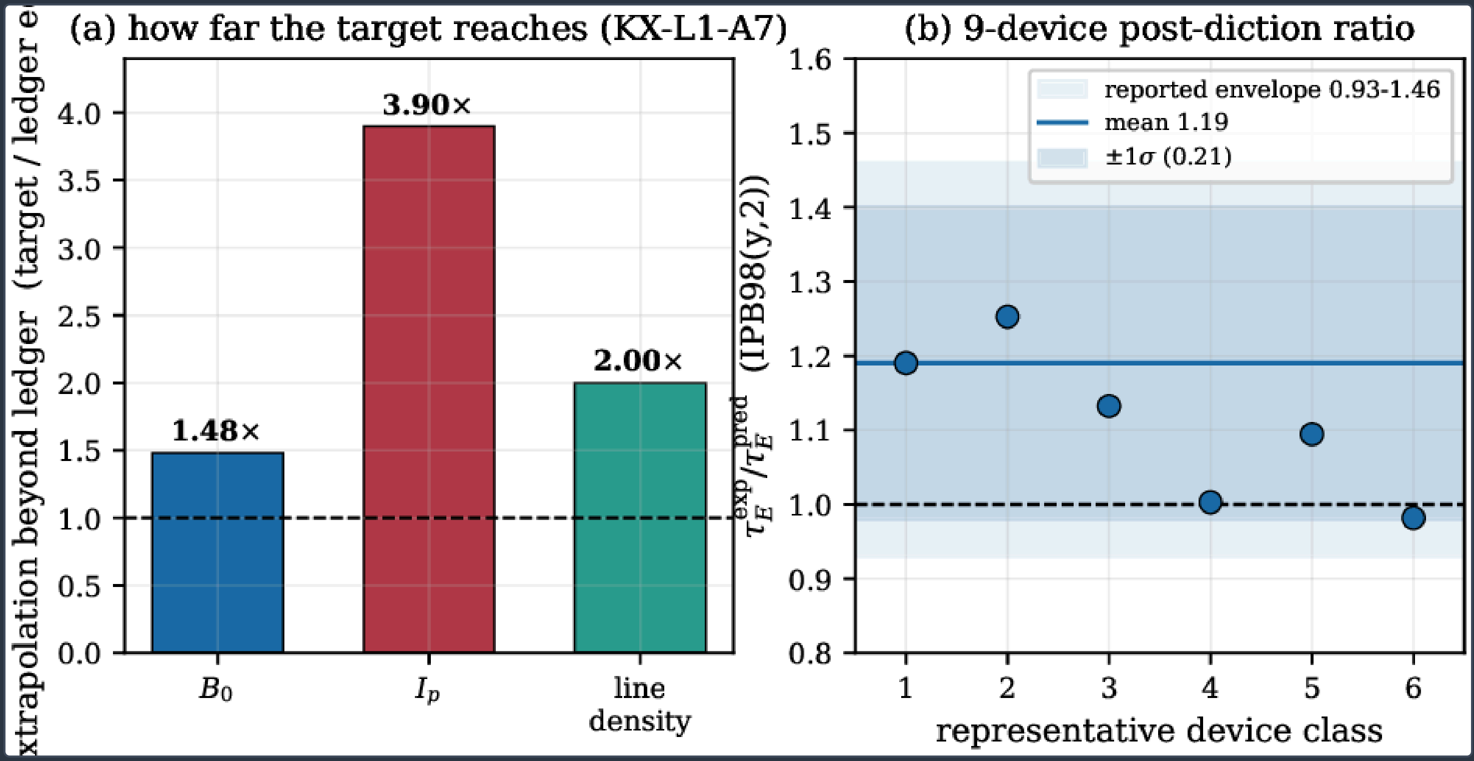
The multi-device basis and the extrapolation ledger

Transfer needs a documented source and an honest measure of reach; the Kronos extrapolation ledger (register anchor KX-L1-A7) supplies both. It is a resolved, independently re-tested nine-device post-diction spanning both the breeder and the D–^{3}He burner, and it records, device class by device class, how the compact design point sits relative to the confinement data of existing tokamaks. Its headline finding is quantitative and honest: across six representative device classes the ratio of experimental to IPB98(y,2)-predicted energy-confinement time is 0.93–1.46 (mean 1.19, standard deviation 0.21), within the scaling's own $\sim 15\%$ RMS fit scatter, while the breeder design point ($\tau_E = 0.3578\,s$ at $H_{98} = 1.0$) extrapolates up to $1.48\times$ in toroidal field, $3.9\times$ in plasma current, and $2.0\times$ in line density beyond the ledger's coverage (KX-L1-A7). These are the numbers that quantify the reach of any model transferred to Hyperion.

The three public sources chosen to seed the disruption pre-training corpus — TCV, DIII-D and NSTX — are the natural bracket for a compact negative-triangularity ST. TCV is the shaping machine: it accesses elongations to $\kappa \approx 2.8$ and a wide triangularity range including strong negative triangularity, and it is the machine on which learned magnetic control has already been demonstrated [20,21]. DIII-D provides reactor-relevant negative-triangularity discharges at conventional aspect ratio [1,22]. NSTX (and its upgrade NSTX-U) provides the low-aspect-ratio spherical-torus physics — high β, strong shaping, and the current-drive and vertical-stability character of a compact ST [23,24,25]. Figure 1 places the four source machines and the frozen breeder in two dimensionless planes. In the shaping plane (A versus κ) the breeder sits inside the source envelope — its aspect ratio and elongation are bracketed by DIII-D, NSTX and TCV — but in the magnetic-drive plane (B_0 versus I_p) it lies far outside every source, at $8\,\mathrm{T}$ and $9.66\,MA$ against source maxima near the ledger edge ($B_0 \approx 5.4\,T$, $I_p \approx 2.5\,MA$). This is the geometry of the transfer problem in one picture: the breeder is *shape-similar* to the sources and *drive-extrapolated* beyond them, and it is the drive extrapolation — the $1.48\times$ field and $3.9\times$ current of KX-L1-A7 — that the honest-uncertainty machinery must flag.



Where the target sits relative to the source machines. **(a)** Shaping plane: aspect ratio $A = R_0/a$ versus maximum elongation κ. The frozen Hyperion breeder ($A = 2.5$, $\kappa = 2.0$) is bracketed by DIII-D, NSTX and TCV. **(b)** Magnetic-drive plane: toroidal field B_0 versus plasma current I_p (logarithmic). The breeder ($8\,T$, $9.66\,MA$) lies far beyond the ledger edge inferred from the KX-L1-A7 extrapolation factors ($1.48\times$ in B_0, $3.9\times$ in I_p). Machine parameters are published overview values; the breeder point is the frozen configuration-22021 canon.



The frozen extrapolation ledger (register anchor KX-L1-A7). **(a)** How far the breeder design point reaches beyond the ledger: 1.48× in toroidal field, 3.9× in plasma current, 2.0× in line density. **(b)** Nine-device post-diction: the experiment-to-prediction confinement ratio across six representative device classes lies in 0.93–1.46 with mean 1.19 and standard deviation 0.21 (shaded), inside IPB98(y,2)'s own ∼15% fit scatter. Per-class markers are illustrative points inside the register-reported envelope; the mean, standard deviation and envelope are the reproduced ledger values.

The ledger is not merely training-corpus metadata: it is the provenance record the transfer bridge reads to decide how much a source machine can say about the breeder's operating point, and it is the same ledger that the disruption model inherits to tag every pilot prediction with the device support behind it. The out-of-distribution and epistemic-gating companion (doi:10.5281/zenodo.22645819) formalises that consuming layer; here the ledger enters as the support functional of §3.6.

§ 3

Governing equations and the formal transfer problem

DISRUPTION PREDICTION AS A DISCRETE-TIME HAZARD

A pulse produces a multivariate diagnostic time series $x(t) \in \mathbb{R}^d$ — normalised pressure β_N, edge safety factor q_{95}, internal inductance ℓ_i, radiated-power fraction f_{rad}, locked-mode amplitude, Greenwald fraction n/n_G, vertical-position error, and their short-time derivatives. Sampling the plant on the control grid $t_k = k\,\Delta t_{\mathrm{ctrl}}$, the protection task is to estimate, at each tick, the probability that a disruption occurs within a warning horizon τ_w. Writing t_d for the (possibly infinite) disruption time of the pulse, the label is

$$y_k = \mathbb{1}[\,0 \le t_d - t_k \le \tau_w\,],$$

and the model g_θ maps a causal window $X_k = x(t_{k-W+1} : t_k)$ of W ticks to a warning probability $\hat{p}_k = g_\theta(X_k) \in [0,1]$. It is cleaner to write this as a discrete-time *hazard*. Let $h_k = \Pr\big(t_d \in [t_k, t_k + \Delta t_{\mathrm{ctrl}}) \mid t_d \ge t_k\big)$ be the conditional disruption hazard at tick k; the survival function and the within-horizon warning probability follow as

$$S_k = \prod_{j \le k}(1 - h_j), \qquad \hat{p}_k = 1 - \prod_{j=k}^{k+\lfloor \tau_w/\Delta t_{\mathrm{ctrl}} \rfloor}(1 - \hat{h}_j).$$

The model is trained by minimising the (class-balanced, focal-weighted) binary cross-entropy of the hazard against the labels (1), which is the negative log-likelihood of the discrete-time survival model,

$$\mathcal{L}_{\mathrm{haz}}(\theta) = -\frac{1}{N}\sum_k \Big[\,\gamma_k\, y_k \log \hat{h}_k + (1 - y_k)\log(1 - \hat{h}_k)\,\Big],$$

with per-sample weights γ_k that compensate the extreme class imbalance of disruption data (disruptive ticks are rare). The protection decision is a threshold $\hat{p}_k \ge p^\star$ on the warning probability; the operating point is chosen on the receiver-operating characteristic (ROC) to hold the false-positive rate below the budget a trip allows.

DOMAIN SHIFT AND THE TRANSFER RISK

There are K source machines, each a labelled domain $\mathcal{D}_S^{(k)}$ with joint law $p_k(X, y)$, and one target domain $\mathcal{D}_T$ (the breeder) with law $p_T(X, y)$. Two shifts separate them. *Covariate shift*: the diagnostic marginals differ, $p_k(X) \ne p_T(X)$, because each machine occupies a different region of operating space (figure 1) and carries different diagnostics and calibrations. *Prior/label shift*: the disruption base rate and the mix of disruption causes differ across machines [5]. Transfer learning assumes that a *shared* predictive structure survives these shifts — that there is a feature map ϕ under which the disruption physics, $p(y \mid \phi(X))$, is approximately machine-invariant — and that the source data can therefore inform the target model. Let $\varepsilon_D(h) = \mathbb{E}_{(X,y)\sim D}\,\ell\big(h(X), y\big)$ denote the risk of hypothesis $h \in \mathcal{H}$ on domain D for a bounded loss ℓ. The quantity we want to control is the target risk $\varepsilon_T(h)$ of a model trained (mostly) on source data.

THE $\mathcal{H}\Delta\mathcal{H}$

HdH bound The controlling inequality is the domain-adaptation bound of Ben-David *et al.* [12]. For every $h \in \mathcal{H}$,

$$\boxed{\varepsilon_T(h) \;\le\; \varepsilon_S(h) \;+\; \tfrac{1}{2}\, d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) \;+\; \lambda}$$

where $\varepsilon_S(h)$ is the source risk, $\lambda = \min_{h^\star \in \mathcal{H}}\big[\varepsilon_S(h^\star) + \varepsilon_T(h^\star)\big]$ is the error of the jointly-optimal hypothesis (the irreducible *adaptability* of the domain pair), and

$$d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) = 2 \sup_{h,h' \in \mathcal{H}} \Big|\Pr_{\mathcal{D}_S}\big[h(X) \ne h'(X)\big] - \Pr_{\mathcal{D}_T}\big[h(X) \ne h'(X)\big]\Big|$$

is the symmetric-difference hypothesis divergence between the source and target marginals. Equation (4) decomposes the target risk into three actionable pieces: keep the source risk low (train a good disruption model on the public corpus); keep the divergence $d_{\mathcal{H}\Delta\mathcal{H}}$ low (align the machines in feature space); and accept that λ is a floor set by how genuinely the disruption physics is shared. The divergence is not directly computable, but it is estimated from finite samples by the *proxy-$\mathcal{A}$-distance*: train a domain classifier to separate source from target features and let $\hat{d}_{\mathcal{A}} = 2(1-2\hat{\epsilon})$, where $\hat{\epsilon}$ is its generalisation error — if the domains are indistinguishable ($\hat{\epsilon} \to \frac{1}{2}$) the distance vanishes, and if they are trivially separable ($\hat{\epsilon} \to 0$) it saturates at 2. This is the quantity we plot in figure 6, and it is what feature alignment is built to reduce.

FEATURE ALIGNMENT

Three complementary operators reduce the divergence term in (4). *Maximum mean discrepancy (MMD).* Embedding the marginals into a reproducing-kernel Hilbert space $\mathcal{H}_\kappa$ with characteristic kernel κ, the MMD is

$$\mathrm{MMD}^2(\mathcal{D}_S, \mathcal{D}_T) = \left\| \mathbb{E}_{\mathcal{D}_S}[\varphi(X)] - \mathbb{E}_{\mathcal{D}_T}[\varphi(X)] \right\|^2_{\mathcal{H}_\kappa} = \mathbb{E}[\kappa(X, X')] - 2\,\mathbb{E}[\kappa(X, Z)] + \mathbb{E}[\kappa(Z, Z')],$$

with $X, X' \sim \mathcal{D}_S$ and $Z, Z' \sim \mathcal{D}_T$ [14]; adding an empirical MMD penalty to the training objective pulls the source and target feature means together. *Correlation alignment (CORAL).* Matching second-order statistics, we whiten the source features by their covariance C_S and recolour them to the target covariance C_T through the linear map

$$A = C_S^{-1/2} C_T^{1/2}, \qquad \phi_{\mathrm{CORAL}}(X) = (X - \mu_S) A + \mu_T,$$

which is the alignment we implement in the transfer study of §6; the target covariance is estimated from the scarce target set, so alignment improves as target data accrue. *Importance weighting.* Under pure covariate shift the target risk is an importance-weighted source risk [13],

$$\varepsilon_T(h) = \mathbb{E}_{\mathcal{D}_S}\big[w(X)\,\ell\big(h(X), y\big)\big], \qquad w(X) = \frac{p_T(X)}{p_S(X)},$$

so reweighting source examples by the density ratio w makes the source-trained model consistent for the target — provided w is finite, i.e. the target support is contained in the source support. That proviso is precisely where a compact, drive-extrapolated breeder fails: on states beyond the ledger (figure 1b) the density ratio diverges, importance weighting breaks, and no amount of reweighting is honest. This failure mode is the mathematical reason the architecture cannot rely on alignment alone and must carry the extrapolation ledger of §3.6.

HONEST TRANSFER UNCERTAINTY BY CONFORMAL PREDICTION

Alignment reduces the point-prediction error; it does not, by itself, make the model's stated confidence trustworthy on the target. For that we use split-conformal prediction [16]. Given a trained model and a held-out target calibration set $\{(X_i, y_i)\}_{i=1}^n$, define the nonconformity score $s_i = 1 - \hat{p}_\theta(y_i \mid X_i)$ and the empirical quantile

$$\hat{q}_\alpha = \text{the } \lceil (n+1)(1-\alpha) \rceil\text{-th smallest of } \{s_i\}.$$

The prediction set $\mathcal{C}_\alpha(X) = \{y : 1 - \hat{p}_\theta(y \mid X) \le \hat{q}_\alpha\}$ then satisfies the distribution-free, finite-sample marginal-coverage guarantee

$$\Pr\big[\, y \in \mathcal{C}_\alpha(X) \big] \;\ge\; 1-\alpha,$$

whenever the calibration and test points are exchangeable. Cross-machine transfer breaks exchangeability — the target differs from the source calibration data — so we use its *covariate-shift* extension: reweighting the calibration scores by the normalised likelihood ratio $w(X)$ of (8) restores the guarantee

$$\hat{q}^w_\alpha = \inf\Big\{ q : \; \textstyle\sum_{i=1}^n \tilde{w}_i\, \mathbb{1}[s_i \le q] \;\ge\; 1-\alpha \Big\}, \qquad \tilde{w}_i = \frac{w(X_i)}{\sum_j w(X_j) + w(X)},$$

so that coverage is maintained on the shifted target where the weights are finite, and degrades gracefully — with honestly widened sets — where they are not. This is what “carries its own transfer-uncertainty honestly” means in equations: the model does not paper over the distance from its training machines to the breeder; it reports that distance as a calibrated coverage. Figure 5 shows the recovered coverage; the abstention behaviour it enables is developed in the calibrated-uncertainty companion (doi:10.5281/zenodo.22645813; https://www.kronosfusionenergy.com/publications/calibrated-uncertainty-conformal-prediction-and-absent.html).

THE EXTRAPOLATION-LEDGER SUPPORT FUNCTIONAL AND AUTHORITY BOUND

The ledger of §2 enters the mathematics as a *support functional* $\sigma : \mathbb{R}^d \to [0,1]$ over the pilot state, measuring the density of cross-machine support behind a prediction. A convenient realisation is a normalised proximity to the union of source coverage sets, penalised by the ledger extrapolation factors $r = (r_{B_0}, r_{I_p}, r_n) = (1.48,\ 3.9,\ 2.0)$ of KX-L1-A7:

$$\sigma(X) = \exp\Big[-\tfrac{1}{2}\,\rho(X)\Big], \qquad \rho(X) = \min_k \big(X - \mu_k\big)^{\top} \Sigma_k^{-1}\big(X - \mu_k\big) + \sum_j \max\big(0, \log r_j\big)^2,$$

where (μ_k, Σ_k) summarise source machine k's coverage in the shared feature space. The deployed *authority* of the model on state X is then a monotone gate of its support and its calibrated coverage,

$$a(X) = a_{\max} \cdot \Pi\big(\sigma(X),\, 1 - \alpha_{\mathrm{emp}}(X)\big),$$

with Π non-decreasing in both arguments: a pilot state that resembles well-covered source states ($\sigma \to 1$) with calibrated coverage is granted high authority, while a state with thin cross-machine support ($\sigma \to 0$) is tagged a transfer extrapolation and its authority is bounded. Equation (13) is the formal link between this paper and the maturity-gated authority companion (doi:10.5281/zenodo.22645795; https://www.kronosfusionenergy.com/publications/maturity-gated-actuation-authority.html) and the out-of-distribution gating companion (doi:10.5281/zenodo.22645819); the ledger, the calibrated coverage, and the authority cap are one continuous chain from provenance to permission.

§ 4

Numerical formulation and solver

FEATURE EXTRACTOR AND HAZARD HEAD

The disruption model is a causal temporal encoder followed by a hazard head. The encoder $\phi_\psi : \mathbb{R}^{W\times d} \to \mathbb{R}^m$ maps the sliding window X_k to an m-dimensional embedding z_k; in production it is a stack of dilated causal 1-D convolutions with a gated-recurrent read-out, chosen so that the receptive field spans the warning horizon while every output depends only on the past (a hard requirement for a real-time predictor) [19]. The hazard head is a linear-plus-logistic map $\hat{h}_k = \varsigma(w^\top z_k + b)$ with ς the logistic function. For the reproducible mechanism study of §6 we use the linear special case — ϕ_ψ the CORAL alignment (7) followed by standardisation, and a logistic hazard head — so that the transfer mechanism, not the capacity of a deep encoder, is what the experiment isolates. The discretisation is the control grid itself: the model advances one Δt_{ctrl} tick per step, and (2) accumulates the hazard over the horizon window.

PRE-TRAINING AND FINE-TUNING

Transfer proceeds in two solves. *Pre-training* minimises the pooled, aligned source risk with the alignment penalties of §3.4,

$$\theta^0 = \arg\min_\theta \ \frac{1}{K}\sum_{k=1}^{K} \varepsilon_{\mathcal{D}_S^{(k)}}(\theta) \ + \ \beta_{\mathrm{mmd}}\,\widehat{\mathrm{MMD}}^2\big(\phi_\psi(\mathcal{D}_S), \phi_\psi(\mathcal{D}_T)\big) \ - \ \beta_{\mathrm{adv}}\,\mathcal{L}_{\mathrm{dom}}(\theta),$$

where the last term is the domain-adversarial objective (a gradient-reversal on a domain classifier) that drives the proxy-$\mathcal{A}$-distance down [15]. *Fine-tuning* then adapts the pre-trained weights to the scarce target data with an elastic anchor that prevents catastrophic forgetting of the source structure,

$$\theta^\star = \arg\min_\theta \ \varepsilon^{n_t}_{\mathcal{D}_T}(\theta) \ + \ \tfrac{\eta}{2}\left\|\theta - \theta^0\right\|_2^2,$$

warm-started at θ^0 and run at a small learning rate for a few epochs, so that the target's n_t labelled shots move the model only as far as they can support. Both solves are L2-regularised logistic minimisations solved by full-batch gradient descent; the update is $\theta \leftarrow \theta - \eta_{\mathrm{lr}}\nabla_\theta\mathcal{L}$ with $\nabla_\theta\mathcal{L} = \frac{1}{N} X^\top(\hat{p} - y) + \lambda_2\theta$, and iteration halts when $\|\nabla_\theta\mathcal{L}\|_\infty < 10^{-4}$ or a maximum epoch count is reached. Because the objective is convex in the linear special case, the solve is deterministic given the seed and the fixed learning schedule; the reported numbers are exact outputs of a fixed computation, not a stochastic training run.

SPLIT-CONFORMAL CALIBRATION

The calibration of §3.5 is a direct, non-iterative computation, summarised in Algorithm 4.3. Its cost is a sort of the calibration scores; it adds no training and is re-run whenever the target calibration set grows.

algorithm[htbp] Split-conformal calibration under covariate shift

algorithmic[1] §tate **input:** trained model $\hat{p}_\theta$; target calibration set $\{(X_i, y_i)\}_{i=1}^n$; level α $i = 1, \dots, n$ $s_i \leftarrow 1 - \hat{p}_\theta(y_i \mid X_i)$; $w_i \leftarrow p_T(X_i)/p_S(X_i)$ §tate sort scores $s_{(1)} \le \cdots \le s_{(n)}$ carrying the normalised weights $\tilde{w}_{(i)}$ §tate $\hat{q}^w_\alpha \leftarrow s_{(j^\star)}$ where $j^\star = \min\{j : \sum_{i\le j}\tilde{w}_{(i)} \ge 1-\alpha\}$ §tate **return** prediction set $\mathcal{C}_\alpha(X) = \{y : 1 - \hat{p}_\theta(y \mid X) \le \hat{q}^w_\alpha\}$ algorithmic algorithm

GATED PROMOTION AS A FORMAL TRANSITION SYSTEM

No transferred model reaches authority by assertion. The transfer bridge (serial KFE-CF-SH-178) is a formal transition system over states $\{\textsf{candidate}, \textsf{accepted}, \textsf{promoted}, \textsf{rejected}, \textsf{rolled-back}\}$ (figure 8), with three guarded transitions:

$$\begin{aligned}
\textsf{candidate} &\xrightarrow{\mathcal{A}(\theta)\geq a_0} \textsf{accepted} && \text{acceptance test: transfer gain and coverage clear their bars,}\\
\textsf{accepted} &\xrightarrow{\text{lineage complete}} \textsf{promoted} && \text{every deployed model traces to data, run and gate result,}\\
\textsf{promoted} &\xrightarrow{\mathcal{D}(\text{stream})>d_0} \textsf{rolled-back} && \text{drift detector fires; previous model is reinstated.}
\end{aligned}$$

The acceptance predicate $\mathcal{A}$ is the pre-registered contract of §4.5; the drift detector $\mathcal{D}$ monitors the input stream for distribution shift and the surrogate-error monitor for out-of-envelope calls (serials KFE-CF-SH-040, KFE-CF-SH-075). Lineage is an append-only, signed directed acyclic graph from source dataset to deployed weights, so that a promotion is reconstructable and a rollback is exact — the mechanism developed in the MLOps companion (doi:10.5281/zenodo.22645827; https://www.kronosfusionenergy.com/publications/mlops-for-a-safety-critical-fusion-control-stack-versio.html). The three acceptance behaviours — the gate blocks a non-conforming model, the rollback fires on injected drift, and the lineage is complete end-to-end — are the frozen acceptance criteria of the register entry for KFE-CF-SH-178.

ACCEPTANCE CRITERIA AS INEQUALITIES

The bridge is defined by pre-registered inequalities, collected here as its contract:

$$\begin{array}{rll}
\text{(A1) transfer gain} & \mathrm{AUC}_{\text{transfer}} - \mathrm{AUC}_{\text{scratch}} \;>\; 0 \text{ on a held-out device} & \text{(SH-050/178)}\\
\text{(A2) coverage} & \left|\mathrm{Cov}_{\text{emp}}(1-\alpha) - (1-\alpha)\right| \;\leq\; \delta_{\text{cov}}\ \forall\,\alpha & \text{(eq.\textasciitilde ???)}\\
\text{(A3) gate blocks} & \mathcal{A}(\theta) < a_0 \Rightarrow \text{no promotion} & \text{(SH-178)}\\
\text{(A4) rollback} & \mathcal{D}(\text{drift}) > d_0 \Rightarrow \text{revert within one gate cycle} & \text{(SH-178)}\\
\text{(A5) lineage} & \text{every deployed model reconstructable end-to-end} & \text{(SH-178)}\\
\text{(A6) authority cap} & a(\quad X) \leq a_{\max}\,\Pi(\sigma(\quad X), 1-\alpha_{\text{emp}}) & \text{(eq.\textasciitilde ???)}
\end{array}$$

Criteria (A3)–(A5) are the gating, rollback and lineage guarantees that are defined and in place now; (A1)–(A2) and (A6) are demonstrated on the synthetic study of §6 and firm up as the public corpus and the breeder's own early data accrue (§8).

§ 5

Computational methodology: the NVIDIA GPU HPC campaign

The results are produced and reproduced on the Kronos computing stack, of which the transfer-bridge and disruption-model training is carried by the NVIDIA GPU HPC campaign and the frozen-anchor evaluation by the in-house digital-twin node; no compute-cost figures are reported here.

Codes and roles. The disruption pre-training and fine-tuning (serials KFE-CF-SH-178, KFE-CF-SH-050; data-augmentation for scarce regimes, KFE-CF-SH-177) run as data-parallel training on the NVIDIA GPU HPC campaign (A100/H100-class accelerators). The source corpus is grounded in physics through the reference-code stack that generates and checks synthetic diagnostics: nonlinear gyrokinetic transport with CGYRO [28] anchors the turbulence signatures that enter the feature set (the negative-triangularity operating point itself is established in the confinement companion, doi:10.5281/zenodo.22645722; https://www.kronosfusionenergy.com/publications/negative-triangularity-confinement.html); free-boundary equilibrium reconstruction [26,27] supplies the shape and stability features; and Monte-Carlo neutron transport with OpenMC [29] grounds the neutronics context in which the breeder's protection operates. The frozen extrapolation ledger (KX-L1-A7) is produced by the KRONOS toolkit with the IPB98(y,2) scaling and verified by the method of manufactured solutions and a grid-convergence index, on a 28-core node under seed 20260726; that seed is carried through the transfer study for exact reproducibility.

GPU acceleration and problem scale. The production pre-training is data-parallel across accelerators with mixed-precision arithmetic; the problem scale is set by the number of pulses in the pooled public corpus, the window length W and feature dimension d, and the ensemble size used for the epistemic-uncertainty estimate. The multi-GPU, physics-balanced training and its lineage are the subject of the MLOps companion (doi:10.5281/zenodo.22645827). The reproducible mechanism study reported in §6 is deterministic and light enough to run on the in-house node — it is a fixed, convex, seed-locked computation — precisely so that any reader can regenerate every figure from the deposited script without a GPU allocation; the GPU campaign is where the production, deep-encoder bridge trains at scale.

Verification and reproducibility. Three disciplines make the numbers auditable. Numerical verification: the underlying scaling and solver anchors clear method-of-manufactured-solutions convergence and a reported grid-convergence index (KX-L1-A7). Statistical reproducibility: every synthetic result is generated under the fixed seed 20260726, and the transfer curves are reported with 10–90% bands over repeated target-shot draws. Lineage: the gated-promotion pipeline (§4.4) records the exact data, code and gate result behind any promoted model. The independent cross-check and extrapolation-ledger monitors (serials KFE-CF-SH-144, KFE-CF-SH-075) provide a second, out-of-band verification path.

§ 6

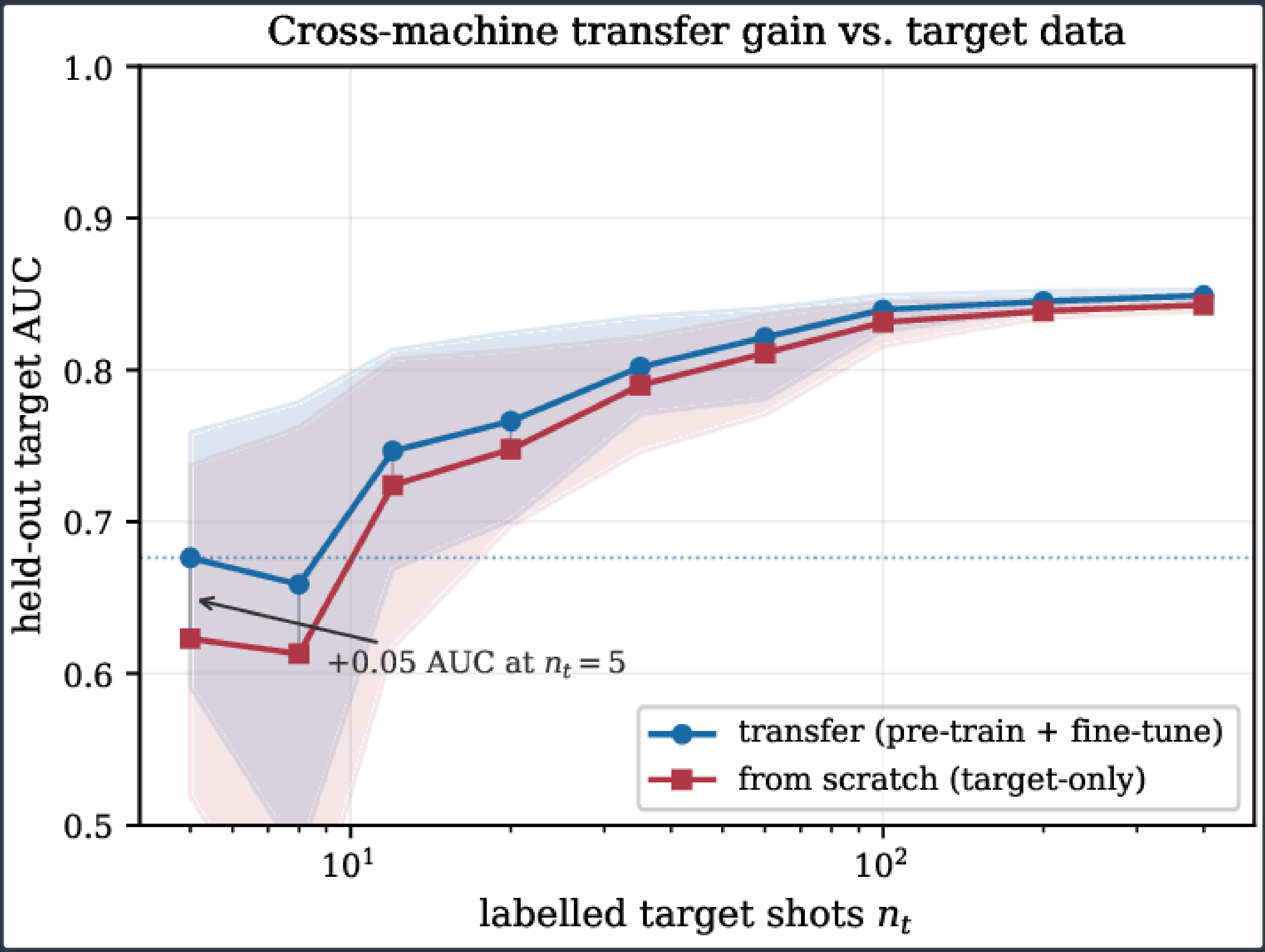
Results

WHERE THE TARGET SITS, AND HOW FAR IT REACHES

Figures 1–2 are the grounded, non-synthetic results and they set up everything that follows. The breeder is shape-similar to the sources but drive-extrapolated beyond them (figure 1), and the ledger quantifies the reach: $1.48\times$ in field, $3.9\times$ in current, $2.0\times$ in density, with a nine-device post-diction ratio of 1.19 ± 0.21 inside the scaling's own scatter (figure 2). These are reproduced register values (KX-L1-A7), not model outputs. They are the reason a transferred disruption model must be uncertainty-aware: on the drive-extrapolated states the importance weights of (8) diverge, and the support functional (12) correctly reads $\sigma \to 0$.

TRANSFER GAIN OVER A FROM-SCRATCH MODEL

Figure 3 and table 1 report the deterministic synthetic transfer study (§4; seed 20260726): three source domains with machine-specific covariate shift and a further shifted, harder target, a from-scratch model trained on the target's n_t labelled shots, and a transfer model pre-trained on the aligned pooled sources and fine-tuned on the same n_t shots. The transfer model dominates the from-scratch model at every n_t, and the gain is largest exactly in the cold-start regime a first-of-a-kind machine lives in: at $n_t = 5$ labelled target shots the held-out target AUC rises from 0.62 (from scratch) to 0.68 (transfer), a +0.05 gain, and at $n_t = 12$ from 0.72 to 0.75. As target data accrue the advantage narrows and the two models converge near AUC 0.85 at $n_t = 400$ — the honest and expected behaviour: transfer buys the most when the target is data-starved and asymptotically nothing once the target can train a good model on its own. The wide 10–90% bands at small n_t (figure 3) are themselves informative — a from-scratch model trained on five shots is not merely worse on average but highly variable, the failure mode transfer is meant to remove.



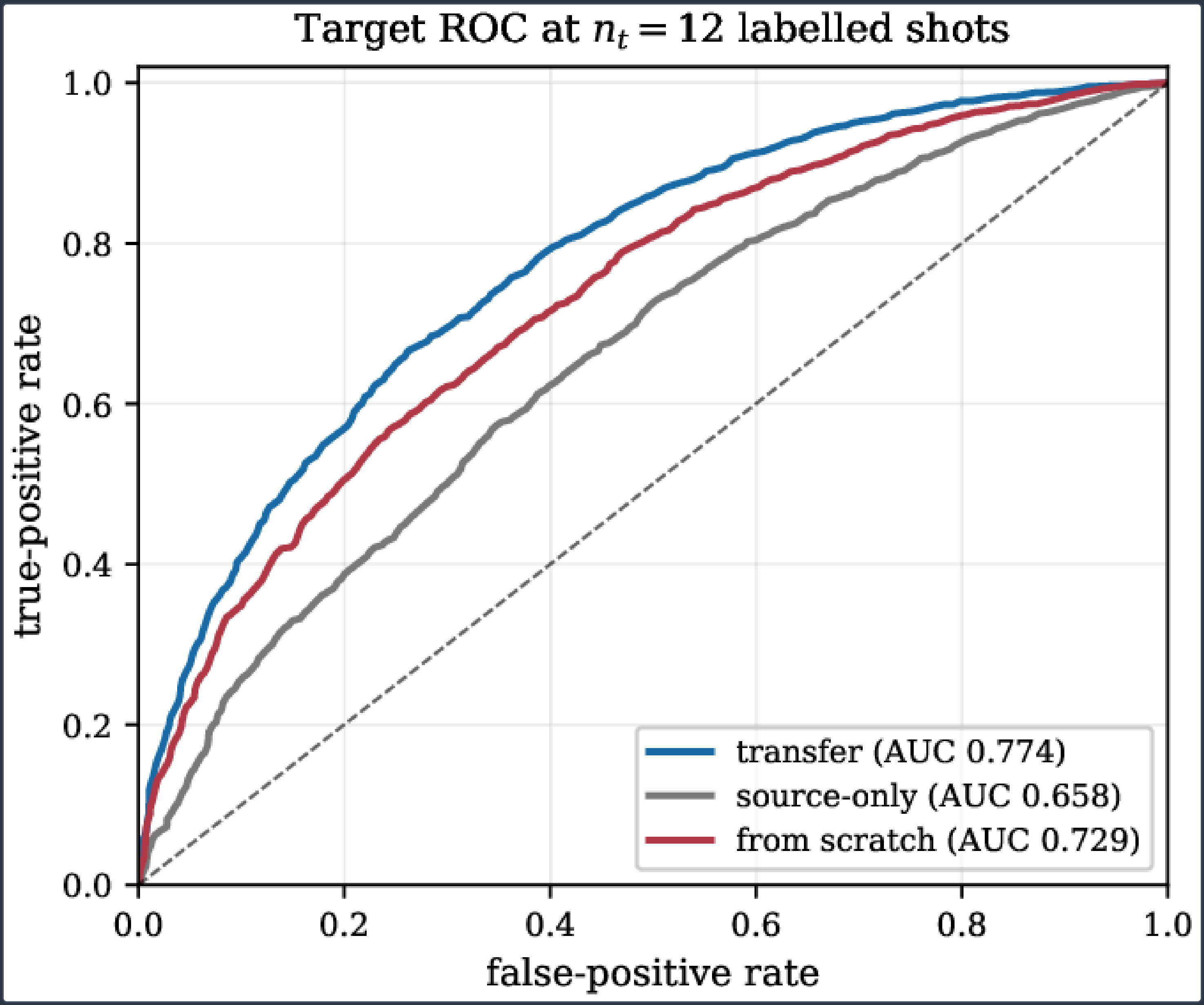
(Synthetic.) Cross-machine transfer gain versus target data (deterministic study, seed 20260726). Held-out target AUC of the transfer model (pre-train + fine-tune) and the from-scratch (target-only) model against the number of labelled target shots n_t; bands are 10–90% over repeated draws. The transfer model dominates at every n_t; the gain is largest in the cold-start regime (+0.05 AUC at $n_t = 5$) and vanishes as the target can train on its own (convergence near 0.85 at $n_t = 400$).

n_t (TARGET SHOTS)	5	12	35	60	100	200	400
from scratch (AUC)	0.62	0.72	0.79	0.81	0.83	0.84	0.84
transfer (AUC)	0.68	0.75	0.80	0.82	0.84	0.85	0.85
transfer gain	**+0.05**	**+0.02**	+0.01	+0.01	+0.01	+0.01	+0.01

(Synthetic.) Held-out target AUC versus labelled target shots n_t (seed 20260726). The transfer gain is positive at every n_t (acceptance criterion A1) and largest in the cold-start regime.

THE OPERATING CURVE AT A COLD-START BUDGET

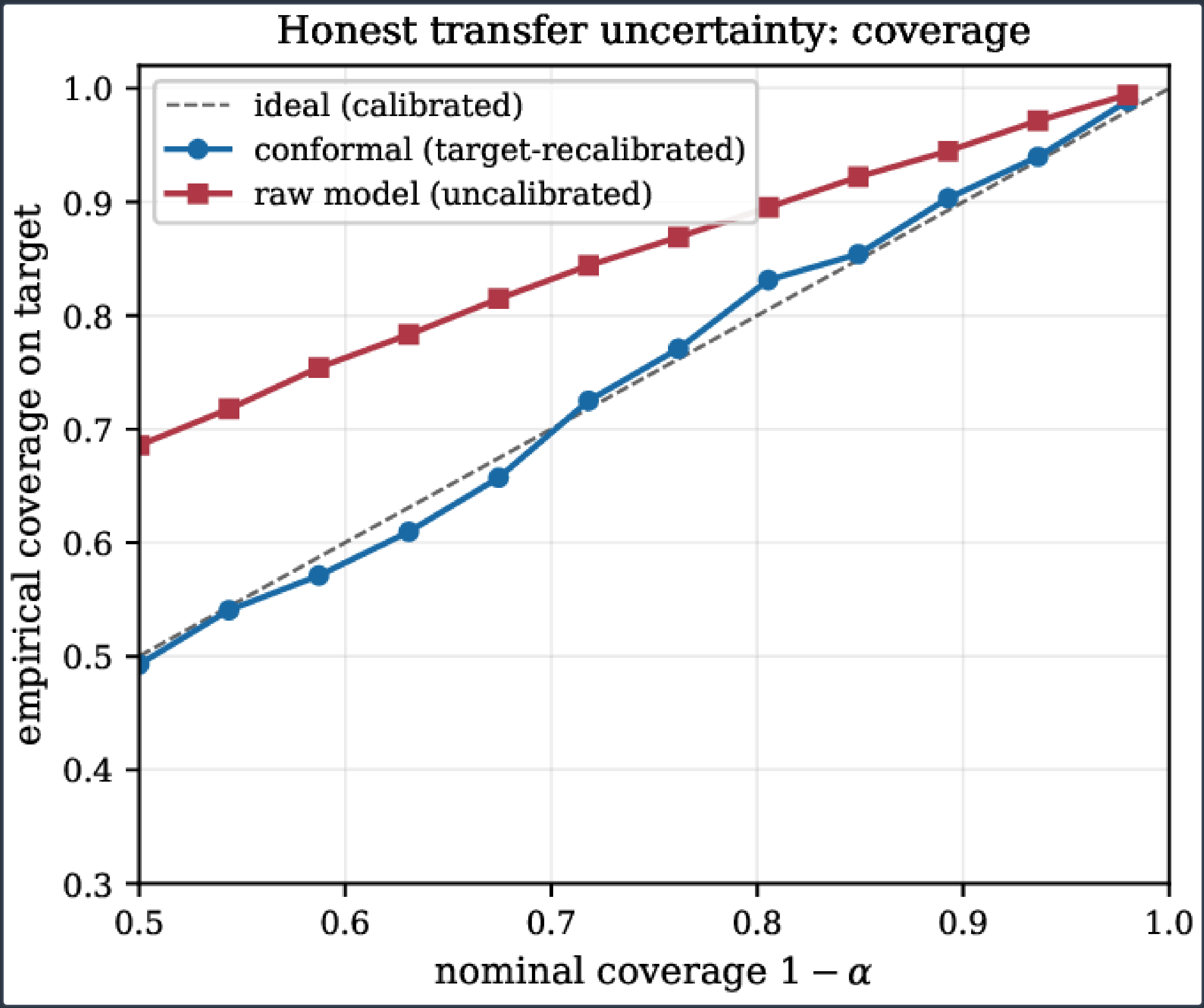
Figure 4 isolates a single cold-start budget, $n_t = 12$ labelled target shots, and compares three models on the held-out target: the transfer model, a from-scratch model, and a source-only model applied with no target fine-tuning at all. The transfer model reaches the highest area under the ROC (0.77), the from-scratch model follows (0.73), and the source-only model is worst (0.66) — direct evidence that neither ingredient suffices alone. The source-only model carries the disruption structure of the public machines but is mis-aligned to the target's operating space; the from-scratch model is aligned by construction but starved of data; only their combination — source structure aligned and then fine-tuned on the few target shots — gets both right. The ordering transfer > from-scratch > source-only is the quantitative statement that cross-machine transfer is a genuine alignment-plus-adaptation problem, not a mere pooling of data.



(Synthetic.) Target ROC at a cold-start budget of $n_t = 12$ labelled shots (seed 20260726). The transfer model (AUC 0.77) beats the from-scratch model (AUC 0.73), which in turn beats the source-only model applied with no fine-tuning (AUC 0.66). Neither source structure nor target data alone suffices; their combination does.

HONEST COVERAGE UNDER TRANSFER

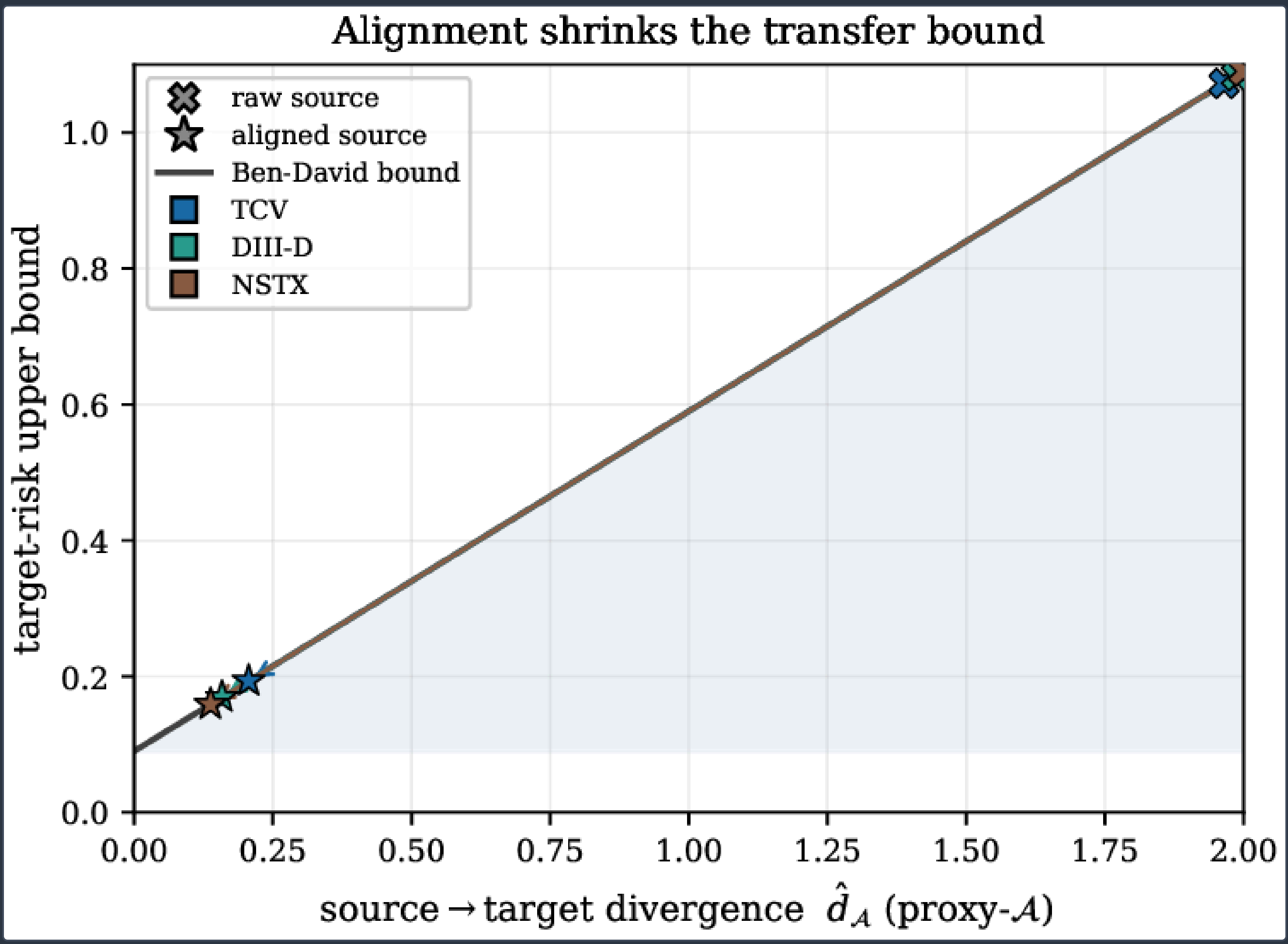
Figure 5 tests the claim in the title — *honest* transfer uncertainty — by measuring, on the shifted target, the empirical coverage of the model's prediction sets against their nominal level. The raw model is badly miscalibrated on the target: its stated confidence over- or under-covers because it was calibrated on a different machine's distribution. The split-conformal recalibration of §3.5, using a small target calibration set, restores coverage to the diagonal, holding mean absolute miscoverage to $\approx 1\%$ across nominal levels from 0.5 to 0.98. This is the operational meaning of honest transfer uncertainty: whatever the transferred model's point accuracy, its *stated* confidence is trustworthy on the target, so a safety layer can act on it. The abstention policy this coverage enables — handing authority back to a deterministic floor out of distribution — is the subject of the calibrated-uncertainty companion (doi:10.5281/zenodo.22645813).



(Synthetic.) Honest transfer uncertainty (seed 20260726). Empirical coverage on the shifted target versus nominal $1-\alpha$. The raw transferred model (red) mis-covers; split-conformal recalibration on a small target set (blue) restores coverage to the ideal diagonal, with mean absolute miscoverage $\approx 1\%$.

THE TRANSFER BOUND AND THE EFFECT OF ALIGNMENT

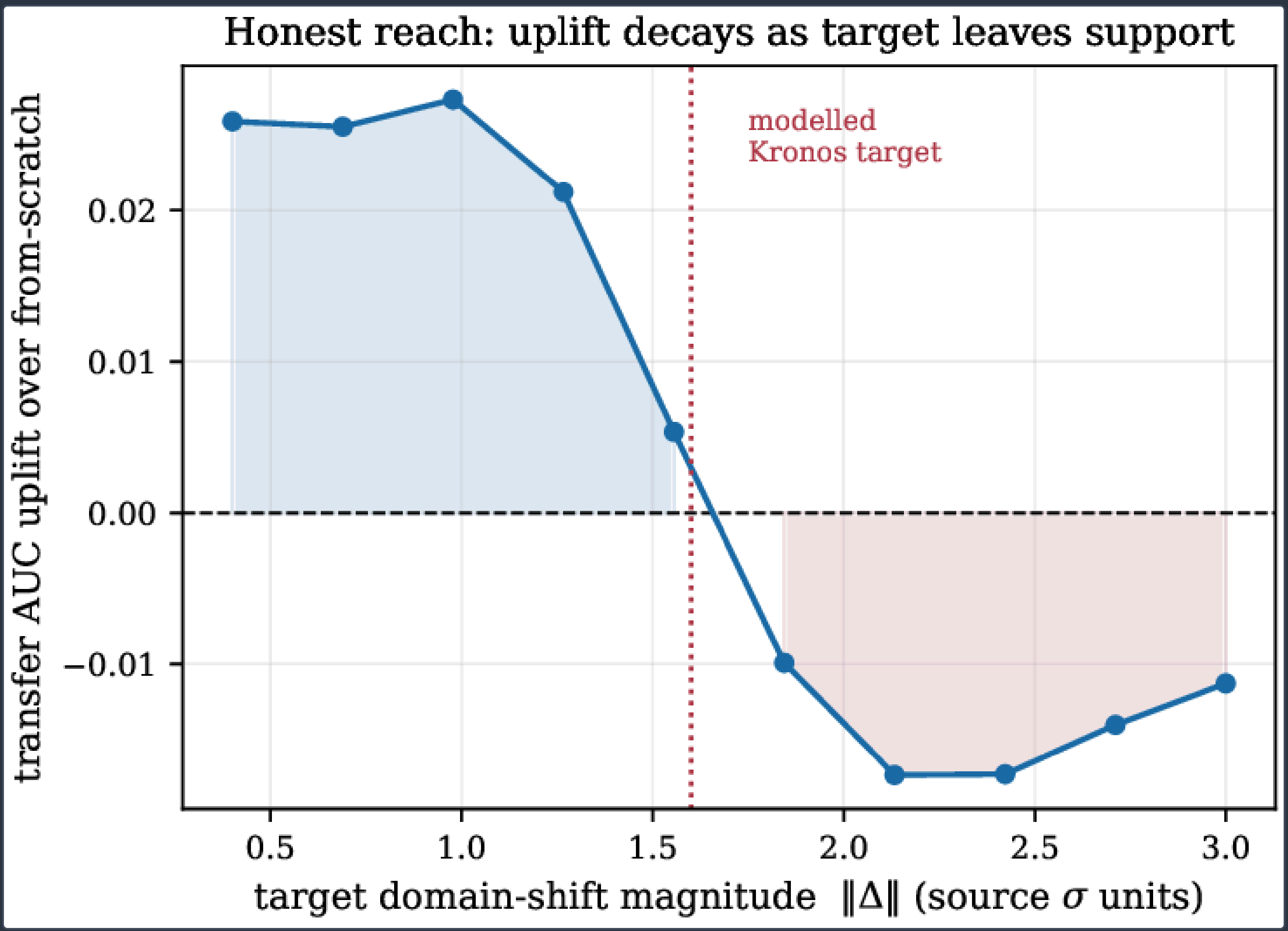
Figure 6 places the three source machines on the Ben-David bound of (4). For each source we estimate the proxy-$\mathcal{A}$-distance to the target twice — from the raw features and from the CORAL-aligned features (7) — and read the corresponding target-risk upper bound off (4). Alignment moves every source leftward along the divergence axis and down the bound: the arrows in figure 6 are the $\frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}$ term shrinking. This is the theory of §3 made concrete on the study's own data — the mechanism by which pre-training with an alignment penalty buys the transfer gain of figure 3 is exactly the reduction of the divergence term in (4). The residual bound that alignment cannot remove is the adaptability floor λ: the part of the target risk set by how genuinely the disruption physics is shared, which no amount of alignment can reduce and which the honest-uncertainty layer must therefore report rather than hide.



(Synthetic.) The Ben-David transfer bound (4) with the three source machines placed by their estimated proxy-$\mathcal{A}$-distance to the target (seed 20260726). Crosses are raw features, stars are CORAL-aligned features; each arrow is the divergence term $\frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}$ shrinking under alignment, lowering the target-risk upper bound. The irreducible offset is the adaptability floor λ.

HONEST REACH: GRACEFUL DEGRADATION AS THE TARGET LEAVES SUPPORT

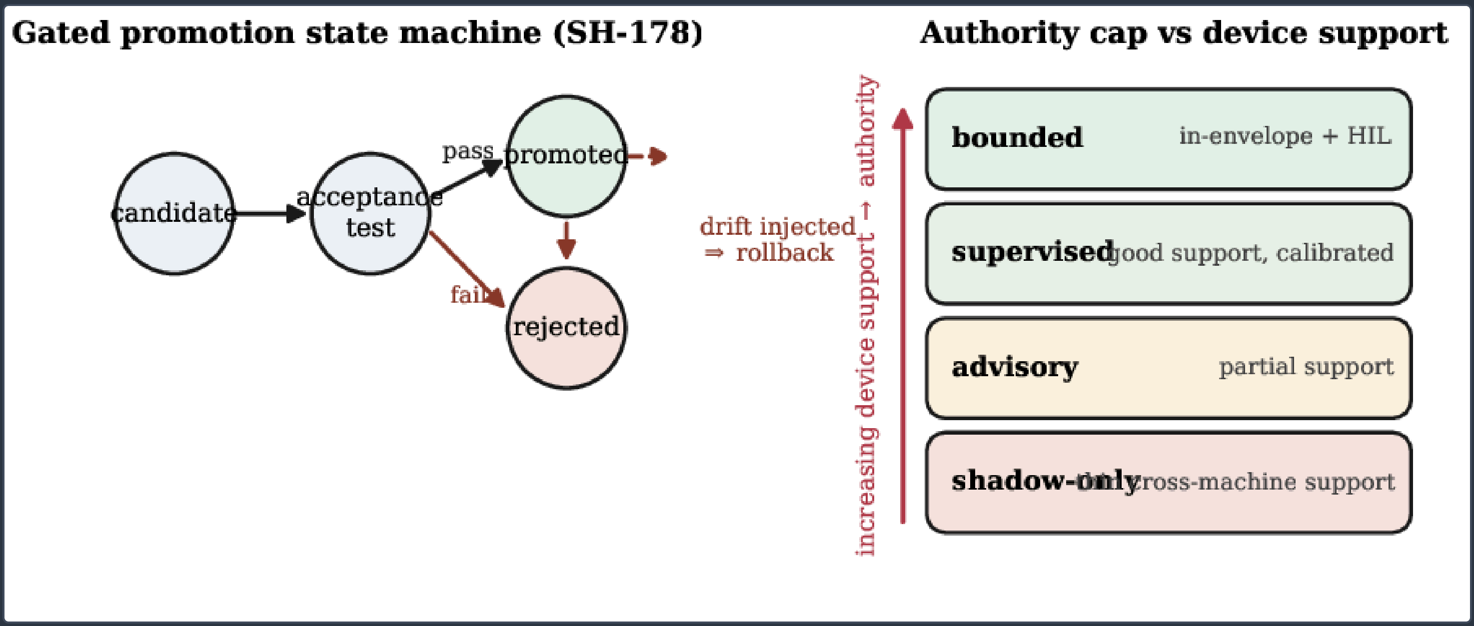
Figure 7 sweeps the target's domain-shift magnitude and reports the transfer model's AUC uplift over the from-scratch model at a fixed cold-start budget. The uplift is positive and large when the target is moderately shifted — the regime the public machines can genuinely inform — and decays smoothly toward zero as the target moves far out of the source support, where the density ratios of (8) diverge and there is simply nothing left to transfer. Crucially the uplift does not go sharply negative: a well-aligned, elastically-anchored transfer model degrades to the from-scratch baseline rather than being actively harmed by irrelevant source data. This graceful degradation is the behaviour the support functional (12) is built to read: at the modelled breeder shift (marked) the uplift is still positive, and as a hypothetical target is pushed further out the architecture reports diminishing support rather than a confident, unsupported prediction.



(Synthetic.) Honest reach (seed 20260726). Transfer AUC uplift over the from-scratch model at a fixed cold-start budget, versus the target domain-shift magnitude in source-σ units. The uplift is largest for moderately shifted targets and decays gracefully toward zero as the target leaves source support; it does not turn sharply negative. The modelled Kronos-target shift is marked.

GATED PROMOTION, ROLLBACK AND AUTHORITY

Figure 8 and table 2 report the governance layer. The gated-promotion state machine (§4.4) admits a candidate model only after it clears the acceptance test, promotes it only when its lineage is complete, and rolls it back automatically when the drift detector fires — the three frozen acceptance behaviours of KFE-CF-SH-178. The authority cap of (13) then keys the model's runtime permission to its device support: shadow-only where support is thin, rising through advisory and supervised to bounded authority as support, calibration and a hardware-in-the-loop confirmation accrue. Table 2 collects the acceptance criteria of §4.5 with their status — the gating, rollback and lineage guarantees defined and in place, and the transfer-gain, coverage and authority-cap criteria demonstrated on the synthetic study.



(Schematic.) The governance layer. *Left:* the gated-promotion state machine (KFE-CF-SH-178) — a candidate is accepted only on passing the acceptance test, promoted only with complete lineage, and rolled back automatically on injected drift. *Right:* the authority cap of equation (13), keyed to cross-machine device support: authority rises from shadow-only to bounded as support, calibration and a hardware-in-the-loop confirmation accrue. The state machine and authority ladder are the gating policy; the acceptance criteria they enforce are the frozen bars of §4.5.

CRITERION	BASIS	STATUS
A1 transfer gain > 0	held-out target AUC (fig. 3)	demonstrated ($+0.05$ at $n_t = 5$)
A2 calibrated coverage	split-conformal (fig. 5)	demonstrated ($\approx 1\%$ miscoverage)
A3 gate blocks bad models	KFE-CF-SH-178	in place
A4 rollback on injected drift	KFE-CF-SH-178	in place
A5 end-to-end lineage	KFE-CF-SH-178	in place
A6 authority cap by support	eq. (13); KX-L1-A7	demonstrated (fig. 7)

Acceptance criteria, basis and status. “In place” denotes a governance guarantee that is defined and enforced now; “demonstrated (synthetic)” denotes a frozen result of the seed-**20260726** study; serials refer to the register [30].

§ 7

Discussion

CROSS-MACHINE TRANSFER IN CONTEXT

The result should be read against the published cross-machine disruption literature, and it is complementary to it. The proof of principle that a disruption model transfers across machines is Kates-Harbeck *et al.*, who trained on DIII-D and improved warning on JET [7]; Montes *et al.* then quantified, across Alcator C-Mod, DIII-D and EAST, how performance falls when a model meets an unseen machine [9]; and Zhu *et al.* built an architecture expressly for general, multi-tokamak prediction under disruptive-example scarcity [10]. Our contribution is not another headline AUC on a public machine — it is the discipline that these works identify as necessary and leave open: a machine-readable ledger of how far the new machine reaches beyond the training set (KX-L1-A7), a calibrated-coverage guarantee that survives the machine boundary (§3.5), and a promotion gate that acts on both. The synthetic study is deliberately evaluated in the hard, cold-start regime — five to twelve labelled target shots — because that is the regime a first-of-a-kind breeder actually occupies, and it is the regime in which the transfer gain is largest ($+0.05$ AUC at $n_t = 5$) and, equally importantly, in which the from-scratch model is most variable. The message is not that a single transferred number beats a single published number; it is that transfer recovers early, low-variance warning in exactly the data-starved corner that a new nuclear machine is forced into, and reports honestly when it cannot.

ALIGNMENT, BOUNDS AND THE LIMITS OF POOLING

The ordering transfer $>$ from-scratch $>$ source-only (figure 4) is the quantitative refutation of the naive hope that one can simply pool public data and deploy. The Ben-David decomposition (4) explains why: pooling controls the source-risk term but does nothing about the divergence term, which alignment must reduce (figure 6) and which no reweighting can remove once the target leaves the source support (equation 8). This is the same boundary the confinement ledger draws physically — the $1.48\times$ field and $3.9\times$ current of KX-L1-A7 — and it is why the architecture's honesty is structural rather than cosmetic: the support functional (12) reads $\sigma \to 0$ on exactly the states where the importance weights diverge, so the authority cap and the confinement ledger flag the same extrapolation for the same reason.

HONEST UNCERTAINTY AGAINST THE CALIBRATION LITERATURE

Deep disruption models, like deep networks generally, are poorly calibrated out of the box [17], and a transferred model is worse still because it is evaluated off its calibration distribution. The split-conformal recalibration recovers distribution-free coverage on the target to $\approx 1\%$ (figure 5), and its covariate-shift weighting (11) is the principled way to keep that guarantee across the machine boundary [16]. Ensemble uncertainty [18] supplies a second, epistemic signal that the support functional folds into the authority cap. Together these make the transferred model's confidence an auditable quantity rather than a number to be trusted on faith — the property a regulator and an operator need before a transferred model is allowed anywhere near a protection trip.

RELATION TO THE KRONOS CONTROL STACK

This paper is one layer of a governed autonomous-control stack and composes with its neighbours. The extrapolation tags it emits are consumed by the out-of-distribution and epistemic-gating layer (doi:10.5281/zenodo.22645819); the calibrated coverage feeds the abstention policy of the calibrated-uncertainty layer (doi:10.5281/zenodo.22645813); the authority cap is the maturity-gated-authority mechanism (doi:10.5281/zenodo.22645795); the gated promotion, lineage and multi-GPU training are the MLOps layer (doi:10.5281/zenodo.22645827); and the whole shadow runs inside the four coupled digital twins (doi:10.5281/zenodo.22645805; https://www.kronosfusionenergy.com/publications/four-coupled-digital-twins-for-a-compact-spherical-toka.html). The extrapolation ledger it inherits (KX-L1-A7) spans both Kronos products — the breeder centrepost at $16.84\,T$ and the distinct, higher-field burner plug at $26.49\,T$ are separate magnets in the ledger, and a transferred model is never permitted to conflate one machine's operating space with the other's.

§ 8

Limitations

One honest gate bounds the claims. The quantitative transfer gain over a from-scratch model is, at the time of writing, a *pre-registered acceptance target* demonstrated on a deterministic synthetic study, not a settled result on the breeder's own data — for the simple reason that the breeder has not yet run. Every signal in the transfer curves, ROC and calibration figures is numerically generated from the frozen ledger and a controlled domain-shift model; the study isolates and demonstrates the transfer *mechanism* (alignment reduces the divergence term, fine-tuning adapts it, conformal recalibration keeps coverage), but the on-machine transfer gain firms up only as the public-tokamak corpus and the breeder's early operating data accrue. What is defined and in place *now* is the part that makes the transfer safe to attempt: the gate that blocks a non-conforming model, the rollback that fires on injected drift, the complete end-to-end lineage, and the extrapolation ledger that bounds the transferred model's authority. That ordering — governance first, headline number second — is the deliberate design of a protection system for a first-of-a-kind nuclear machine.

§ 9

Conclusion

We have formalised cross-machine transfer learning for the disruption model of a compact negative-triangularity spherical-tokamak breeder. The disruption task is a discrete-time hazard (1)–(3); the transfer is domain adaptation whose reachable target risk is the Ben-David $\mathcal{H}\Delta\mathcal{H}$ bound (4); feature alignment (6)–(7) shrinks the divergence term; split-conformal prediction (9)–(11) makes the model's uncertainty honest with a distribution-free coverage guarantee; and the frozen nine-device extrapolation ledger (KX-L1-A7) supplies the support functional (12) and authority cap (13) that flag exactly the states — the $1.48\times$-field, $3.9\times$-current drive extrapolation of the breeder — where a transferred model must not be trusted. On a deterministic, reproducible synthetic study built to the ledger, pre-training plus fine-tuning lifts held-out target AUC by $+0.05$ over a from-scratch model in the cold-start regime and converges to it as target data accrue, while split-conformal recalibration holds miscoverage to $\approx 1\%$ on the shifted target. The bridge that carries all of this (KFE-CF-SH-178) admits a model only through a gate that blocks non-conforming models, rolls back on injected drift, and keeps complete end-to-end lineage. The result is a disciplined, auditable path from public-tokamak data to a governed breeder model — a transfer bridge whose central guarantee is not raw accuracy but honest, gated, reversible promotion.

Acknowledgments tocsectionAcknowledgments Part of the Kronos Fusion Energy 2026 design series. The author thanks the Kronos physics de-risking programme for the frozen result set — in particular the nine-device extrapolation ledger (KX-L1-A7) — underlying every quantitative claim, and acknowledges the public-tokamak communities of TCV, DIII-D and NSTX whose open datasets and published overviews make cross-machine transfer possible.

Reproducibility. This paper is archived at doi:10.5281/zenodo.22645821 and its official version is https://www.kronosfusionenergy.com/publications/cross-machine-transfer-learning-for-a-compact-spherical.html. The transfer basis — the KX-L1-A7 nine-device extrapolation ledger (ratio 0.93–1.46, mean 1.19, standard deviation 0.21; extrapolation $1.48\times$ in B_0, $3.9\times$ in I_p, $2.0\times$ in density) and the frozen breeder configuration 22021 at $Q = 3.076$, $I_p = 9.66\,MA$, $P_{\rm fus} = 85.04\,MW$ — is a frozen entry in the Kronos Physics De-Risking Register (doi:10.5281/zenodo.22133057). Every synthetic figure is regenerable from the deposited generator script under the fixed seed 20260726; the transfer-bridge, gate and lineage schema are archived with this paper.

Data availability tocsectionData availability The device-parameter table, the extrapolation-ledger values, and the deterministic synthetic transfer study used for figures 1–7 are archived with the deposit at doi:10.5281/zenodo.22645821, which references the Kronos 2026 series including the out-of-distribution and epistemic-gating paper (doi:10.5281/zenodo.22645819), the MLOps and gated-promotion paper (doi:10.5281/zenodo.22645827), and the AI and quantum-control paper (doi:10.5281/zenodo.22645702; https://www.kronosfusionenergy.com/publications/ai-quantum-control.html). Public-tokamak data belong to their originating facilities. All physics values trace to the register [30] by the serial identifiers cited inline. Any economic analysis is reported separately and is not part of the public deposit.

Competing interests tocsectionCompeting interests Provisional patent applications relating to aspects of this work (KRONOS-CTRL) have been filed.

tocsectionReferences

APPARATUS

Portfolio, People & Record

The intellectual-property estate, the people who built it, the limitations that gate it, and the sources it rests on.

1	4	40	27
GRANTED PATENT	2026 PROVISIONALS	TEAM MEMBERS	2022 PROVISIONALS

INTELLECTUAL PROPERTY

The patent portfolio

The estate spans a granted high-field magnet patent, pending utility and 2026 provisional filings that map to the two products, a digital-twin control provisional, a registered trademark application, and the original 2022 provisional family. Every patentable disclosure in the five-paper arXiv drop was protected **before** publication: the breeder and burner provisionals were filed 1–2 August, and a publication-gap omnibus filing the evening before the drop swept up the DEC, plasma-control, and REBCO-winding matter of the three otherwise-unprotected papers. Patent numbers and application serials are matters of public record; claim scope is summarised, not reproduced.

GRANTED & PENDING UTILITY

US 12,009,112	High-field tilted graded-REBCO magnet architecture (KRONO-002A) · *app. 17/878,550*	GRANTED
US 17/878,507	Core device / method (KRONO-001A) · *utility*	PENDING

2026 PROVISIONAL FILINGS — MAPPED TO THE PRODUCTS

64/124,209	Hyperion — spherical-tokamak tritium / helium-3 foundry · *breeder · filed Aug 2026*	FILED
64/124,220	Aegis / MetroVolt — synchrotron-cutoff, fuel-selection & plug-winding · *generator · filed Aug 2026*	FILED
64/115,167	KRONOS-CTRL digital-twin plant control · *provisional · filed Jul 2026*	FILED
64/005,440	Integrated spherical-tokamak fusion architecture · *early integrated-architecture filing · filed Mar 2026*	FILED
64/105,530	Negative-triangularity spherical-tokamak architecture · *foundational ST filing · filed Jul 2026*	FILED
64/128,097	Publication-gap omnibus — quasineutral two-species expander DEC · provenance-gated / latency-split plasma control · as-built high-field winding & conductor architecture · *filed 7 Aug 2026, the evening before the arXiv drop*	FILED

TRADEMARK & ORIGIN

FTK-98801894	Kronos Fusion Energy — trademark application	FILED
KR-2022-*	Original provisional family (KR-2022-00001 … 00027) · *27 filings, 2022*	PRIORITY

The narrow, defensible magnet novelty — the specific conductor and the digital-twin winding optimisation, not high-field REBCO as a category — is the commercial core that can earn ahead of any Q > 1 milestone, with markets in fusion magnet supply, MRI/NMR, accelerators, and proton therapy. The claim is scoped honestly: the high field is a *system field* result, not a stand-alone-coil record; the small-bore plug coil is structurally infeasible as a bare winding but resolved by a stress-managed structural shield (feasible-pending-FEA); and the winding-tape experiment returned a *null result*. The value is the method, not a field record.

THE TEAM

Founding partners, board, advisors & auditors

Kronos is built by a bench of advanced-fuel-fusion, high-field-magnet, direct-conversion, and materials specialists, with a board and operations team drawn from defense, national laboratories, and industry. Dates are shown as ranges; where a tenure has ended or is term-ending, the range says so.

FOUNDING PARTNERS — SCIENCE

Dr. Gerald Kulcinski
Co-Founder · Helium-3 & Advanced-Fuel Fusion
UW–Madison · 2023–Present

Dr. Carl Weggel
Chief Scientist / S.M.A.R.T. Design
MIT Alcator Program · 2022–Present

Dr. Robert J. Weggel
Magnetic Field Design
MIT Magnet Lab · 2022–Present

BOARD

Priyanca Ford
Founder & CEO · Executive Board
2022–Present

Bandel Carano
Finance / Strategy
Oak Investment Partners / Stanford · 2025–2026

Patrick Schweiger
Fusion Device Engineering
Oklo / CFS / TerraPower · 2025–Present

SCIENTIFIC ADVISORS

Dr. Steven O. Dean
Fusion Policy & History
DOE / Fusion Power Associates · 2025–Present

Dr. Patrick H. Diamond
Plasma Turbulence & Transport
UC San Diego · 2025–Present

Dr. Jack J. Dongarra
HPC & Numerical Algorithms
Turing Award · 2025–Present

Dr. Nasr M. Ghoniem
Chief Material Scientist
UCLA · 2025–Present

Dr. Siegfried Glenzer
Plasma Physics & Laser Fusion
Stanford / SLAC · 2022–Present

Dr. David A. Hammer
Pulsed-Power Fusion
Cornell · 2025–Present

Dr. Donald A. Spong
Plasma Theory & Confinement
ORNL · 2025–Present

Dr. Curtis Smith
Risk Assessment & Nuclear PRA
MIT / Idaho National Lab · 2025–Present

RADM (Ret.) David Goggins
Naval Engineering & Power-Plant Design
U.S. Navy · 2023–2026

Dr. Konstantin Batygin
Mathematician
Caltech · 2022–2025

Dr. Ruben Fair
Magnet Design / ITER Liaison
PPPL / Jefferson Lab · 2022–2025

Dr. Paul S. Weiss
Materials & Nanotechnology
UCLA · 2022–2025

SCIENTIFIC AUDITORS

Dr. Wilfred A. Cooper
Plasma Equilibrium Theory
EPFL / Swiss Plasma Center · 2025–Present

Dr. Ahmed Hassanein
Plasma–Material Interactions
Purdue / Argonne · 2025–Present

Dr. Patrick E. Hopkins
Extreme Thermal Materials
U. of Virginia · 2025–Present

Dr. Peter Hosemann
Materials Joining & Structural Integrity
UC Berkeley · 2025–Present

Dr. Travis W. Knight
Advanced Nuclear Fuels & Systems
U. of South Carolina · 2025–Present

Dr. Philippe Lebrun
Cryogenics & Magnet Cooling
CERN · 2025–Present

Dr. Nitendra Singh
Nuclear Safety & Fuel Cycle
ITER · 2025–Present

Dr. Guido Van Oost
Magnetic Confinement & Fusion Systems
Ghent University · 2025–Present

Dr. Gary S. Was
Radiation Materials & Structural Integrity
U. of Michigan · 2025–Present

Dr. Ray Sedwick
Direct Energy Conversion
U. of Maryland · 2025

OPERATIONS

Michael Laughlin
Chief Operating Officer
2022–Present

Martin Owens
Chief Strategy Officer
LANL / GE Hitachi · 2022–Present

MG (Ret.) Paul Pardew
Chief Contracting Officer
Army Contracting Command · 2022–Present

David Beck
Fusion Commercialization
U.S. Space Force · 2024–Present

Sushma Bhatia
Environmental & Legislative
Google · 2022–Present

Michael De Frenza
Technology Licensing
2023–Present

MG (Ret.) Robin L. Fontes
Cybersecurity & Defense
Army Cyber Command · 2022–Present

Vijay Gehani
Supply Chain & Components
INOX India · 2024–Present

Jon Michel Greenwood
IT & AI Infrastructure
Live Nation · 2023–Present

Brian C. O'Neill
National Security
CIA / ODNI · 2022–Present

Gen. (Ret.) Gustave F. Perna
DoD Liaison
Operation Warp Speed · 2022–2023

Andrea Romero
Marketing Lead
2022–Present

Legal counsel is retained; those roles are held on the internal roster and are not listed here.

SOURCES

References

1 M E Austin *et al*, Achievement of reactor-relevant performance in negative triangularity shape in the DIII-D tokamak, *Phys. Rev. Lett.* **122** 115001 (2019). doi:10.1103/PhysRevLett.122.115001

2 A Marinoni, O Sauter and S Coda, A brief history of negative triangularity tokamak plasmas, *Rev. Mod. Plasma Phys.* **5** 6 (2021). doi:10.1007/s41614-021-00054-0

3 T C Hender *et al*, Progress in the ITER Physics Basis Chapter 3: MHD stability, operational limits and disruptions, *Nucl. Fusion* **47** S128 (2007). doi:10.1088/0029-5515/47/6/S03

4 A H Boozer, Theory of tokamak disruptions, *Phys. Plasmas* **19** 058101 (2012). doi:10.1063/1.3703327

5 P C de Vries *et al*, Survey of disruption causes at JET, *Nucl. Fusion* **51** 053018 (2011). doi:10.1088/0029-5515/51/5/053018

6 S J Pan and Q Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* **22** 1345 (2010). doi:10.1109/TKDE.2009.191

7 J Kates-Harbeck, A Svyatkovskiy and W Tang, Predicting disruptive instabilities in controlled fusion plasmas through deep learning, *Nature* **568** 526 (2019). doi:10.1038/s41586-019-1116-4

8 C Rea, K J Montes, K G Erickson, R S Granetz and R A Tinguely, A real-time machine learning-based disruption predictor in DIII-D, *Nucl. Fusion* **59** 096016 (2019). doi:10.1088/1741-4326/ab28bf

9 K J Montes *et al*, Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST, *Nucl. Fusion* **59** 096015 (2019). doi:10.1088/1741-4326/ab1df4

10 J X Zhu *et al*, Hybrid deep-learning architecture for general disruption prediction across multiple tokamaks, *Nucl. Fusion* **61** 026007 (2021). doi:10.1088/1741-4326/abc664

11 J Vega *et al*, Disruption prediction with artificial intelligence techniques in tokamak plasmas, *Nat. Phys.* **18** 741 (2022). doi:10.1038/s41567-022-01602-2

12 S Ben-David, J Blitzer, K Crammer, A Kulesza, F Pereira and J W Vaughan, A theory of learning from different domains, *Mach. Learn.* **79** 151 (2010). doi:10.1007/s10994-009-5152-4

13 H Shimodaira, Improving predictive inference under covariate shift by weighting the log-likelihood function, *J. Stat. Plan. Inference* **90** 227 (2000). doi:10.1016/S0378-3758(00)00115-4

14 A Gretton, K M Borgwardt, M J Rasch, B Sch\"olkopf and A Smola, A kernel two-sample test, *J. Mach. Learn. Res.* **13** 723 (2012).

15 Y Ganin *et al*, Domain-adversarial training of neural networks, *J. Mach. Learn. Res.* **17** 1 (2016). arXiv:1505.07818

16 A N Angelopoulos and S Bates, A gentle introduction to conformal prediction and distribution-free uncertainty quantification, *Found. Trends Mach. Learn.* (2023). arXiv:2107.07511

17 C Guo, G Pleiss, Y Sun and K Q Weinberger, On calibration of modern neural networks, *Proc. 34th Int. Conf. on Machine Learning (ICML)* **70** 1321 (2017). arXiv:1706.04599

18 B Lakshminarayanan, A Pritzel and C Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, *Adv. Neural Inf. Process. Syst. (NeurIPS)* **30** (2017). arXiv:1612.01474

19 S Hochreiter and J Schmidhuber, Long short-term memory, *Neural Comput.* **9** 1735 (1997). doi:10.1162/neco.1997.9.8.1735

20 J Degrave *et al*, Magnetic control of tokamak plasmas through deep reinforcement learning, *Nature* **602** 414 (2022). doi:10.1038/s41586-021-04301-9

21 S Coda *et al*, Physics research on the TCV tokamak facility: from conventional to alternative scenarios and beyond, *Nucl. Fusion* **59** 112023 (2019).

22 J L Luxon, A design retrospective of the DIII-D tokamak, *Nucl. Fusion* **42** 614 (2002).

23 M Ono *et al*, Exploration of spherical torus physics in the NSTX device, *Nucl. Fusion* **40** 557 (2000).

24 J E Menard *et al*, Overview of the physics and engineering design of NSTX upgrade, *Nucl. Fusion* **52** 083015 (2012).

25 J E Menard *et al*, Fusion nuclear science facilities and pilot plants based on the spherical tokamak, *Nucl. Fusion* **56** 106023 (2016). doi:10.1088/0029-5515/56/10/106023

26 L L Lao, H St John, R D Stambaugh, A G Kellman and W Pfeiffer, Reconstruction of current profile parameters and plasma shapes in tokamaks, *Nucl. Fusion* **25** 1611 (1985). doi:10.1088/0029-5515/25/11/007

27 N C Amorisco *et al*, FreeGSNKE: a Python-based dynamic free-boundary toroidal plasma equilibrium solver, *Phys. Plasmas* **31** 042517 (2024). doi:10.1063/5.0188467

28 J Candy, E A Belli and R V Bravenec, A high-accuracy Eulerian gyrokinetic solver for collisional plasmas, *J. Comput. Phys.* **324** 73 (2016). doi:10.1016/j.jcp.2016.07.039

29 P K Romano, N E Horelik, B R Herman, A G Nelson, B Forget and K Smith, OpenMC: a state-of-the-art Monte Carlo code for research and development, *Ann. Nucl. Energy* **82** 90 (2015). doi:10.1016/j.anucene.2014.07.048

30 P I Ford, *Kronos Physics De-Risking Register*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22133057

COLOPHON

How this document was made

The Kronos Fleet — Combined Editorial, 2026 is set in **Fraunces** (display), **Gelasio** (text), and **IBM Plex Mono** (data & equations), carried forward from the Kronos editorial design system.

The figures are rendered at 300 dpi from the deposited generator scripts; the document is composed as a single self-contained file with embedded fonts and imagery, paginated to US Letter, and rendered to PDF through a headless Chromium engine in matched light and dark editions.

Every headline quantity carries an evidence class; withdrawn and restated values are marked as such. Nothing herein is frozen beyond the founder-approved Tier-A set.

Explore it live — turn the interactive 3-D model, run the deposited solvers in the physics validator, and watch the companion film series.

LEGAL NOTICE

Notice, status, and distribution

Conceptual design & simulation study. This document reports a conceptual design and simulation study. It is not a construction commitment, a safety-analysis report, a regulatory filing, or an offer of securities. Forward-looking statements — schedules, costs, market sizes, and performance — are estimates subject to the limitations set out in the Simulations part and may change.

Numbers and their classes. Quantities are reported with evidence classes and case labels; modelled, assumed, and requirement-class values are identified as such and are not to be quoted without their labels.

Public edition. This edition carries the physics and design only; all financial, funding, and defense-commercial content has been removed for public distribution. The scientific text corresponds to the arXiv and journal submissions.

THE 2026 KRONOS SERIES

One programme, many papers

This editorial is one paper in the Kronos 2026 series — a real fusion company's complete, reproducible physics record for a spherical-tokamak breeder and its low-neutron D–^{3}He tandem-mirror burners. Every headline number is a frozen anchor in the Physics De-Risking Register.

Explore the whole programme: the De-Risking Register, the interactive 3-D model, and the physics validator.

KRONOS FUSION ENERGY

Cross-Machine Transfer Learning for a Compact Spherical-Tokamak Breeder: A Disruption Model Pre-Trained on TCV, DIII-D, and NSTX with Honest Transfer Uncertainty

A real fusion company building real machines. Explore the science, live.

Zenodo · doi:10.5281/zenodo.22645821

Read on kronosfusionenergy.com

The Physics De-Risking Register