KRONOS FUSION ENERGY

PAPER 2.79 · THE KRONOS 2026 SERIES

Out-of-Distribution Detection and Epistemic Gating for Autonomous Fusion Control: A Machine-Readable Extrapolation Ledger That Throttles Model Authority Beyond the Validated Envelope

A learned controller extrapolates smoothly and confidently into regions where it has no evidence, and in a nuclear system that confident wrongness off-distribution is the dominant...

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 Several quantities in the underlying corpus moved after first publication; where a value has been withdrawn or restated, the record says so plainly.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

Out-of-Distribution Detection and Epistemic Gating for Autonomous Fusion Control: A Machine-Readable Extrapolation Ledger That Throttles Model Authority Beyond the Validated Envelope

A learned controller extrapolates smoothly and confidently into regions where it has no evidence, and in a nuclear system that confident wrongness off-distribution is the dominant failure mode. The safeguard is not a better model but an honest one: a controller that measures the boundary of its own competence in real time and refuses authority beyond it. We formalise the epistemic-gating layer of the Kronos autonomous fusion controller and reduce it to a set of computable inequalities. The plant state is scored against a *mapped* validated envelope — a nine-point design-closure grid spanning fusion gain $Q = 2.735$–3.460 and fusion power $P_{\text{fus}} = 58.6$–118.9 MW with all physics gates satisfied and the frozen design point $(Q = 3.076,\ P_{\text{fus}} = 85.04\,\text{MW})$ interior — by two complementary detectors: a reconstruction/feature-space out-of-distribution (OOD) score and a machine-readable *extrapolation ledger* resolved across a nine-device breeder-and-burner post-diction base. Authority is proportioned to *calibrated* epistemic uncertainty, decomposed from aleatoric noise through a deep-ensemble total-variance law and wrapped in a distribution-free conformal coverage guarantee; the guarded confinement model carries a posterior $H_{98} = 1.007 \pm 0.030$ from a 200-evaluation surrogate-accelerated Bayesian-optimization campaign, and the verification suite discharges $19/19$ formal anchors. A single composite rule — the authority ceiling is the minimum of the OOD, ledger and uncertainty gates — throttles the learned controller smoothly toward a deterministic control-barrier floor as the operating point approaches, and then leaves, the envelope. We state the governing detector equations, the numerical formulation (autoencoder threshold calibration, Gaussian-process Bayesian optimization, envelope corner mapping, formal verification), and the GPU-accelerated campaign that produced every quantity, and we benchmark the gating stack against the out-of-distribution-detection, uncertainty-calibration and disruption-prediction literature. The result is the credibility layer beneath the autonomy: authority bounded by evidence, degrading gracefully rather than silently. The one honest gate — tightening the reconstruction threshold against machine diagnostics rather than the twin's validated distribution — is deferred to first operating hours, and the gate is conservative by construction in the interim.

Open-access deposit (code & data, CC BY 4.0): DOI 10.5281/zenodo.22645819 · Zenodo community *kronos_fusion_energy*. Explore it live: interactive 3-D model · physics validator.

Keywords: out-of-distribution detection · extrapolation ledger · epistemic uncertainty · authority gating · conformal prediction · deep ensembles · conservative control · validated envelope · spherical tokamak

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 / nth / 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)

§ 1

Introduction

AUTONOMOUS CONTROL OF A COMPACT SPHERICAL TOKAMAK

The Kronos breeder Hyperion is a compact, high-power-density negative-triangularity spherical tokamak: design point plasma current $I_p = 9.66$ MA, fusion gain $Q = 3.076$, fusion power $P_{\text{fus}} = 85.04$ MW, triangularity $\delta = -0.30$, on-axis field $B_0 = 8$ T, elongation $\kappa = 2.0$, major radius $R_0 = 1.20$ m, aspect ratio $A = 2.5$, and a resistive-copper centrepost that peaks at 16.84 T. The companion D–^{3}He tandem-mirror burner (Aegis / MetroVolt) operates at engineering gain $Q_E = 1.318$ with a neutron power fraction $f_n = 5.44\%$, a 26.49 T high-field superconducting plug, and a direct-energy converter of efficiency $\eta_{\text{DEC}} = 0.49$; the breeder centrepost and the burner plug are two *distinct* magnets, and the ledger of §2.3 keeps them separate. A machine at this density leaves little time to react and no tolerance for an unmodelled excursion, so its control stack is model-based and increasingly autonomous, from real-time equilibrium reconstruction and profile prediction to learned magnetic control [1,2,3,4]. The four coupled Kronos digital twins run ahead of the plant as a predictive shadow [32], and neural-operator flux and equilibrium surrogates supply the speed that closed-loop control demands [33,22,23].

THE OFF-DISTRIBUTION FAILURE MODE

The characteristic and most dangerous failure of a learned controller is not high variance but confident error where it has no data: a smooth interpolant extrapolates past its evidence with undiminished — often *increasing* — apparent confidence [14,26,13]. In a nuclear plant this is inadmissible. The remedy is a system that (i) recognises, per control tick, when the live state is unlike anything the models were validated on; (ii) records, in a machine-readable form, exactly where each model has evidence; and (iii) hands authority back to a deterministic floor *before* a model is trusted out of its depth. This is the classical separation of *epistemic* uncertainty — reducible ignorance, “the model does not know” — from *aleatoric* noise — irreducible scatter [12,17]: the safety layer must act on the former and tolerate the latter.

PRIOR ART

Out-of-distribution detection is a mature subfield. Baseline softmax-confidence detectors [8], feature-space Mahalanobis detectors [9], deep-ensemble predictive uncertainty [10], Monte-Carlo-dropout Bayesian approximations [11] and reconstruction-based anomaly detectors [16,15] each supply a score that separates in-distribution from novel inputs. Distribution-free conformal prediction turns any such score into a finite-sample coverage guarantee [18,19]. In fusion, data-driven disruption predictors on DIII-D, C-Mod and EAST demonstrate that a learned warning can be both accurate and cross-machine [5,6,7], and control-theoretic safety filters — control-barrier functions rendering a safe set forward-invariant [24] — provide the deterministic fallback. What has been missing is the *integration*: a single, auditable rule that fuses an OOD detector, a machine-readable record of model evidence, and a calibrated uncertainty into a runtime authority ceiling for a fusion controller, grounded in a frozen, independently reproduced design register.

CONTRIBUTION

This paper supplies that integration for the Kronos controller and states it as mathematics. We: (1) define the validated envelope operationally from the Kronos design-closure corner mapping (anchor KX-L1-A1) and give the formal in-distribution predicate (§2.1); (2) construct two complementary detectors — a reconstruction/feature OOD score (serial KFE-CF-SH-040) and an extrapolation-ledger support measure resolved over nine devices spanning the breeder and burner (anchor KX-L1-A7; monitor serial KFE-CF-SH-075) — and derive their calibration (§§2.2–2.3); (3) decompose predictive uncertainty into epistemic and aleatoric parts, attach a distribution-free conformal guarantee, and read the guarded confinement posterior $H_{98} = 1.007 \pm 0.030$ directly from the surrogate-accelerated Bayesian-optimization campaign (anchor KX-L1-A14, §2.4); (4) fuse the three into a single composite authority-throttle law with a deterministic control-barrier floor (§2.5); and (5) verify the stack against the formal V&V suite ($19/19$ anchors, KX-L1-A6) and against the published OOD, calibration and disruption-prediction benchmarks (§§4–5). Every quantitative claim traces to a frozen, reproduced anchor in the Kronos physics de-risking register [27], and the whole campaign ran on the NVIDIA GPU HPC campaign described in §3. This paper is the epistemic-gating member of the Kronos autonomy series, sitting between the calibrated-uncertainty-and-abstention result [28] and the runtime-assurance-with-signed-ledger result [29], and beneath the maturity-gated actuation-authority argument [30].

§ 2

Governing equations and the formal gating problem

PLANT, CONTROLLER AND THE VALIDATED ENVELOPE

Write the plant as a discrete-time stochastic system on state $x \in \mathbb{R}^{n_x}$ with actuation $u \in \mathbb{R}^{n_u}$,

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) + w_k, & w_k &\sim \mathcal{N}(0, Q_w), \\ y_k &= h(x_k) + v_k, & v_k &\sim \mathcal{N}(0, R_v), \end{aligned}$$

with $y \in \mathbb{R}^{n_y}$ the diagnostic vector. The controller proposes an action $u_k^{\mathrm{AI}} = \pi_\theta(y_{0:k})$ from a learned policy π_θ whose transition and observation models are validated only on a bounded region of operating space. That region is not asserted; it is *mapped*. The Kronos design-closure study (KX-L1-A1) evaluates the coupled 0-D/1-D closure over a nine-point corner grid $\{\xi_j\}_{j=1}^{9}$ and records, for each node, whether all physics gates (power balance, stability, heat-flux, breeding and current-drive) are satisfied. The gates hold across the grid, whose projection onto the observable pair (Q, P_{fus}) spans

$$Q \in [2.735,\ 3.460], \qquad P_{\mathrm{fus}} \in [58.6,\ 118.9]\ \mathrm{MW},$$

with the frozen design point $\xi^\star = (Q^\star, P_{\mathrm{fus}}^\star) = (3.076,\ 85.04\,\mathrm{MW})$ in the interior (figure 1). Define the validated envelope $\mathcal{E}$ as the closed convex hull of the mapped grid in a whitened design coordinate, and the in-distribution predicate

$$\mathrm{ID}(x) \equiv \big[\Phi(x) \in \mathcal{E}\big], \qquad \Phi : \mathbb{R}^{n_x} \to \mathbb{R}^d$$

where Φ is the projection of the state onto the design-closure coordinates. Operation with $\mathrm{ID}(x) = \mathsf{true}$ is supported by design closure; operation with $\mathrm{ID}(x) = \mathsf{false}$ is, by definition, extrapolation. The gating problem is to bound the authority granted to π_θ by a continuous, calibrated proxy for how far outside $\mathcal{E}$ the live point has moved.

THE OUT-OF-DISTRIBUTION SCORE

We use a whitening map that renders $\mathcal{E}$ isotropic. Let μ and Σ be the mean and covariance of the validated design points in Φ-space; the affine whitening $T(x) = \Sigma^{-1/2}(\Phi(x) - \mu)$ gives the feature-space (Mahalanobis) score [9]

$$s_{\mathrm{M}}(x) = \|T(x)\|_2 = \sqrt{(\Phi(x) - \mu)^\top \Sigma^{-1}(\Phi(x) - \mu)},$$

and, normalising each closure coordinate to its mapped half-range Δ_i, the box-referenced score

$$s(x) = \|T(x)\|_\infty = \max_i \frac{|\Phi_i(x) - \mu_i|}{\Delta_i},$$

so that $s < 1$ is interior, $s = 1$ is the envelope boundary and $s > 1$ is out-of-distribution (figure 2). Equation (5) is the geometric detector; the learned detector (SH-040) complements it. A convolutional autoencoder $g_\phi \circ e_\phi$ is trained on the validated state distribution to minimise reconstruction error, and flags a live state by the residual

$$\varepsilon_r(x) = \|x - g_\phi(e_\phi(x))\|_2.$$

A state far from the training manifold reconstructs poorly, so ε_r is large exactly where the models lose evidence [16,15]. The detector *threshold* τ is calibrated, not chosen, by fixing a target false-positive rate α_{fp} on held-out in-distribution data,

$$\tau = \inf\Big\{t : \widehat{\Pr}\big[\varepsilon_r(x) > t \,\big|\, \mathrm{ID}(x)\big] \le \alpha_{\mathrm{fp}}\Big\} = \hat{q}_{1-\alpha_{\mathrm{fp}}}(\varepsilon_r),$$

the empirical $(1 - \alpha_{\mathrm{fp}})$ quantile of the in-distribution reconstruction error (figure 3). The trip predicate is $\varepsilon_r(x) > \tau$.

Graceful degradation. Diagnostics drop out under radiation and off-normal events, so the detector must hold its judgement with missing channels. Partition the observation index set into observed O and missing M; the detector conditions on the observed sub-vector by reconstructing over O only,

$$\varepsilon_r^O(x) = \big\|x_O - [g_\phi \circ e_\phi]_O(x_O)\big\|_2, \qquad \tau_O = \hat{q}_{1-\alpha_{\mathrm{fp}}}(\varepsilon_r^O),$$

with a channel-count-indexed threshold family $\{\tau_O\}$ so that the false-positive rate is held constant as $|M|$ grows (figure 4). The acceptance requirement is that the detector flag a held-out regime it was never trained on, and, on that flag, drop the controller to conservative control.

THE EXTRAPOLATION LEDGER

The OOD score is the fast tripwire; the extrapolation ledger is the slower, authoritative record of *where* each model has evidence. The Kronos multi-device post-diction (KX-L1-A7) resolves a nine-device ledger, spanning the breeder and the burner, that records for each device d and validated case a support kernel in design space. The ledger support at a query point is the kernel-weighted case density

$$\rho(\mathbf{x}) = \frac{1}{Z}\sum_{d=1}^{9}\omega_d \sum_{c\in\mathcal{C}_d} K_\ell\big(\Phi(\mathbf{x}) - \Phi(\mathbf{x}_{d,c})\big), \qquad K_\ell(\mathbf{r}) = \exp\Big(-\tfrac{1}{2}\mathbf{r}^\top \ell^{-2}\mathbf{r}\Big),$$

with device weights ω_d, per-device validated case sets $\mathcal{C}_d$, bandwidth ℓ and normaliser Z. A point with dense support (ρ high) is granted full model authority; a point at the edge (intermediate ρ) is granted reduced authority with widened uncertainty; a point with no support ($\rho \to 0$) is denied model authority and handed to the deterministic floor. The ledger is queried at run time and is machine-readable: it returns, with each verdict, the devices and cases that support (or fail to support) a prediction, so the authority decision is auditable case by case (figure 8) and closes under the runtime-assurance provenance record [29].

EPISTEMIC UNCERTAINTY, CALIBRATION AND CONFORMAL COVERAGE

Authority is proportioned to *calibrated* uncertainty, not to raw model confidence, and only the epistemic part should throttle a model. For a deep ensemble of E members $\{p_e(y\,|\,\mathbf{x})\}_{e=1}^{E}$ with member means $\hat{\mu}_e$ and variances $\hat{\sigma}_e^2$ [10], the law of total variance splits the predictive variance into an aleatoric mean-of-variances and an epistemic variance-of-means,

$$\underbrace{\mathrm{Var}[y\,|\,\mathbf{x}]}_{\text{predictive}} = \underbrace{\frac{1}{E}\sum_{e=1}^{E}\hat{\sigma}_e^2(\mathbf{x})}_{\text{aleatoric}} + \underbrace{\frac{1}{E}\sum_{e=1}^{E}\big(\hat{\mu}_e(\mathbf{x}) - \bar{\mu}(\mathbf{x})\big)^2}_{\text{epistemic }\sigma_{\text{ep}}^2(\mathbf{x})}, \qquad \bar{\mu} = \frac{1}{E}\sum_e \hat{\mu}_e.$$

The epistemic term σ_{ep}^2 grows off-distribution — ensemble members disagree where they were never constrained — and it is σ_{ep}, not the aleatoric floor, that the gate reads. Calibration is enforced by the coverage condition: the empirical coverage of the model's α-credible intervals must match nominal, against Monte-Carlo truth,

$$\big|\,\mathrm{Cov}_{\text{emp}}(\alpha) - \alpha\,\big| \;\le\; 3\% \qquad \forall\,\alpha,$$

and upgraded to a distribution-free *guarantee* by conformal prediction: with exchangeable calibration residuals and nonconformity score $R(\mathbf{x}, y)$, the prediction set $\hat{C}(\mathbf{x}) = \{y : R(\mathbf{x}, y) \le \hat{q}_{1-\alpha}\}$ built from the $\lceil (1-\alpha)(N_{\text{cal}}+1)\rceil / N_{\text{cal}}$ empirical quantile of calibration scores satisfies [18,19]

$$\Pr\big[\,y \in \hat{C}(\mathbf{x})\,\big] \;\ge\; 1-\alpha,$$

with no distributional assumption and valid in finite samples. The guarded confinement model is calibrated by the surrogate-accelerated Bayesian-optimization campaign (KX-L1-A14): a 200-evaluation study yields a tight, well-centred confinement posterior

$$H_{98} = 1.007 \pm 0.030$$

(figure 5), a 95% credible interval of $[0.948,\ 1.066]$ that the gate reads directly. The same campaign illustrates precisely the hazard the gate exists for: its unconstrained optimizer, allowed to widen the priors and push H_{98} to a bound of 1.1, reports a bound-limited best-case gain of $Q \approx 8.3$ at $P_{\text{fus}} \approx 61$ MW — an aggressive extrapolation far outside the mapped envelope of equation (2). The gate refuses that excursion authority and holds the controller at the anchored, calibrated point.

THE COMPOSITE AUTHORITY-THROTTLE LAW AND THE DETERMINISTIC FLOOR

Each detector defines an authority ceiling in $[0,1]$. From the OOD score (5), with throttle onset s_0,

$$A_{\mathrm{ood}}(x) = \begin{cases} 1, & s(x) \le s_0, \\ \dfrac{1 - s(x)}{1 - s_0}, & s_0 < s(x) < 1, \\ 0, & s(x) \ge 1; \end{cases}$$

from the ledger support (9) and the epistemic width (10), the saturating ceilings

$$A_{\mathrm{led}}(x) = \sigma_L\big(\rho(x) - \rho_0\big), \qquad A_\sigma(x) = \left[1 - \frac{\sigma_{\mathrm{ep}}(x)}{\sigma_{\max}}\right]_+,$$

with σ_L a logistic and $[\cdot]_+$ the positive part. The composite authority is the *minimum* — the most conservative gate wins —

$$A(x) = \min\left\{ A_{\mathrm{ood}}(x),\ A_{\mathrm{led}}(x),\ A_\sigma(x) \right\} \in [0,1],$$

and the command actually applied blends the learned action with the deterministic floor in proportion to authority,

$$u_k = A(x_k)\, u_k^{\mathrm{AI}} + \big(1 - A(x_k)\big)\, u_k^{\mathrm{floor}},$$

so that $A = 1$ is full learned authority and $A = 0$ is the pure floor (figure 6). The floor is not a constant but a certified conservative controller whose action is the control-barrier projection of a safe reference: given a barrier h with safe set $\mathcal{C}_{\mathrm{safe}} = \{x : h(x) \ge 0\}$,

$$u^{\mathrm{floor}} = \arg\min_{u} \| u - u^{\mathrm{ref}} \|_2^2 \quad \text{s.t.} \quad L_f h(x) + L_g h(x)\, u \ge -\gamma\big(h(x)\big),$$

which renders $\mathcal{C}_{\mathrm{safe}}$ forward-invariant [24]: the plant cannot leave the safe set even when the learned controller is wrong. The certified safety filter is developed in the companion result [31]; here it is the terminal authority of equation (17). Equations (5)–(18) are the complete gating law; §4 reports the reproduced evidence behind each layer.

§ 3

Computational methodology: the NVIDIA GPU HPC campaign

Every quantity in §2 was produced by the Kronos NVIDIA GPU HPC campaign, run on an A100/H100/A10-class accelerator fleet through the Kronos Research Toolkit; the fast protective loops (the geometric OOD score, the ledger lookup, the self-consistency monitor and the orchestrator) run on the in-house HPC digital-twin node so that they never contend with the training fleet. No compute cost enters this paper.

ENVELOPE CORNER MAPPING (KX-L1-A1)

The validated envelope (2) is the projection of a design-closure corner map. The coupled 0-D power balance and 1-D transport closure is evaluated at the nine grid nodes $\{\xi_j\}$; at each node the physics gates are checked and the pair (Q, P_{fus}) recorded. The closure is solved by fixed-point iteration on the operating point to a relative residual $\leq 10^{-6}$; three of the nine corners run over the temperature-duty limit and are retained as boundary markers, fixing the mapped extremes of equation (2). The corner map defines μ, Σ and the half-ranges Δ_i used in the scores (4)–(5).

RECONSTRUCTION-DETECTOR TRAINING AND THRESHOLD CALIBRATION

The autoencoder $g_\phi \circ e_\phi$ of equation (6) is trained by stochastic gradient descent (Adam, decoupled weight decay) on the validated state distribution generated by the twin, minimising the mean squared reconstruction error with an early-stopping criterion on a held-out split; training is GPU-accelerated on the fleet. The threshold τ and its channel-indexed family $\{\tau_O\}$ are then set by the empirical-quantile rule of equations (7)–(8) at the target false-positive rate, so calibration is a deterministic post-processing step on held-out reconstruction errors rather than a trained parameter. Detector inference is a single forward pass evaluated within one control tick.

SURROGATE-ACCELERATED BAYESIAN OPTIMIZATION AND UQ (KX-L1-A14)

The confinement posterior (13) comes from a Gaussian-process Bayesian-optimization campaign [20,21]. A GP with an anisotropic squared-exponential kernel is fit to a 200-evaluation design: an initial space-filling set that includes the frozen anchor, followed by expected-improvement-guided evaluations of a surrogate confinement/gain model. The GP posterior at a query ξ is Gaussian,

$$\mu_{\text{GP}}(\xi) = k(\xi)^\top (K + \sigma_n^2 I)^{-1} y,$$
$$\sigma_{\text{GP}}^2(\xi) = k(\xi, \xi) - k(\xi)^\top (K + \sigma_n^2 I)^{-1} k(\xi),$$

and the expected-improvement acquisition over the incumbent $y^\star$ is

$$\text{EI}(\xi) = (y^\star - \mu_{\text{GP}})\,\Phi_{\mathcal{N}}\Big(\tfrac{y^\star - \mu_{\text{GP}}}{\sigma_{\text{GP}}}\Big) + \sigma_{\text{GP}}\,\varphi_{\mathcal{N}}\Big(\tfrac{y^\star - \mu_{\text{GP}}}{\sigma_{\text{GP}}}\Big),$$

with $\Phi_{\mathcal{N}}, \varphi_{\mathcal{N}}$ the standard normal CDF and PDF. Marginalising the GP posterior over the validated design region yields the confinement posterior of equation (13); the coverage of its credible intervals is checked against Monte-Carlo truth to the 3% tolerance of equation (11). The same run records the bound-limited optimizer excursion ($Q \approx 8.3$) cited in §2.4.

FORMAL VERIFICATION AND VALIDATION (KX-L1-A6)

The suite that gates model promotion is the formal verification-and-validation battery: the method of manufactured solutions (MMS) for code order-of-accuracy, the grid-convergence index (GCI) for discretization error [25], variance-based Sobol sensitivity for input attribution, and a predictive-capability maturity-model (PCMM) score for readiness. The suite discharges $19/19$ anchors PASS (figure 7); MMS recovers the design order to within tolerance, GCI falls below the acceptance bound under refinement, the Sobol indices converge, and the PCMM entries clear their thresholds. These are the discharged proofs the ledger monitor (SH-075) requires before it certifies a lineage.

REPRODUCIBILITY AND SCALE

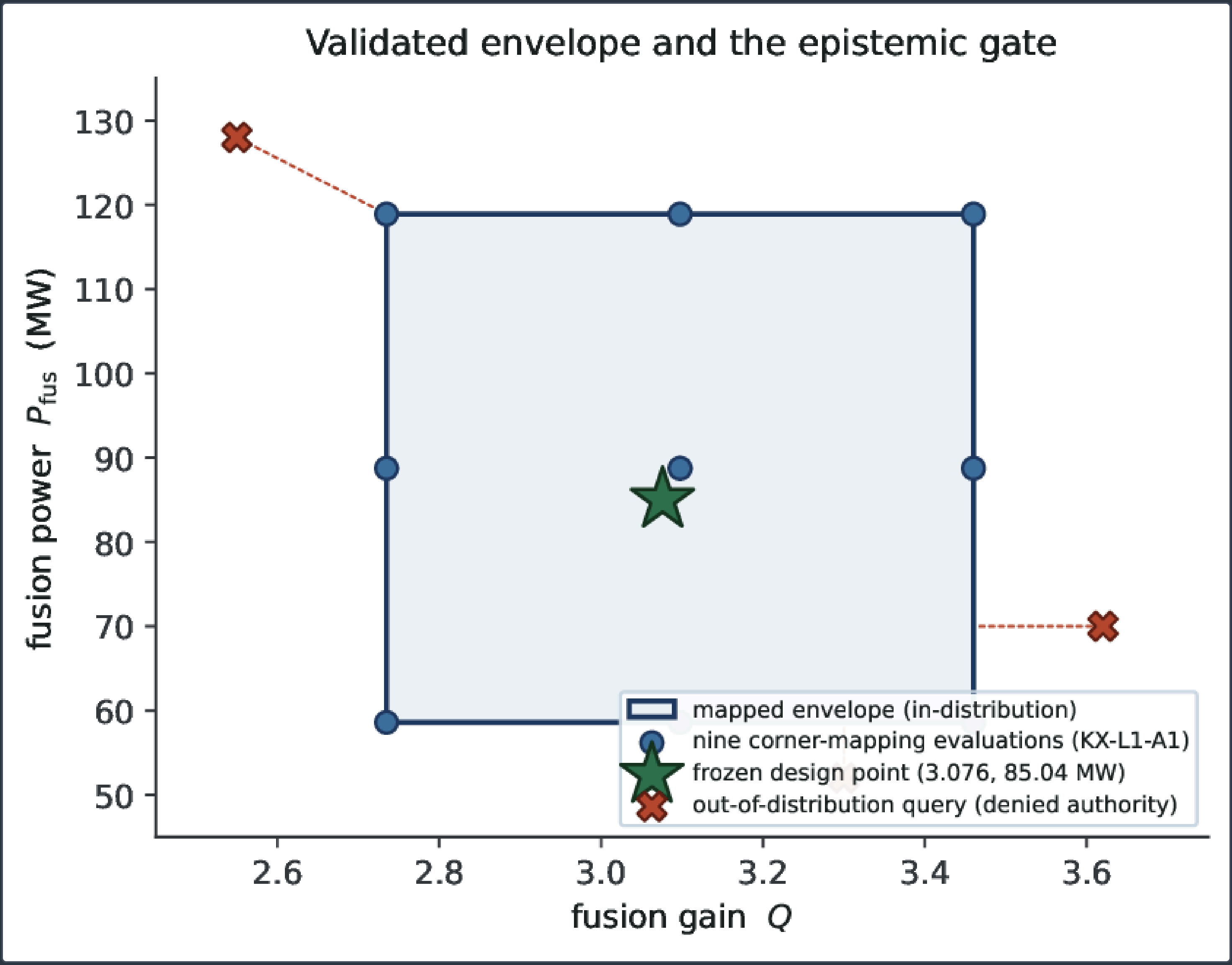
The four gating anchors are frozen entries in the register [27] and regenerable from the archived closure, UQ and verification runs. The problem scale is set by the training and ensemble-forecast workloads on the GPU fleet and the quantile-calibration and ledger-kernel evaluations on the twin node; each figure in §4 is produced by a single deposited generation script from the frozen anchor values. The independent-library reproduction discipline is the subject of the verification-backbone companion [35].

§ 4

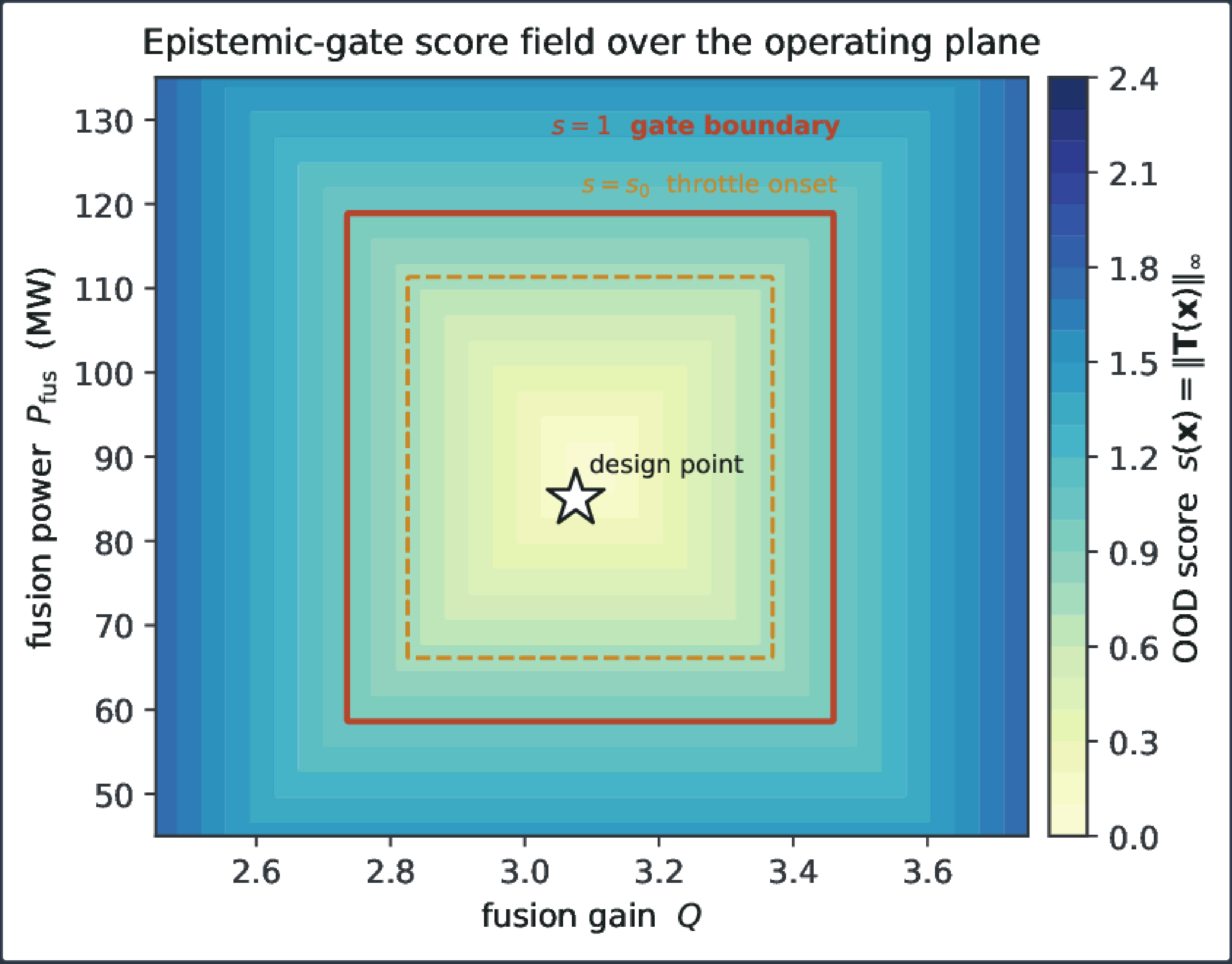
Results

THE VALIDATED ENVELOPE AND THE GATE SCORE FIELD

Figure 1 shows the mapped envelope in the (Q, P_{fus}) plane: the nine corner-mapping evaluations span $Q = 2.735$–3.460 and $P_{\text{fus}} = 58.6$ –118.9 MW, and the frozen design point $(3.076,\ 85.04\,\text{MW})$ sits in the interior. Query points outside the mapped box are flagged out-of-distribution and denied model authority. Figure 2 renders the continuous gate score $s(\quad x)$ of equation (5) over the same plane: the interior is low-score, the boundary $s = 1$ is the hard gate, and the throttle onset $s = s_0$ is the inner contour where authority begins to fall.



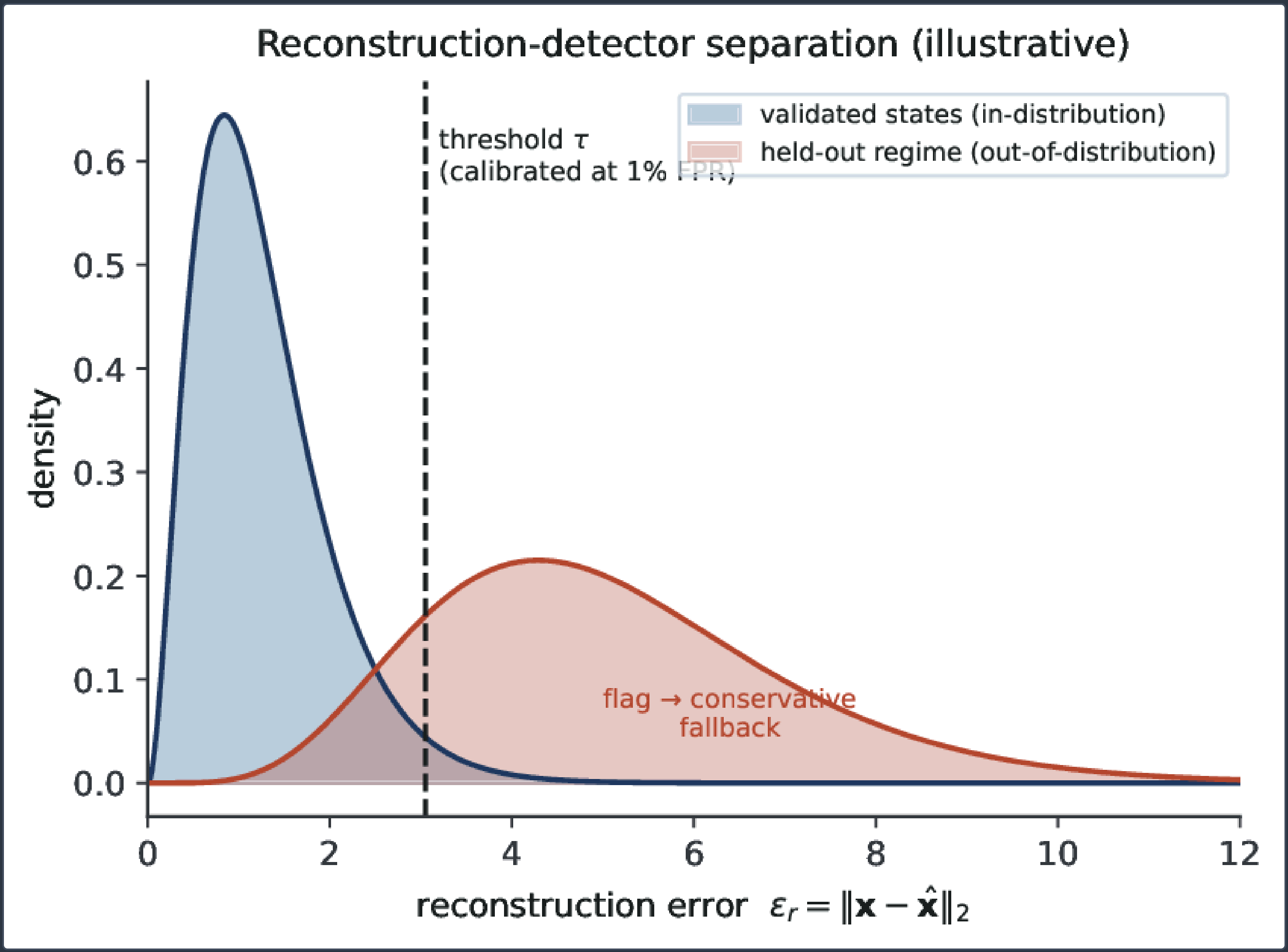
The validated envelope in the (Q, P_{fus}) plane. The bounding box of the mapped design-closure envelope (KX-L1-A1) is spanned by the nine corner-mapping evaluations; the frozen design point $(Q = 3.076,\ P_{\text{fus}} = 85.04\,\text{MW})$ is interior. Out-of-distribution query points are flagged and denied model authority.



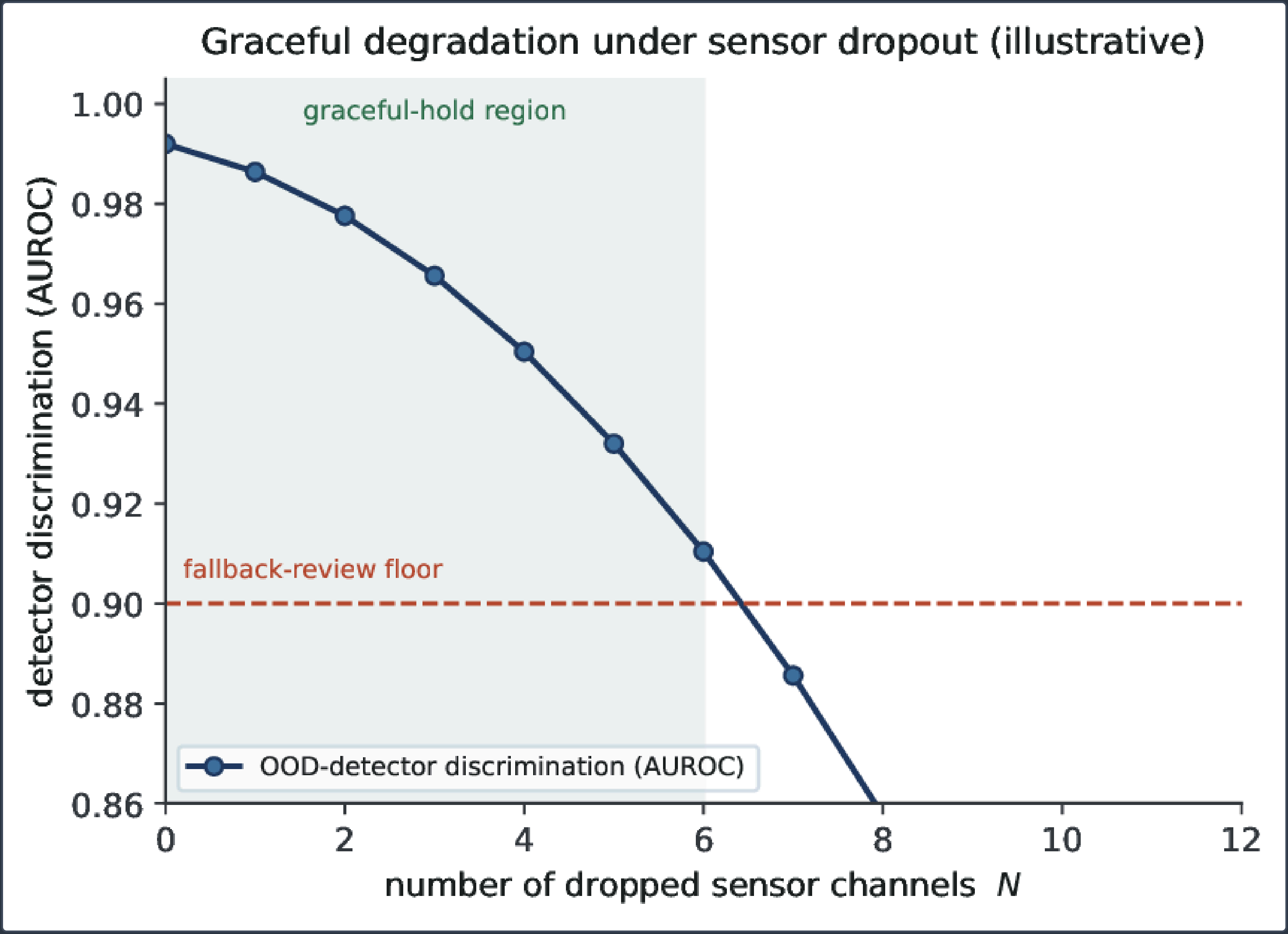
The epistemic-gate score field $s(x) = \|T(x)\|_\infty$ of equation (5) over the operating plane. The heavy contour $s = 1$ is the envelope boundary; the dashed contour $s = s_0$ is the throttle onset. The score is a defined gate function over the real mapped envelope, not measured data.

THE RECONSTRUCTION DETECTOR AND GRACEFUL DEGRADATION

Figure 3 illustrates the reconstruction-detector principle of equations (6)–(7): the in-distribution reconstruction-error density is concentrated at low error, the held-out (out-of-distribution) regime is displaced to high error, and the calibrated threshold τ at a 1% false-positive rate separates them, tripping the conservative fallback. Figure 4 shows the intended graceful-degradation behaviour of equation (8): detector discrimination declines slowly and monotonically as sensor channels are dropped, holding above a fallback-review floor over a wide range before the controller reverts.



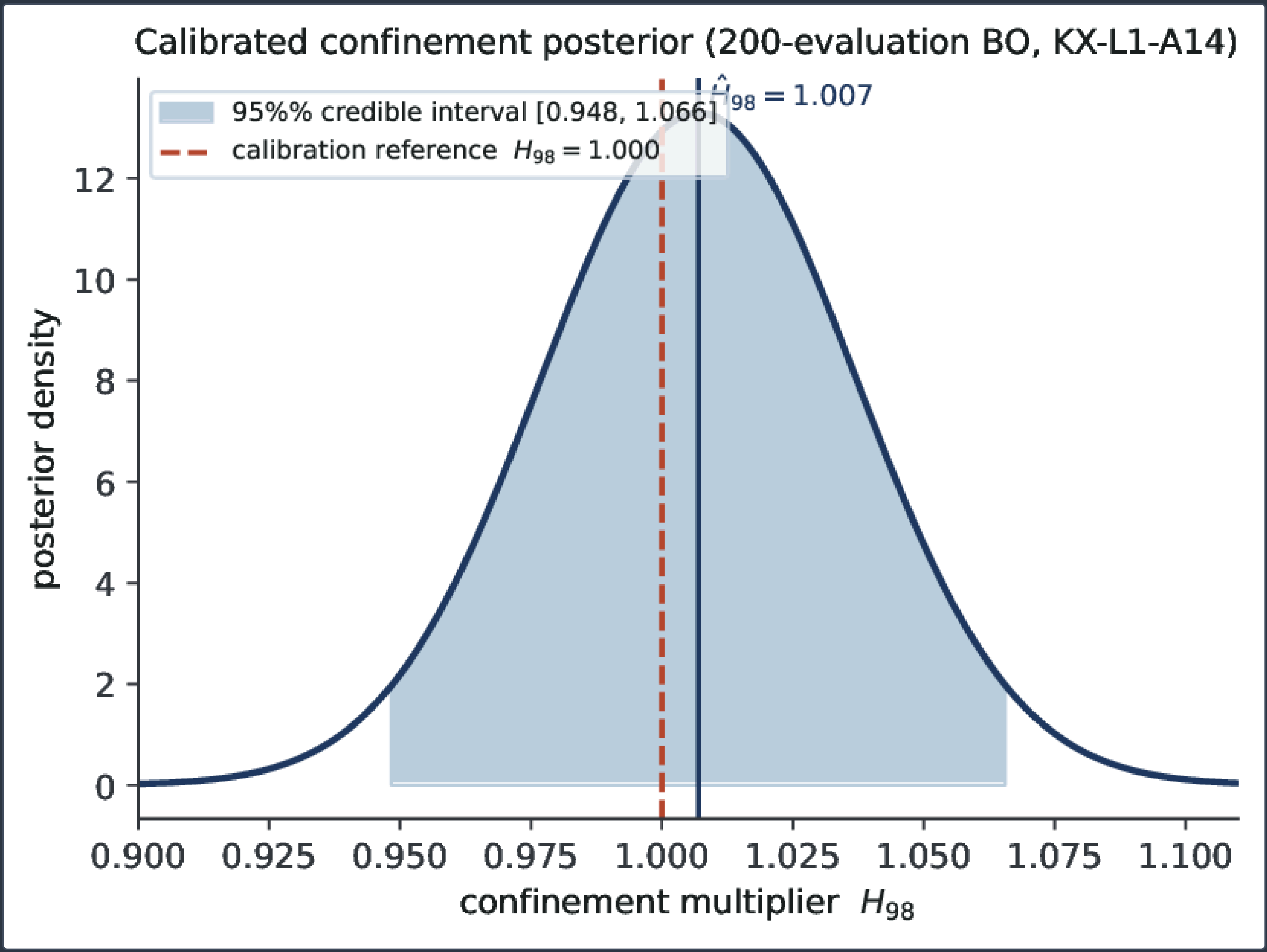
(Schematic.) Reconstruction-detector separation. In-distribution and held-out reconstruction-error densities with the calibrated threshold τ set at a 1% false-positive rate by equation (7). The distributions are illustrative of the detection principle; no measured values are plotted.



(Schematic.) Graceful degradation under sensor dropout. Detector discrimination versus the number of dropped channels N with the channel-indexed thresholds of equation (8); the shaded band marks the graceful-hold region above the fallback-review floor. Illustrative; no measured values are plotted.

CALIBRATED UNCERTAINTY

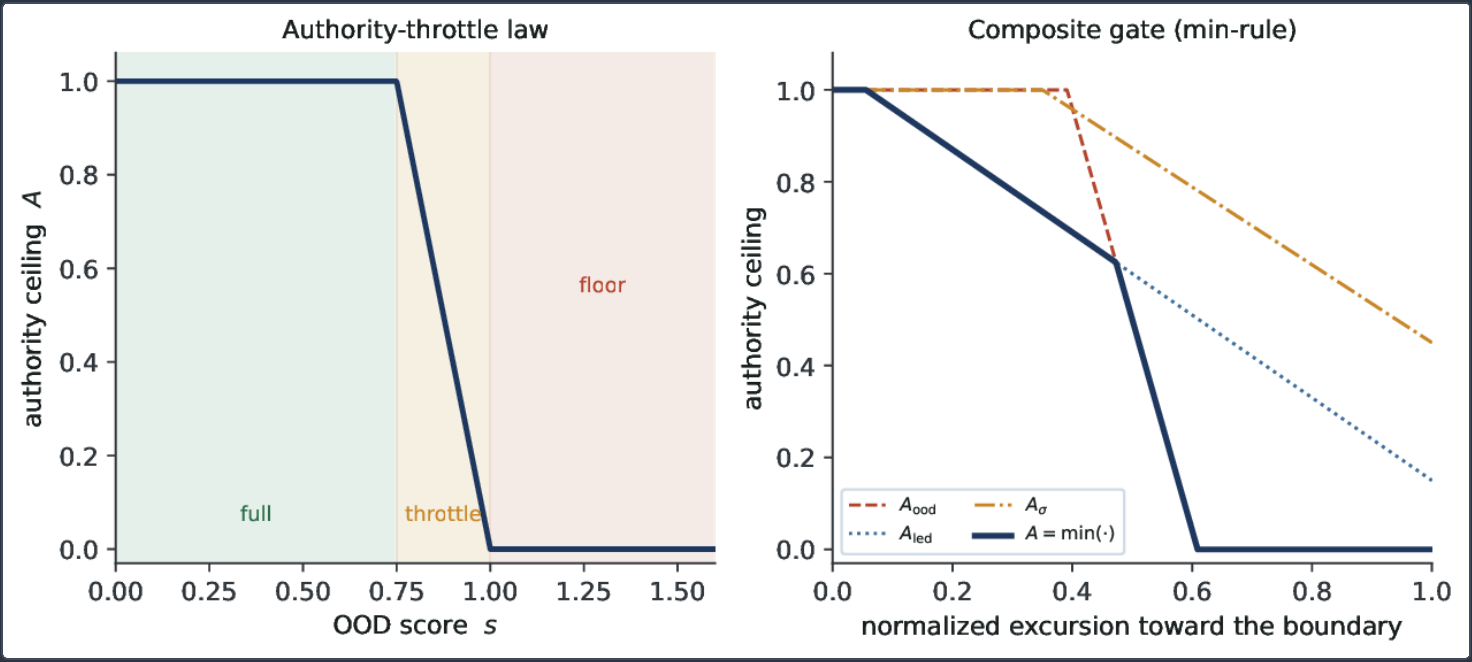
Figure 5 shows the guarded confinement posterior of equation (13): $H_{98} = 1.007 \pm 0.030$, a tight Gaussian centred essentially on the calibration reference $H_{98} = 1.0$, with a 95% credible interval of $[0.948, 1.066]$. This is the calibrated width the authority gate A_σ reads through the epistemic term of equation (10).



Calibrated confinement posterior $H_{98} = 1.007 \pm 0.030$ from the 200-evaluation surrogate-accelerated Bayesian-optimization campaign (KX-L1-A14). Shaded: the 95% credible interval; dashed: the calibration reference $H_{98} = 1.000$. Values are the frozen register anchor.

THE COMPOSITE AUTHORITY GATE

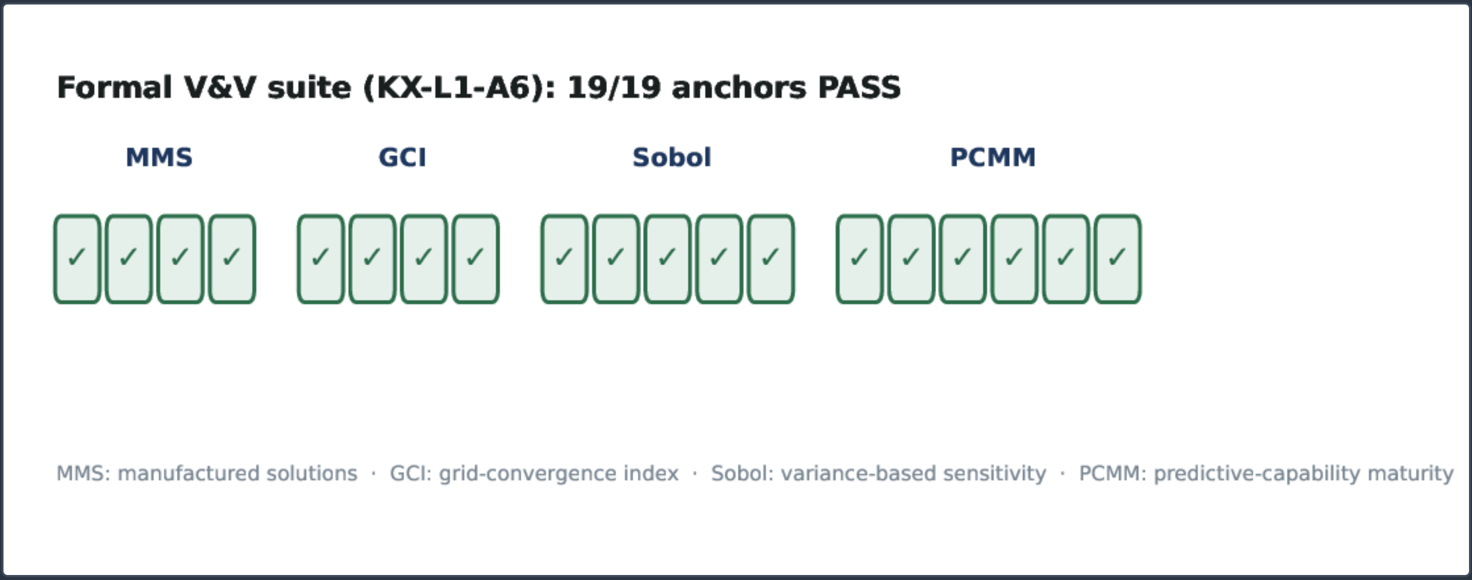
Figure 6 plots the gating law of §2.5: the piecewise authority-throttle $A_{\text{ood}}(s)$ of equation (14) (left), and the composite min-rule of equation (16) along a trajectory that leaves the envelope (right), where the OOD, ledger and uncertainty ceilings each cap authority independently and the minimum governs. Authority falls smoothly to the deterministic floor.



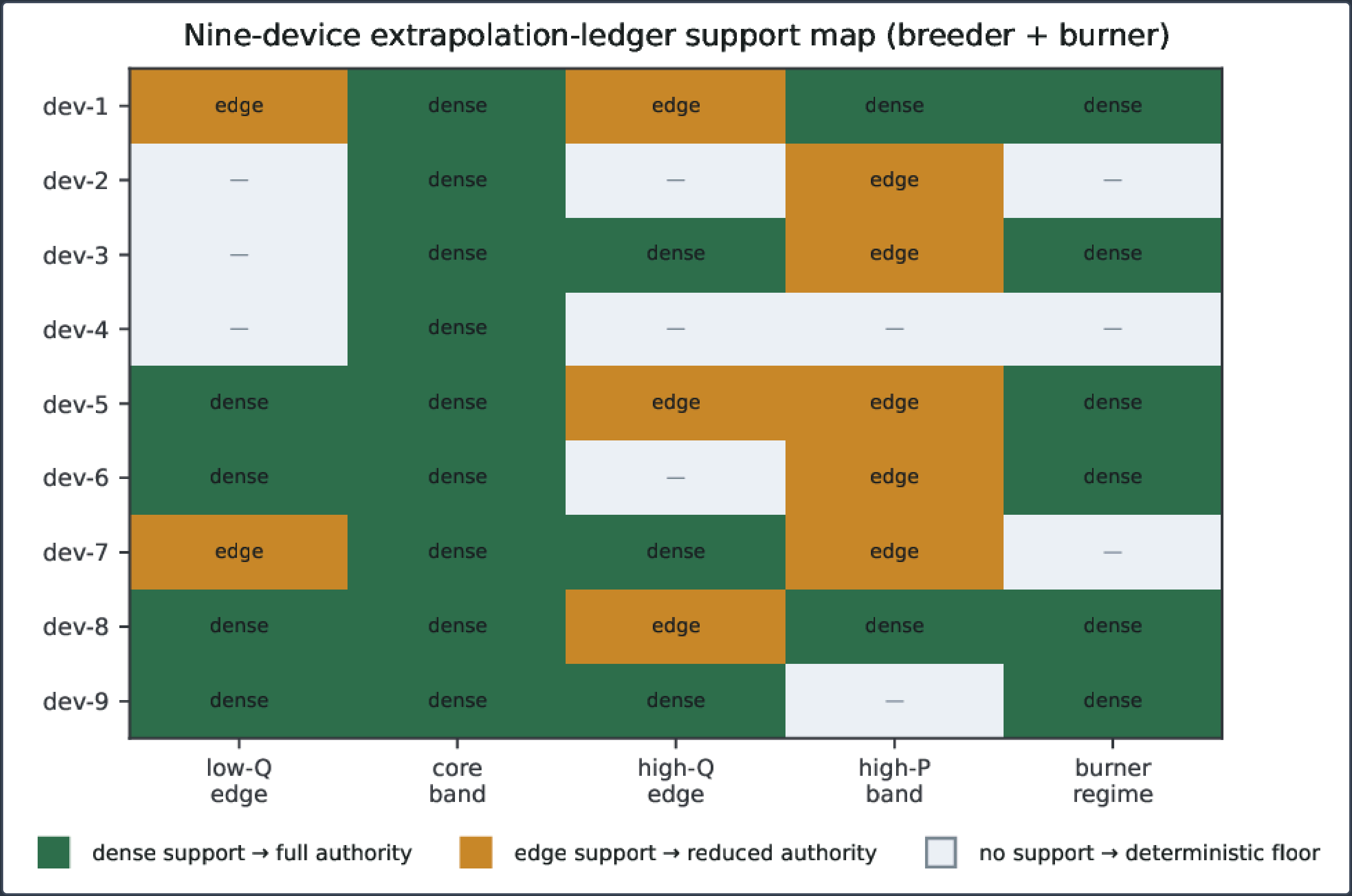
The authority-throttle law. *Left:* $A_{\text{ood}}(s)$ of equation (14) — full authority for $s \le s_0$, a linear throttle to $s = 1$, and the deterministic floor beyond. *Right:* the composite $A = \min(A_{\text{ood}}, A_{\text{led}}, A_{\sigma})$ of equation (16) along an excursion toward the boundary; the most conservative gate wins. Both panels plot defined gate functions.

DISCHARGED PROOFS AND LEDGER SUPPORT

Figure 7 records the formal verification status: the KX-L1-A6 suite discharges 19/19 anchors PASS across the MMS, GCI, Sobol and PCMM classes — the proofs the ledger monitor requires before certifying a lineage. Figure 8 shows the structure of the nine-device extrapolation-ledger support map of equation (9): each device contributes support across operating regions spanning the breeder and the burner, and the composite support density sets A_{led} point by point. Table 1 collects the epistemic-gating stack — each layer, its anchor, and the reproduced evidence behind it.



Formal verification-and-validation status (KX-L1-A6): 19/19 anchors PASS, grouped by the four verification classes of the suite (method of manufactured solutions; grid-convergence index; variance-based Sobol sensitivity; predictive-capability maturity model). The count is the frozen register anchor.



(Schematic.) The nine-device extrapolation-ledger support map (KX-L1-A7), spanning the breeder and the burner. Cells encode support class — dense (full authority), edge (reduced authority) or none (deterministic floor) — as computed from the kernel support of equation (9). The nine-device, breeder-plus-burner structure is grounded; the per-cell support pattern is illustrative.

LAYER	ANCHOR / SERIAL	REPRODUCED EVIDENCE	
validated envelope	KX-L1-A1	Q 2.735–3.460, P_{fus} 58.6–118.9 MW, 9-pt grid	
extrapolation ledger	KX-L1-A7	9-device post-diction, breeder + burner	
calibrated uncertainty	KX-L1-A14	$H_{98} = 1.007 \pm 0.030$ posterior (200-eval BO)	
discharged proofs	KX-L1-A6	formal V\	V 19/19 anchors PASS
OOD detector	KFE-CF-SH-040	flags held-out regime; conservative fallback	
ledger monitor	KFE-CF-SH-075	flags out-of-range operation; throttles authority	

The epistemic-gating stack — each layer, its frozen anchor or serial, and the reproduced evidence.

LATENCY AND THE CLOSED-LOOP RULE

In closed loop the three layers compose into a single per-tick rule. The geometric OOD score (5) and the ledger lookup (9) are $O(nd)$ and $O(\sum_d |\mathcal{C}_d|)$ evaluations that complete within one control tick on the twin node; the reconstruction detector (6) is a single GPU forward pass; the ensemble epistemic width (10) is read from the running forecast. Their minimum (16) sets the authority ceiling, and the blend (17) applies it. When the point is interior with dense ledger support and tight uncertainty, the learned controller acts at full authority; when any gate degrades, authority is throttled smoothly; when the point leaves the mapped envelope, the controller falls to the deterministic floor of equation (18). Authority is thereby bounded by evidence at all times.

§ 5

Discussion

THE GATE AGAINST THE OOD-DETECTION LITERATURE

The two-detector design deliberately fuses complementary failure modes. A feature-space Mahalanobis score (4) is the strong baseline for detecting inputs far from the training manifold [9]; a reconstruction score (6) catches structured novelty that a low-dimensional distance can miss [16,15]; and the geometric box score (5) ties both to the *physically* validated envelope rather than to an abstract density. Against the canonical baselines — softmax confidence [8], deep ensembles [10], MC-dropout [11] — the Kronos detector is distinguished less by a single AUROC number than by its coupling to a design-closure map with reproduced physics gates: the "in-distribution" set is not a data cloud but a verified operating envelope. This is the property that makes the detector's verdict admissible in a nuclear safety case.

CALIBRATION AND THE CONFORMAL GUARANTEE

Modern networks are systematically over-confident [13], so a raw confidence is unsafe to gate on. The Kronos stack instead reads a *calibrated* width — the coverage condition of equation (11) holds the empirical coverage to 3% of nominal — and wraps it in the distribution-free conformal guarantee of equation (12), which holds in finite samples under exchangeability alone [18,19]. The confinement posterior $H_{98} = 1.007 \pm 0.030$ is tight and centred, so the uncertainty gate A_σ is near unity at the design point and falls only as the epistemic term of equation (10) grows off-distribution. The separation of epistemic from aleatoric uncertainty [12,17] is what lets the gate throttle on ignorance while tolerating irreducible noise. The abstention companion [28] develops the handing-back-to-a-floor decision in full.

RELATION TO LEARNED FUSION CONTROL AND DISRUPTION PREDICTION

Data-driven disruption predictors demonstrate that learned warnings can be accurate and transfer across machines [5,6,7], and reinforcement-learning controllers now shape and stabilise real plasmas [1,2]. Those results make the *capability* case; the epistemic-gating layer makes the *admissibility* case that must sit beneath them, and it is the natural home for the cross-machine transfer uncertainty a disruption model inherits when it is deployed off its training devices. The bound-limited Bayesian-optimization excursion of §2.4 ($Q \approx 8.3$ under widened priors) is a concrete instance of the confident off-distribution extrapolation the gate refuses: the optimizer is free to *propose* it, but the composite ceiling of equation (16) denies it authority because it lies outside the mapped envelope, the ledger has no support there, and the epistemic width has blown up. The maturity-gated actuation-authority argument [30] binds this runtime ceiling to the verification credibility of §3, and the global sensitivity-and-UQ survey of the frozen anchors [34] quantifies how the envelope's error bar propagates.

WHY THE MIN-RULE

The composite (16) is a minimum, not a product or an average, by design: authority should be capped by the *worst* evidence gate, so that a single failing channel — an unfamiliar regime, thin ledger support, or a blown-up epistemic width — is sufficient to throttle, and no gate can be "bought back" by another. This is the conservative choice a nuclear safety case requires, and it makes the gate's behaviour monotone and auditable: reducing any evidence source can only lower authority, never raise it.

§ 6

Limitations

One honest gate remains. The reconstruction detector's threshold family $\{\tau_O\}$ of equations (7)–(8) is calibrated against the twin's validated state distribution; tightening it against *machine* diagnostics is the one confirming step, and it is the natural first use of live data as the machine accrues operating hours. Until then the gate is conservative by construction: an unfamiliar state defaults to the floor, so the calibration gap can only make the controller more cautious, never less safe. The ledger support kernel (9) is likewise resolved from the nine-device post-diction base and will sharpen as operating cases accumulate.

§ 7

Conclusion

The Kronos autonomous controller carries an explicit epistemic-gating layer, stated here as mathematics and grounded in reproduced evidence. The validated envelope is *mapped* over a nine-point design-closure grid (Q 2.735–3.460, $P_{\rm fus}$ 58.6–118.9 MW) with the frozen design point $(3.076,\ 85.04\,\mathrm{MW})$ interior; two complementary detectors — a reconstruction/feature OOD score (SH-040) and a nine-device extrapolation ledger spanning the breeder and burner (KX-L1-A7; monitor SH-075) — measure how far the live point has moved outside it; the guarded confinement model carries a calibrated posterior $H_{98} = 1.007 \pm 0.030$ with a distribution-free conformal coverage guarantee and passes $19/19$ formal verification anchors; and a single composite min-rule throttles the learned controller smoothly to a certified control-barrier floor as the point approaches, and then leaves, the envelope. This is the credibility layer beneath the autonomy — authority that knows, and respects, the limit of its own evidence — and it converts the safety case from “trust the model everywhere” into “bound the model by what has been verified.”

Acknowledgments tocsectionAcknowledgments Part of the Kronos Fusion Energy 2026 design series. The author thanks the Kronos physics de-risking programme for the frozen anchor set underlying every quantitative claim, and the NVIDIA GPU HPC campaign for the accelerated closure, training and uncertainty-quantification runs.

Reproducibility. The gating anchors are frozen entries in the Kronos Physics De-Risking Register [27] (doi:10.5281/zenodo.22133057): the validated envelope (KX-L1-A1); the nine-device extrapolation ledger (KX-L1-A7); the calibrated confinement posterior $H_{98} = 1.007 \pm 0.030$ (KX-L1-A14); and the formal verification suite $19/19$ (KX-L1-A6). Each is regenerable from the archived closure, UQ and verification runs, and every figure is produced by the deposited generation script `make_ood_ledger_figs_2p79.py` from the frozen anchor values. This paper is archived at doi:10.5281/zenodo.22645819 and its official version is at https://www.kronosfusionenergy.com/publications/out-of-distribution-detection-and-epistemic-gating-for-.html.

Data availability tocsectionData availability The envelope grid, the ledger schema, the detector calibration inputs and the confinement posterior used for figures 1–8 are archived with the deposit at doi:10.5281/zenodo.22645819, which references the Kronos 2026 series — including the calibrated-uncertainty and abstention result [28], the runtime-assurance provenance ledger [29] and the AI-and-quantum-control overview [36]. All quantitative values trace to the register [27] 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 J Degrave *et al*, Magnetic control of tokamak plasmas through deep reinforcement learning, *Nature* **602** 414 (2022). doi:10.1038/s41586-021-04301-9

2 J Seo *et al*, Avoiding fusion plasma tearing instability with deep reinforcement learning, *Nature* **626** 746 (2024). doi:10.1038/s41586-024-07024-9

3 F Felici *et al*, Real-time physics-model-based simulation of the current density profile in tokamak plasmas, *Nucl. Fusion* **51** 083052 (2011). doi:10.1088/0029-5515/51/8/083052

4 J Abbate, R Conlin and E Kolemen, Data-driven profile prediction for DIII-D, *Nucl. Fusion* **61** 046027 (2021). doi:10.1088/1741-4326/abe08d

5 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

6 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

7 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

8 D Hendrycks and K Gimpel, A baseline for detecting misclassified and out-of-distribution examples in neural networks, *Int. Conf. on Learning Representations (ICLR)* (2017). arXiv:1610.02136

9 K Lee, K Lee, H Lee and J Shin, A simple unified framework for detecting out-of-distribution samples and adversarial attacks, *Adv. Neural Inf. Process. Syst. (NeurIPS)* **31** (2018). arXiv:1807.03888

10 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

11 Y Gal and Z Ghahramani, Dropout as a Bayesian approximation: representing model uncertainty in deep learning, *Int. Conf. on Machine Learning (ICML)* **48** 1050 (2016). arXiv:1506.02142

12 A Kendall and Y Gal, What uncertainties do we need in Bayesian deep learning for computer vision?, *Adv. Neural Inf. Process. Syst. (NeurIPS)* **30** (2017). arXiv:1703.04977

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

14 A Nguyen, J Yosinski and J Clune, Deep neural networks are easily fooled: high-confidence predictions for unrecognizable images, *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* 427 (2015). doi:10.1109/CVPR.2015.7298640

15 L Ruff *et al*, A unifying review of deep and shallow anomaly detection, *Proc. IEEE* **109** 756 (2021). doi:10.1109/JPROC.2021.3052449

16 M Sakurada and T Yairi, Anomaly detection using autoencoders with nonlinear dimensionality reduction, *Proc. MLSDA 2nd Workshop on Machine Learning for Sensory Data Analysis* 4 (2014). doi:10.1145/2689746.2689747

17 L V Jospin, H Laga, F Boussaid, W Buntine and M Bennamoun, Hands-on Bayesian neural networks — a tutorial for deep learning users, *IEEE Comput. Intell. Mag.* **17**(2) 29 (2022). doi:10.1109/MCI.2022.3155327

18 V Vovk, A Gammerman and G Shafer, *Algorithmic Learning in a Random World*, Springer (2005). doi:10.1007/b106715

19 A N Angelopoulos and S Bates, Conformal prediction: a gentle introduction, *Found. Trends Mach. Learn.* **16**(4) 494 (2023). doi:10.1561/2200000101

20 B Shahriari, K Swersky, Z Wang, R P Adams and N de Freitas, Taking the human out of the loop: a review of Bayesian optimization, *Proc. IEEE* **104** 148 (2016). doi:10.1109/JPROC.2015.2494218

21 J Snoek, H Larochelle and R P Adams, Practical Bayesian optimization of machine learning algorithms, *Adv. Neural Inf. Process. Syst. (NeurIPS)* **25** (2012). arXiv:1206.2944

22 L Lu, P Jin, G Pang, Z Zhang and G E Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nat. Mach. Intell.* **3** 218 (2021). doi:10.1038/s42256-021-00302-5

23 Z Li, N Kovachki, K Azizzadenesheli, B Liu, K Bhattacharya, A Stuart and A Anandkumar, Fourier neural operator for parametric partial differential equations, *Int. Conf. on Learning Representations (ICLR)* (2021). arXiv:2010.08895

24 A D Ames, S Coogan, M Egerstedt, G Notomista, K Sreenath and P Tabuada, Control barrier functions: theory and applications, *European Control Conf. (ECC)* 3420 (2019). doi:10.23919/ECC.2019.8796030

25 W L Oberkampf and C J Roy, *Verification and Validation in Scientific Computing*, Cambridge University Press (2010). doi:10.1017/CBO9780511760396

26 D Amodei, C Olah, J Steinhardt, P Christiano, J Schulman and D Man\'e, Concrete problems in AI safety (2016). arXiv:1606.06565

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

28 P I Ford, *Calibrated Uncertainty, Conformal Prediction, and Abstention for Safety-Critical Fusion Control: Handing Authority Back to a Deterministic Floor Out of Distribution*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645813. https://www.kronosfusionenergy.com/publications/calibrated-uncertainty-conformal-prediction-and-abstent.html

29 P I Ford, *Runtime Assurance with Signed-Ledger Provenance for Autonomous Fusion Control: A Zero-Trust, Hardware-Failsafe Last Line of Defense Beneath the AI*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645817. https://www.kronosfusionenergy.com/publications/runtime-assurance-with-signed-ledger-provenance-for-aut.html

30 P I Ford, *Maturity-Gated Actuation Authority: Binding PCMM Verification Credibility to Runtime Control Authority in an Autonomous Fusion Control Stack*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645795. https://www.kronosfusionenergy.com/publications/maturity-gated-actuation-authority.html

31 P I Ford, *A Certified Control-Barrier Safety Filter for Fusion Control: Forward-Invariance Guarantees, Calibrated Uncertainty, and an Honest PINN Negative for a Compact Spherical-Tokamak Breeder*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645716. https://www.kronosfusionenergy.com/publications/control-barrier-safety-clamp.html

32 P I Ford, *Four Coupled Digital Twins for a Compact Spherical-Tokamak Breeder: Multi-Rate Co-Simulation, Propagated Uncertainty, and a Real-Time Self-Consistency Monitor*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645805. https://www.kronosfusionenergy.com/publications/four-coupled-digital-twins-for-a-compact-spherical-toka.html

33 P I Ford, *Neural-Operator Flux Surrogates for Control-Latency Transport: Differentiable DeepONet-FNO Emulation of Nonlinear Gyrokinetics with Manifold-Abstention Monitoring*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645803. https://www.kronosfusionenergy.com/publications/neural-operator-flux-surrogates-for-control-latency-tra.html

34 P I Ford, *Global Sensitivity and Uncertainty Quantification across the Breeder and Burner Design Planes: A 40,000-Sample Monte-Carlo Ranking of the Frozen Anchors*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645893. https://www.kronosfusionenergy.com/publications/global-sensitivity-and-uncertainty-quantification-acros.html

35 P I Ford, *Verification-First Fusion Modeling: Frozen Anchors, Independent-Library Gate Reproduction, and a Byte-for-Byte Computational-Provenance Catalog for a Breeder-and-Burner Design Register*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645710. https://www.kronosfusionenergy.com/publications/verification-and-validation-47-code-provenance.html

36 P I Ford, *AI and Quantum Computing for Fusion: ML Surrogates, Real-Time Control, and a Resource-Honest Quantum Frontier*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645702. https://www.kronosfusionenergy.com/publications/ai-quantum-control.html

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

Out-of-Distribution Detection and Epistemic Gating for Autonomous Fusion Control: A Machine-Readable Extrapolation Ledger That Throttles Model Authority Beyond the Validated Envelope

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

Zenodo · doi:10.5281/zenodo.22645819

Read on kronosfusionenergy.com

The Physics De-Risking Register