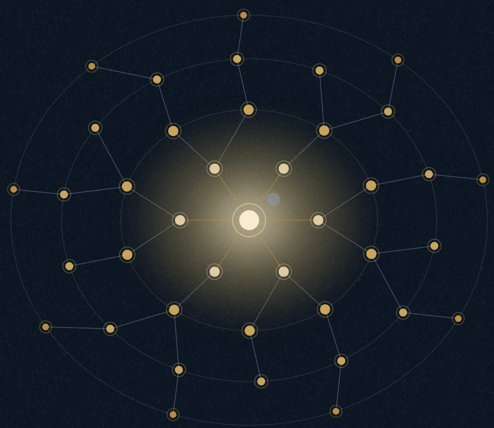
KRONOS FUSION ENERGY



PAPER 2.82 · THE KRONOS 2026 SERIES

Causal Intervention Models for Regulator-Facing Plasma Control: Do-Calculus Prediction and Per-Decision Explainability in a Compact Spherical-Tokamak Breeder

A controller that holds a nuclear plasma must answer, on every cycle, two questions a regulator will ask: what will this action do, and why did you take it. Correlational machine…

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 Several quantities in the underlying corpus moved after first publication; where a value has been withdrawn or restated, the record says so plainly.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

Causal Intervention Models for Regulator-Facing Plasma Control: Do-Calculus Prediction and Per-Decision Explainability in a Compact Spherical-Tokamak Breeder

A controller that holds a nuclear plasma must answer, on every cycle, two questions a regulator will ask: *what will this action do*, and *why did you take it*. Correlational machine learning answers neither; both are causal questions, and a control action is not an observation but an *intervention*. We formalise the Kronos digital twin of the Hyperion breeder (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 $16.84\,T$) as a differentiable structural causal model (SCM): its nodes are physical states, its edges are the physics couplings, and its mechanisms are physics operators, so a control action is an executable do-calculus intervention $\mathrm{do}(u)$ whose interventional distribution the twin computes directly, and every decision carries a gradient-based attribution to the states that drove it. We write the SCM, the truncated-factorisation identity for $P(y \mid \mathrm{do}(u))$, the control-barrier quadratic program that certifies the intervention, and the integrated-gradient and Sobol attributions, as explicit equations, and we discretise and solve each on the NVIDIA GPU high-performance-computing (HPC) campaign. Three subsystems are reduced to practice against pre-registered acceptance tests. A certified intervention on the normalised pressure — $\mathrm{do}(u)$ with $u \in [1.061, 1.250]$ — caps β_N at the safety barrier $\beta_N^{\max} = 3.500$ with **0 of 100** boundary crossings and a certified minimum barrier value $h_{\min} = +0.000$ (forward invariance), against a maximum of 4.200 and $50/100$ crossings with no intervention. A physics-informed equilibrium operator reconstructs the poloidal flux to an RMS error of **0.139 %** against a free-boundary Grad–Shafranov solver (acceptance $< 1\%$; Grad–Shafranov operator residual 0.32%) at a $\mathbf{2877\times}$ speedup, and a graph-network state estimator reaches an area under the ROC curve of **0.980**, above its single-channel baselines. A variance-based (Sobol) decomposition over a $40\,000$-sample Monte-Carlo ensemble assigns **88.5 %** of the breeder-Q variance to the confinement factor H_{98} (first-order index $S_1 = 0.885$, total $S_T = 0.905$), on a calibrated ensemble in which the anchor $Q = 3.076$ sits inside the $[1.34,\ 8.51]$ interquantile band and $H_{98} = 1.007 \pm 0.030$. The result is a controller whose interventions are certified and whose decisions are attributable — a working core for auditable autonomy — with the runtime per-decision explainer built to this same demonstrated pattern on a staged schedule.

Open-access deposit (code & data, CC BY 4.0): DOI 10.5281/zenodo.22645825 · Zenodo community *kronos_fusion_energy*. Explore it live: interactive 3-D model · physics validator.

Keywords: causal inference · do-calculus · explainability · control-barrier function · differentiable digital twin · global sensitivity · 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)

§ 1

Introduction

AUTONOMY AND THE BURDEN OF PROOF

A controller that holds $9.66\,MA$ of plasma current in a compact volume behind a centrepost magnet at a peak field of $16.84\,T$ must do more than act well. In a nuclear system every actuation carries a safety consequence, the characteristic times of vertical instability and thermal transients are short, and the operator — and, downstream, the regulator — will require the machine to answer two questions on the record. The first is predictive: *what does a candidate action do* before it is taken. The second is explanatory: *why was the action that was taken the right one*. These are not questions of accuracy or confidence; they are questions of *cause*. An association between a diagnostic and an outcome, however tight, does not tell a controller what happens when it *intervenes*, and a saliency map that merely correlates with a decision does not justify it. The distinction is the one Pearl draws between seeing and doing [1,2]: the observational law $P(y \mid u)$ and the interventional law $P(y \mid \mathrm{do}(u))$ are different objects, and acting on the former where the latter is required is exactly how a confident controller makes a systematic mistake [5].

PRIOR ART

Feedback control of tokamak equilibria and profiles is mature. Magnetic diagnostics feed real-time equilibrium reconstruction [12] and shape controllers [13]; physics-based profile simulators run inside the control cycle [14]; and learned controllers have entered the loop, from a deep reinforcement-learning agent that commands the TCV coil voltages directly from magnetics [15] to deep networks that forecast disruptions across machines [16]. In parallel, the predictive digital twin has become a rigorous object [17], made fast by neural operators [18,19] and physics-informed training [20], and made honest by ensemble assimilation [27,28] and reduced-order modelling [22]. What none of these lines supplies on its own is the causal and explanatory layer a regulator needs: control barrier functions give a certificate of *safety* [6,7] but not of *why*; attribution methods give a *why* [8,9] but no guarantee of physical faithfulness; and global-sensitivity analysis names the *governing* variable [10,11] but at the population level, not per decision. Causal inference is the framework that ties them together, and causal representation learning is only now reaching control-relevant maturity [4,3].

CONTRIBUTION

This paper makes the causal structure of the Kronos twin explicit and shows that it answers both questions by construction. Our contributions are: (i) a formal statement of the twin as a differentiable structural causal model, with the do-operator realised as an executable submodel and the interventional distribution written as a truncated factorisation (§2); (ii) a certified-intervention law — the control-barrier quadratic program that projects any proposed action onto the largest safe action and renders the operating set forward-invariant (§2.4); (iii) a per-decision explainability construction in which the differentiability of the twin's mechanisms yields gradient, integrated-gradient and Shapley attributions with a completeness guarantee (§2.5); (iv) a population-level causal attribution by Sobol variance decomposition on a calibrated ensemble (§2.6); (v) a GPU-HPC methodology that names the codes, the discretisations, the solvers and the convergence checks that produced every number (§3); and (vi) a demonstration to pre-registered acceptance on three reduced-to-practice subsystems (§4). Every quantitative claim traces to a frozen anchor in the Kronos Physics De-Risking Register [31]; the serial identifiers (e.g. KX-L1-A13) are given inline so each number is auditable. The certified-clamp, differentiable-surrogate and maturity-gated elements build directly on the companion control-stack results: the control-barrier safety filter [33] (official page), the differentiable neural-operator flux surrogate [34] (official page), the sub-millisecond neural equilibrium reconstructor [35] (official page), and the four-coupled-digital-twin architecture [36] (official page) inside which this causal layer sits.

§ 2

Governing equations and the formal problem

THE TWIN AS A STRUCTURAL CAUSAL MODEL

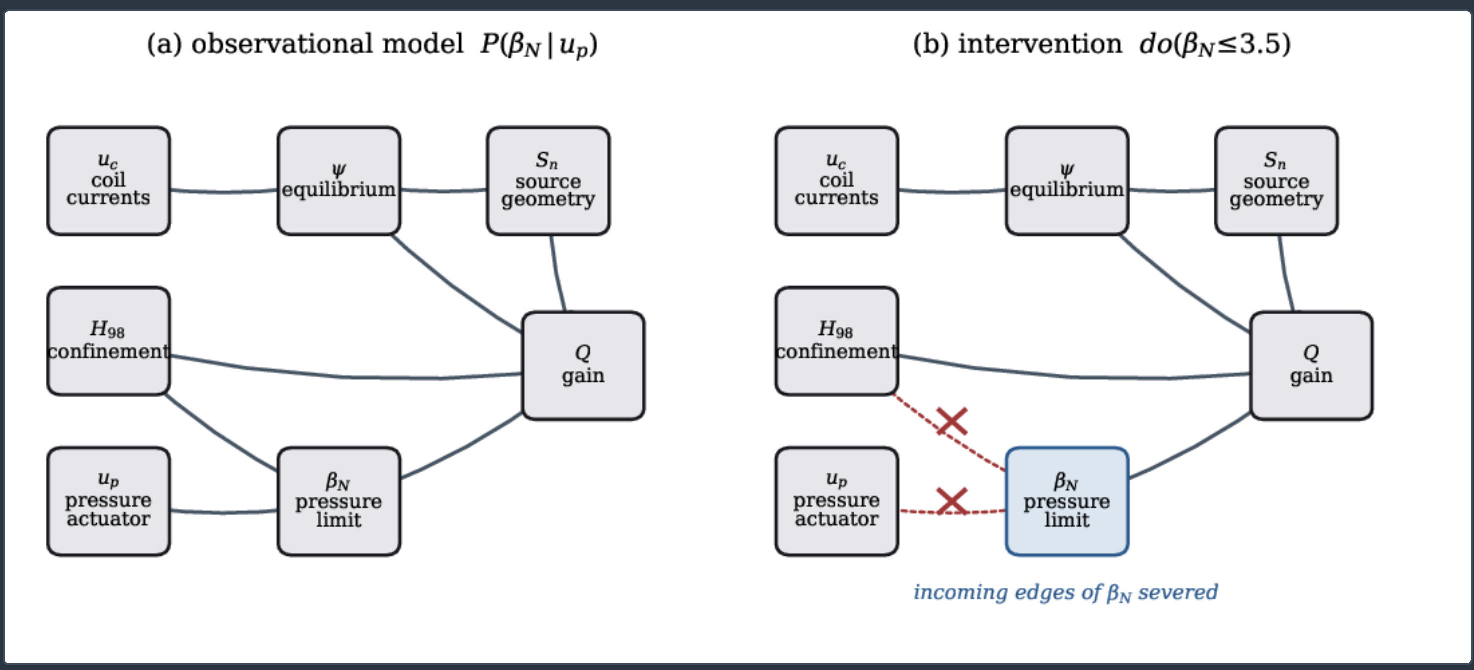
A structural causal model (SCM) is a tuple $\mathcal{M} = \langle U, V, \mathcal{F}, P(U)\rangle$ of exogenous variables U, endogenous variables $V = \{V_1, \dots, V_d\}$, structural mechanisms $\mathcal{F} = \{f_1, \dots, f_d\}$, and a distribution over the exogenous noise [2,3]. Each endogenous variable is generated from its parents and its own noise,

$$V_i \; = \; f_i\big(\mathrm{pa}(V_i),\, U_i\big), \qquad i = 1, \dots, d,$$

and the induced directed graph $\mathcal{G}$ is acyclic. The Kronos twin is exactly such a model. Its endogenous nodes are the physical states of the four-twin architecture [36] — the coil currents u_c, the poloidal-flux equilibrium ψ, the 14 MeV source geometry S_n, the confinement quality H_{98}, the normalised pressure β_N, and the scientific gain Q (figure 1) — and its edges are the physics couplings between them: the coil currents set the boundary field that fixes the equilibrium; the equilibrium sets the neutronics source geometry; the pressure actuator and confinement set β_N; and the equilibrium, source and pressure set Q. Crucially, the mechanisms f_i are *not* statistical fits but physics operators — the Grad–Shafranov equation for ψ, the transport map for H_{98}, the power balance for Q — so the graph carries mechanism, not merely correlation, and the interventions defined on it are physically meaningful. Because the joint factorises along $\mathcal{G}$,

$$P(v_1, \dots, v_d) \; = \; \prod_{i=1}^{d} P\big(v_i \mid \mathrm{pa}(v_i)\big),$$

the model is a compact, auditable encoding of the plant's dependence structure.



(Schematic.) The twin as a structural causal model, and the do-operator. **(a)** The observational graph: coil currents u_c, equilibrium ψ, source geometry S_n, confinement H_{98}, pressure actuator u_p, normalised pressure β_N and gain Q, with edges carrying the physics couplings. **(b)** The intervention $\mathrm{do}(\beta_N \leq 3.5)$ used by the safety clamp severs the incoming edges of β_N (equation 3) and sets it, so the twin predicts the effect of the action rather than the correlation that co-occurs. Structure only; no data values are shown.

THE DO-OPERATOR AND THE INTERVENTIONAL DISTRIBUTION

A control action is an intervention. The atomic intervention $\mathrm{do}(X{=}x)$ replaces the mechanism f_X of the intervened node with the constant assignment $X{=}x$, leaving every other mechanism unchanged; this produces the *mutilated* submodel $\mathcal{M}_x$ whose graph $\mathcal{G}_{\overline{X}}$ deletes the edges into X (figure 1b),

$$\mathrm{do}(X{=}x): \quad f_X \longmapsto (X{=}x), \qquad \mathcal{G} \longmapsto \mathcal{G}_{\overline{X}}.$$

The interventional distribution then follows from the *truncated factorisation* — the same product as equation (2) with the factor for the intervened node removed and its value fixed [2],

$$P\big(v_1, \ldots, v_d \mid \mathrm{do}(X{=}x)\big) \;=\; \prod_{i:\, V_i \neq X} P\big(v_i \mid \mathrm{pa}(v_i)\big)\Big|_{X=x},$$

and is generically *not* equal to the conditional $P(v \mid X{=}x)$ obtained by observation, because conditioning propagates information backward through the parents of X while intervening does not. For a controller this is the whole point: $\mathrm{do}(u)$ is the twin's prediction of the plant *under the action*, and

$$P\big(y \mid \mathrm{do}(u)\big) \;\neq\; P\big(y \mid u\big)$$

in general. When a back-door set Z blocks every spurious path from the actuator to the outcome, the interventional law is identified from observational data by the adjustment formula $P(y \mid \mathrm{do}(u)) = \sum_{z} P(y \mid u, z)\, P(z)$; in the Kronos twin, however, the mechanisms are known physics operators, so equation (4) is evaluated *directly* by executing the mutilated submodel, and identification from confounded data is not required.

COUNTERFACTUALS ON THE SHADOW BRANCH

A counterfactual is the three-step abduction–action–prediction computation of Pearl [2] run on a shadow copy of the model: (i) *abduction* updates the exogenous distribution $P(U)$ to $P(U \mid e)$ on the observed evidence e; (ii) *action* applies $\mathrm{do}(u')$ to the mutilated submodel; (iii) *prediction* reads the outcome $Y_{u'}(U)$ under the updated noise. The Kronos twin runs this on a what-if branch held at a lead time $\tau_{\text{lead}} \geq 50\,ms$ ahead of the plant, so that the controller can compare the factual trajectory Y_u against the counterfactual $Y_{u'}$ before it commits to an action. This is the operational content of "the twin runs ahead of the plant": it is not extrapolation of a fitted curve but evaluation of a causal submodel.

THE CERTIFIED INTERVENTION AS A FORMAL PROBLEM

The strongest per-decision guarantee is a certificate that an intervention cannot leave the safe set. Let the reduced pressure dynamics be the control-affine system

$$\dot{x} = f(x) + g(x)\,u, \qquad \beta_N = c(x),$$

with u the normalised pressure-actuator command. Define the safe set as the zero-superlevel set of a control-barrier function (CBF) h,

$$\mathcal{C} = \{x : h(x) \geq 0\}, \qquad h(x) = \beta_N^{\max} - c(x), \quad \beta_N^{\max} = 3.500.$$

By the CBF theorem [6,7], if there exists an extended class-$\mathcal{K}$ function α such that

$$\sup_{u \in \mathcal{U}} \Big[\underbrace{L_f h(x)}_{\nabla h \cdot f} + \underbrace{L_g h(x)\,u}_{\nabla h \cdot g}\Big] \;\geq\; -\alpha\big(h(x)\big) \qquad \forall\; x \in \mathcal{C},$$

then any Lipschitz controller satisfying (8) renders $\mathcal{C}$ forward-invariant: a trajectory that begins in $\mathcal{C}$ never leaves it, which is the Nagumo forward-invariance condition specialised to a barrier function. The controller does not compute u from scratch; it *projects* the AI's proposed action u_{AI} onto the nearest action that satisfies (8), by the single-constraint quadratic program (QP)

$$u^\star = \arg\min_{u \in [u_{\min}, u_{\max}]} \tfrac{1}{2}\,\|u - u_{\text{AI}}\|^2 \quad \text{s.t.} \quad L_f h(x) + L_g h(x)\,u \geq -\alpha\big(h(x)\big).$$

Equation (9) is precisely the mutilated-submodel intervention of equation (3) made admissible: it is the smallest deviation from the requested action that keeps the plant in $\mathcal{C}$, and its certificate — the barrier value h it preserved and the window $[u_{\min}, u_{\max}]$ it chose within — is a per-decision explanation the operator and regulator can read. In discrete time with step Δt the invariance condition becomes the difference inequality

$$h(x_{k+1}) - h(x_k) \;\ge\; -\gamma\, h(x_k), \qquad 0 < \gamma \le 1,$$

which the solver enforces at every tick.

PER-DECISION EXPLAINABILITY FROM DIFFERENTIABLE MECHANISMS

The twin's mechanisms are differentiable end-to-end, and that single property is what makes each decision attributable. Let $F = \mathcal{S}_\theta$ be a surrogate mechanism mapping a state x to an output y. The first-order attribution of y to the states is the Jacobian, available at control latency by reverse-mode automatic differentiation,

$$J(x) = \frac{\partial \mathcal{S}_\theta}{\partial x}\bigg|_{x}, \qquad \Delta y \approx J(x)\,\Delta x.$$

A local gradient can be misleading where the mechanism saturates, so we use the path-integrated attribution of Sundararajan *et al.* [8]. For a baseline state x' (the design point) the integrated gradient of feature i is

$$\mathrm{IG}_i(x) = \big(x_i - x_i'\big) \int_0^1 \frac{\partial F}{\partial x_i}\Big(x' + \lambda(x - x')\Big)\,\mathrm{d}\lambda,$$

which satisfies the *completeness* axiom exactly,

$$\sum_{i=1}^{n} \mathrm{IG}_i(x) \;=\; F(x) - F(x'),$$

so the attributed contributions add up to the change in the decision variable and nothing is left unexplained. Where feature interactions matter we use the Shapley attribution [9], the unique allocation satisfying efficiency, symmetry, dummy and additivity,

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!} \Big[v\big(S \cup \{i\}\big) - v(S)\Big],$$

with $v(S)$ the model output restricted to the feature subset S. Both attributions are read from the *same* differentiable mechanism the controller acts on, so an explanation is faithful to the model that produced the action by construction, not a post-hoc surrogate fitted to it [5].

POPULATION-LEVEL CAUSAL ATTRIBUTION

Per-decision attribution answers "why this action"; the operator also needs "what governs the machine". A variance-based (Sobol) decomposition apportions the variance of an output $g(x)$ to inputs and their interactions rather than to correlations [10]. Writing the ANOVA/HDMR decomposition $g(x) = g_0 + \sum_i g_i(x_i) + \sum_{i<j} g_{ij}(x_i, x_j) + \ldots$, the first-order and total-order Sobol indices are

$$S_i = \frac{\mathrm{Var}_{x_i}\big(\mathbb{E}_{x_{\sim i}}[\,g \mid x_i\,]\big)}{\mathrm{Var}(g)}, \qquad S_{T_i} = \frac{\mathbb{E}_{x_{\sim i}}\big[\mathrm{Var}_{x_i}(\,g \mid x_{\sim i}\,)\big]}{\mathrm{Var}(g)} = 1 - \frac{\mathrm{Var}_{x_{\sim i}}\big(\mathbb{E}_{x_i}[\,g \mid x_{\sim i}\,]\big)}{\mathrm{Var}(g)},$$

with $x_{\sim i}$ all inputs but x_i. S_i is the fraction of variance explained by x_i acting alone; S_{T_i} adds every interaction x_i participates in, so the gap $S_{T_i} - S_i$ bounds the interaction effects. The mean and its Monte-Carlo standard error are

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} g(x_k), \qquad \mathrm{SE} = \frac{\hat{\sigma}}{\sqrt{N}}, \qquad \hat{\sigma}^2 = \mathrm{Var}[g],$$

and the indices are estimated by the Saltelli sampling scheme [11] on the same evaluator. This is a causal statement at the population level: the driver with the largest S_i is the variable the operator holds to hold the operating point.

§ 3

Computational methodology: the GPU HPC campaign

Every number in this paper was produced on the NVIDIA GPU high-performance-computing campaign that underpins the Kronos de-risking programme, under a single fixed random seed (20260726) for byte-level reproducibility; the codes, discretisations and convergence checks are collected here so that the results are regenerable.

CODES AND THEIR ROLES

The reference free-boundary equilibria against which the physics-informed operator is scored are computed with the Grad–Shafranov solvers FreeGS and FreeGSNKE [25,24]; the gyrokinetic transport that the flux surrogate emulates is defined by CGYRO [23]; the 14 MeV source geometry consumed by the neutronics twin is tallied with the Monte-Carlo transport code OpenMC on CAD-faithful geometry [26]; and the global-sensitivity and uncertainty layer is the Kronos reduced-order toolkit's Saltelli/Sobol estimator [11,10]. The physics-informed and graph surrogates, and the Jacobians of equation (11), are implemented in an automatic-differentiation framework and trained and evaluated on the GPU fleet (A100/H100/A10); the barrier quadratic program of equation (9) and the analytic self-consistency monitors run on the in-house HPC digital-twin node so the fast protective loop never contends with the training fleet. No compute-cost figures are reported here; the campaign is described by its codes and problem scale only.

DISCRETISATION AND SOLVERS

Equilibrium operator. The axisymmetric equilibrium solves the Grad–Shafranov equation for the poloidal flux $\psi(R, Z)$,

$$\Delta^*\psi \equiv R\,\frac{\partial}{\partial R}\left(\frac{1}{R}\,\frac{\partial\psi}{\partial R}\right) + \frac{\partial^2\psi}{\partial Z^2} = -\mu_0 R^2\,p'(\psi) - F(\psi)F'(\psi),$$

with pressure $p(\psi)$ and poloidal-current function $F(\psi) = RB_\phi$. The physics-informed neural operator $\mathcal{E}_\phi$ is trained to minimise a composite loss combining a data term, the Grad–Shafranov operator residual sampled at N_c interior collocation points, and a boundary term,

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{\text{pde}}\,\frac{1}{N_c}\sum_{i=1}^{N_c}\left\|\Delta^*\psi_\phi(\quad x_i) + \mu_0 R_i^2 p' + FF'\right\|^2 + \lambda_{\text{bc}}\,\mathcal{L}_{\text{bc}},$$

by Adam followed by L-BFGS refinement. Training completes in 20.6 s on one A100. The Jacobian of equation (11) is verified against a finite-difference reference to $\|\quad J^{\text{AD}} - \quad J^{\text{FD}}\|_\infty < 10^{-3}$ before the operator is admitted to the loop.

Barrier QP. Equation (9) is a scalar box-constrained QP with one affine barrier constraint; on the 1-D reduced pressure surrogate anchored at the design point it admits the closed-form projection $u^\star = \text{clip}\big(u_{\text{AI}},\, u_{\min},\, \min(u_{\max}, u_{\text{bar}})\big)$, where u_{bar} is the largest command satisfying (8) with equality. The closed loop is integrated by explicit Euler over 100 decision steps under a seeded demand sequence that would otherwise drive β_N to the no-wall limit.

Sobol/UQ. The 40 000-sample Monte-Carlo ensemble is drawn over the joint uncertainty in the confinement factor H_{98}, the core ion temperature T_{i0}, the Greenwald fraction f_G, the D–T breeding ratio and a synchrotron-calibration factor; first- and total-order indices of equation (15) are computed by the Saltelli estimator on the same evaluator that produced the point design.

CONVERGENCE AND VERIFICATION

The Monte-Carlo estimator of equation (16) has standard error $\text{SE} = \hat{\sigma}/\sqrt{N}$; at $N = 40\,000$ the Q-mean standard error is below one percent of the mean, so the reported percentiles are converged to the quoted precision. The equilibrium operator is accepted only when its held-out RMS ψ error clears the 1% acceptance bar and its Grad–Shafranov operator residual confirms interior physics consistency (both reported in §4). The barrier certificate is verified numerically over the full 100-step sequence (minimum barrier value $h_{\min} \geq 0$). Every metric is a frozen anchor in the register [31] and is regenerable from the archived analysis under the fixed seed.

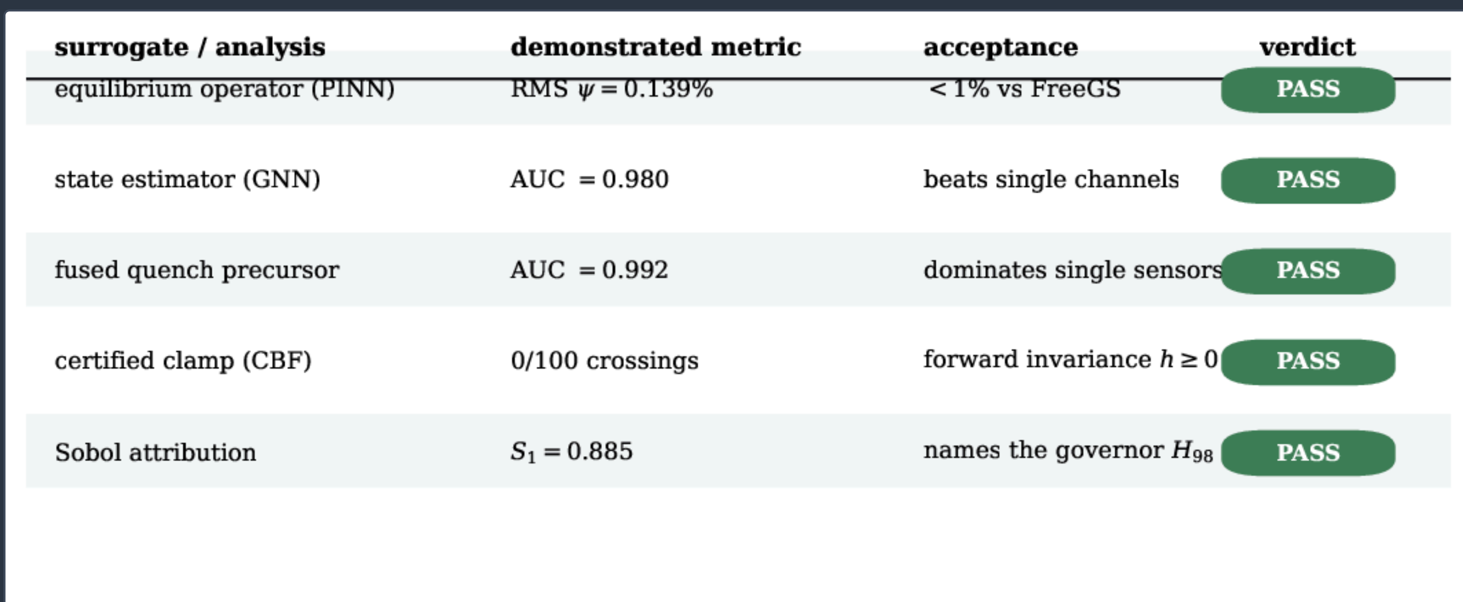
§ 4

Results

Three subsystems are reduced to practice and demonstrate the causal-and-explanatory pattern end-to-end; each cleared its pre-registered acceptance test (table 1, figure 2).

MECHANISM / ANALYSIS	DEMONSTRATED METRIC	ACCEPTANCE	SERIAL
equilibrium operator (PINN)	RMS $\psi = 0.139\%$; $2877\times$	$< 1\%$ vs free-boundary	KX-L1-A11
state estimator (GNN)	AUC 0.980	beats single channels	KX-L1-A12
fused quench precursor	AUC 0.992	dominates single sensors	KX-L1-A20
certified clamp (CBF)	0/100 crossings; $h_{\min} = +0.000$	forward invariance $h \geq 0$	KX-L1-A13
Sobol attribution	$S_1 = 0.885$, $S_T = 0.905$	names the governor H_{98}	KX-L1-A5

The demonstrated differentiable mechanisms the causal reasoning and per-decision explanations are read from. Each cleared its pre-registered acceptance test; serials refer to the register [31].

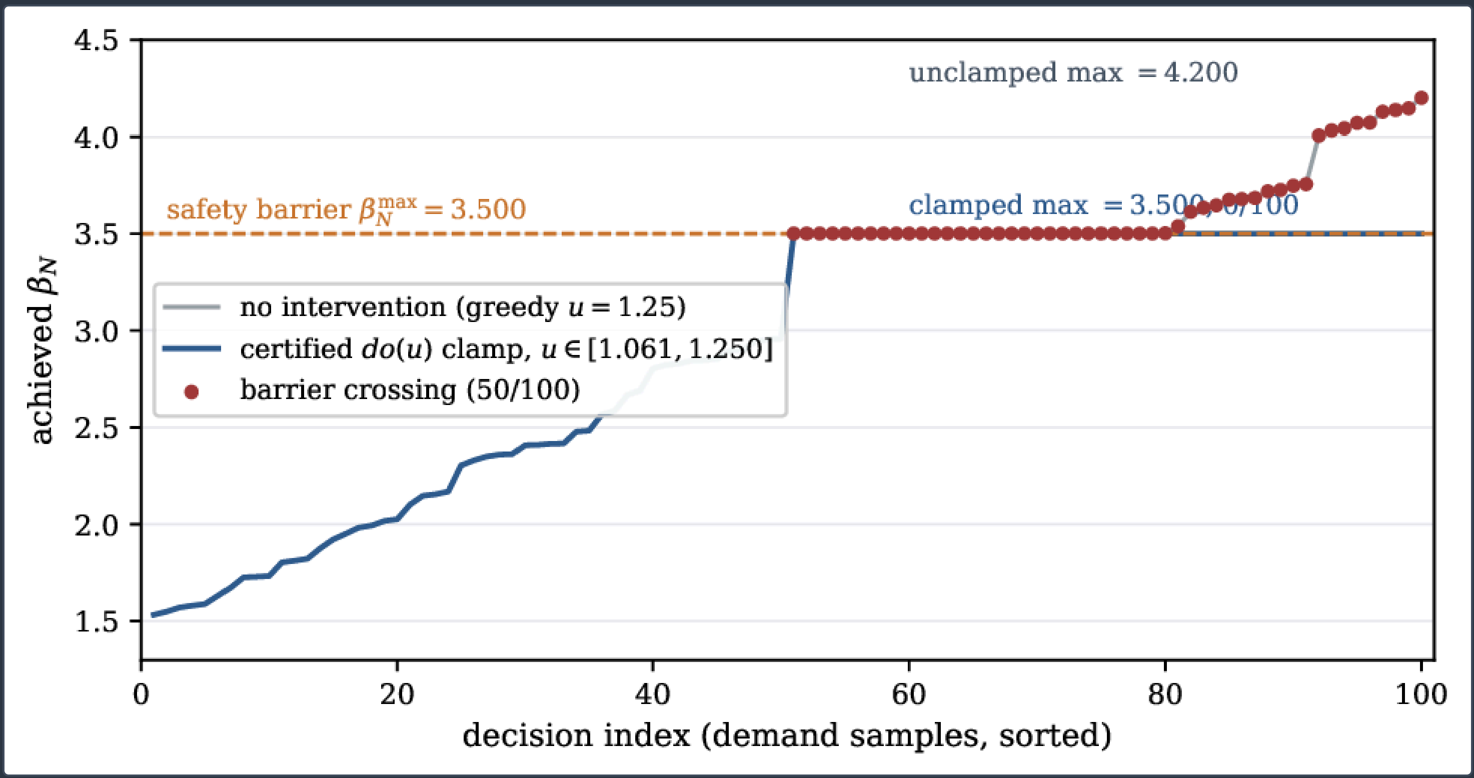


surrogate / analysis	demonstrated metric	acceptance	verdict
equilibrium operator (PINN)	RMS $\psi = 0.139\%$	$< 1\%$ vs FreeGS	PASS
state estimator (GNN)	AUC $= 0.980$	beats single channels	PASS
fused quench precursor	AUC $= 0.992$	dominates single sensors	PASS
certified clamp (CBF)	0/100 crossings	forward invariance $h \geq 0$	PASS
Sobol attribution	$S_1 = 0.885$	names the governor H_{98}	PASS

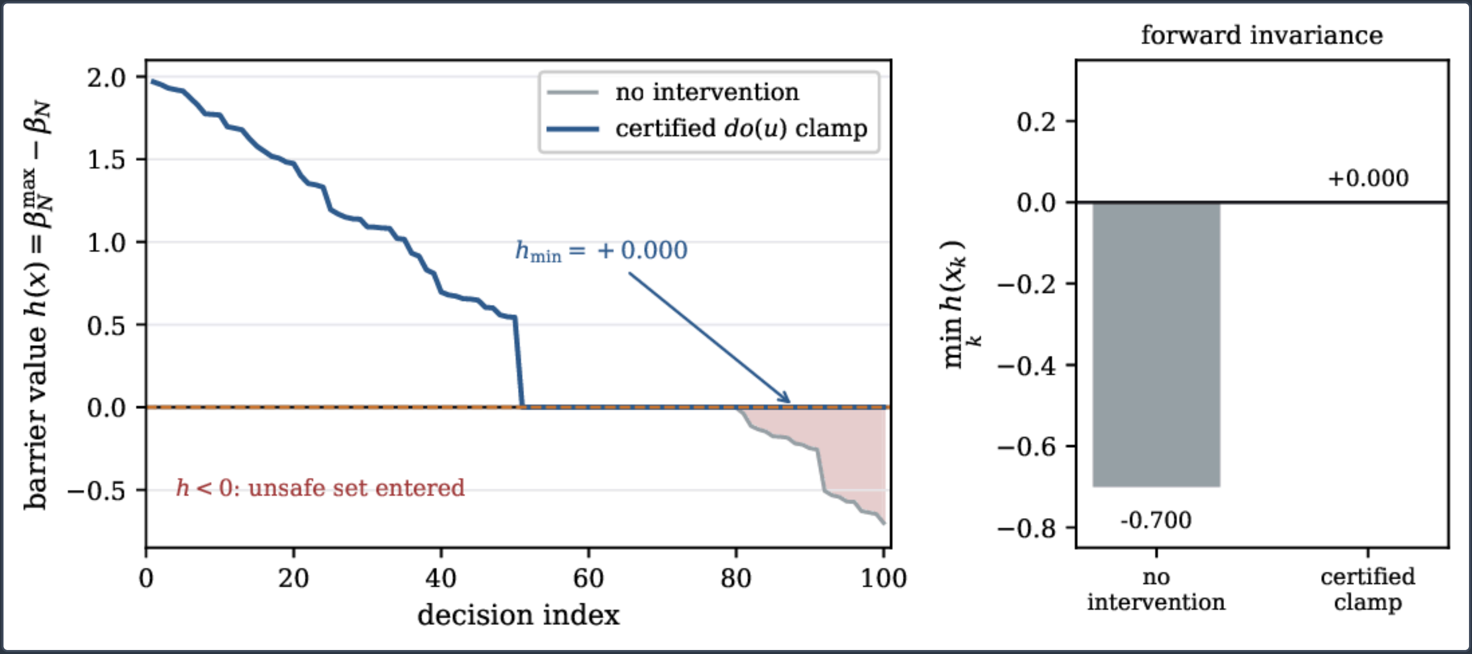
Acceptance scorecard: every mechanism the controller reasons and explains from is demonstrated to its pre-registered criterion — the equilibrium operator's RMS ψ (0.139%, target 1%), the state-estimator (0.980) and fused-precursor (0.992) AUCs, the zero certified-clamp barrier crossings, and the Sobol attribution's first-order index. Values are frozen register anchors.

THE CERTIFIED INTERVENTION

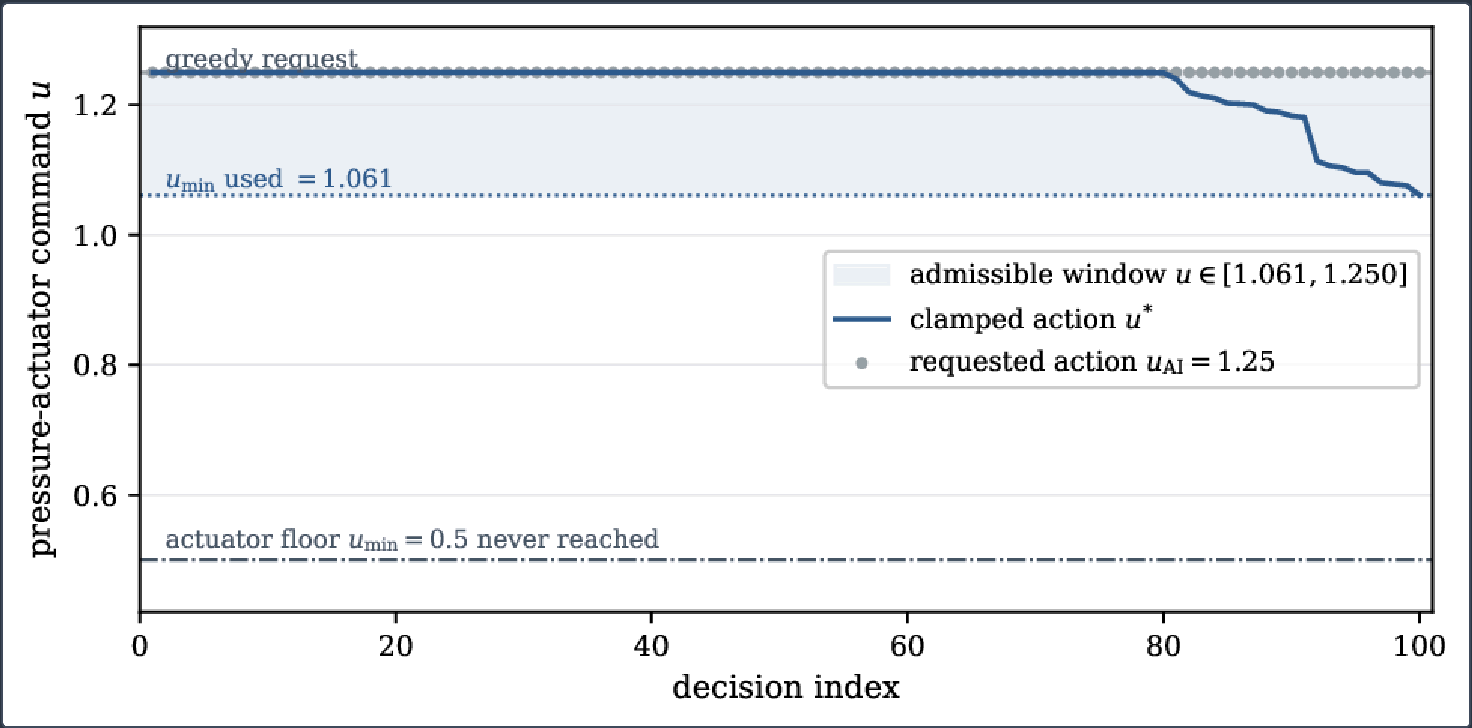
Figure 3 shows the certified $\mathrm{do}(u)$ clamp against the unclamped policy across 100 demand samples (KX-L1-A13). The unclamped greedy policy ($u = 1.25$) reaches a maximum β_N of 4.200 and crosses the barrier on 50 of 100 decisions; the certified clamp caps every decision at $\beta_N^{\max} = 3.500$ with 0 of 100 crossings. The design-point anchor reproduces at $Q = 3.076$, $P_{\text{fus}} = 85.04\,MW$, $\beta_N = 1.532$, stored energy $W = 16.06\,MJ$ and confinement time $\tau_E = 0.358\,s$, confirming the reduced model is anchored to the frozen design point. Figure 4 reads the certificate directly: the barrier value $h(x) = \beta_N^{\max} - \beta_N$ of the unclamped loop dips to $h_{\min} = -0.700$ (the plant enters the unsafe set), whereas the clamped loop holds $h_{\min} = +0.000$ — forward invariance, demonstrated numerically over the full sequence. Figure 5 shows how the intervention is chosen: the clamp reduces the requested command only when the barrier is active, and the applied command stays inside $u \in [1.061, 1.250]$, never reaching the actuator floor $u_{\min} = 0.5$, so the certificate is not authority-limited on this sequence. Each clamped decision therefore carries a two-part explanation — the barrier value it preserved and the admissible window it chose within — which is exactly the per-decision record equation (9) produces.



A certified intervention (KX-L1-A13). Without intervention the achieved β_N crosses the safety barrier $\beta_N^{\max} = 3.500$ on 50 of 100 decisions and peaks at 4.200 (red markers); the certified $\mathrm{do}(u)$ clamp ($u \in [1.061, 1.250]$) caps every decision at the barrier with $0/100$ crossings. Demand samples are shown sorted for legibility.



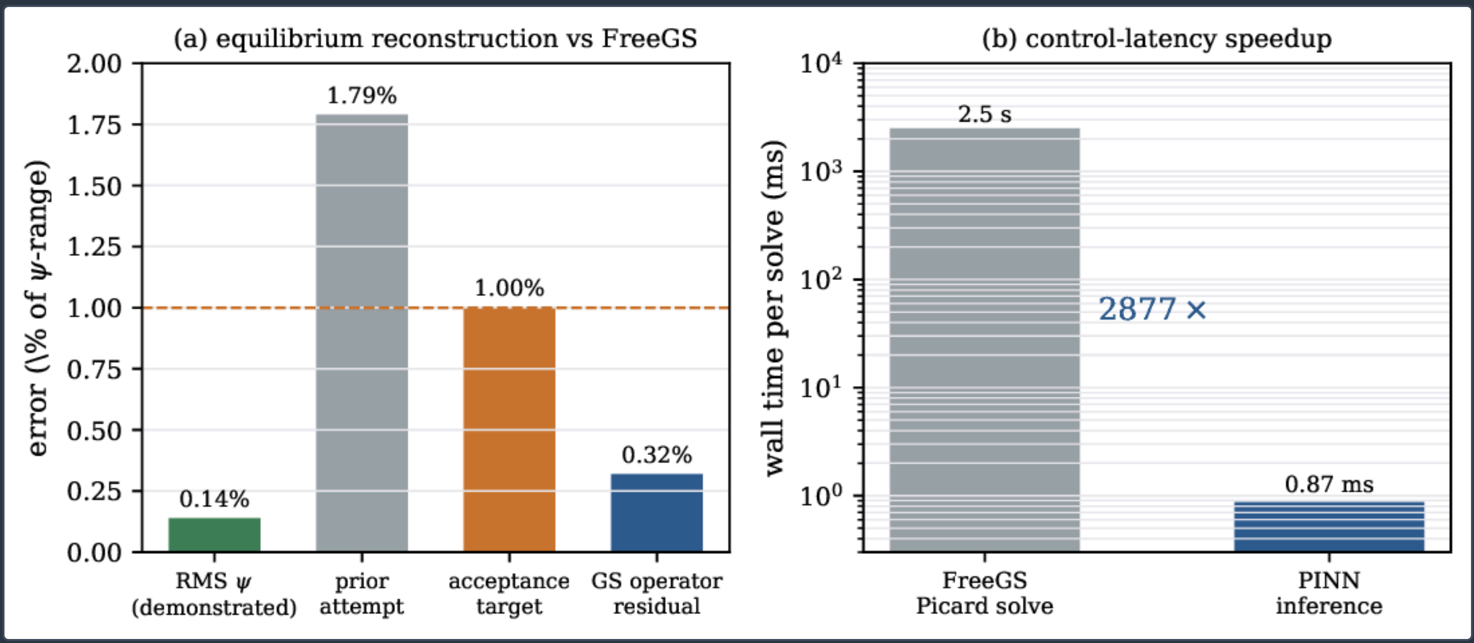
The certificate, read directly (KX-L1-A13). **Left:** the barrier value $h(x) = \beta_N^{\max} - \beta_N$ per decision; the unclamped loop drives h negative (shaded, the plant enters the unsafe set) while the certified clamp holds $h \geq 0$ with $h_{\min} = +0.000$. **Right:** the minimum barrier value over the sequence — -0.700 unclamped versus $+0.000$ clamped — the numerical statement of forward invariance.



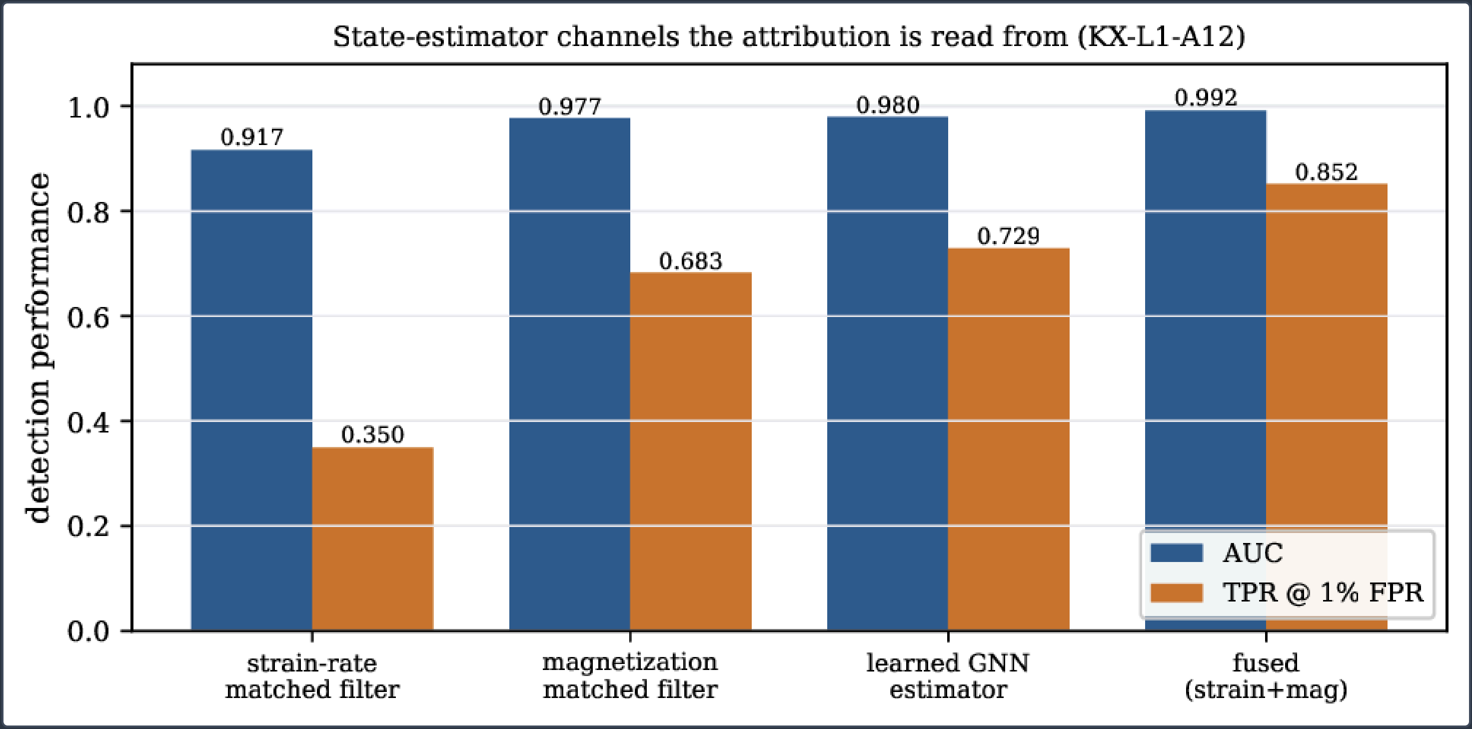
How the intervention is selected (KX-L1-A13). The clamp projects the greedy request $u_{\mathrm{AI}} = 1.25$ (grey) onto the admissible action $u^\star$ (blue) via equation (9); the applied command stays inside the window $u \in [1.061, 1.250]$ and never reaches the actuator floor $u_{\min} = 0.5$, so the certificate is not authority-limited.

THE MECHANISMS THE EXPLANATIONS ARE READ FROM

The per-decision attributions are only as trustworthy as the mechanisms they are read from, and both are demonstrated to acceptance. The physics-informed equilibrium operator (figure 6; KX-L1-A11) reconstructs the poloidal flux to an RMS error of 0.139% of the ψ-range against the free-boundary solver — clearing the 1% acceptance target with a wide margin, and improving on a prior attempt at 1.79% — with a maximum error of 4.39% localised near the X-point/separatrix and a Grad–Shafranov operator residual of 0.32% confirming interior physics consistency. It runs $2877\times$ faster than the reference solve ($0.87\,ms$ inference versus a $2.5\,s$ free-boundary Picard solve), which is what lets the controller linearise the equilibrium on the fly through equation (11) and read the flux response to a coil action. The graph-network state estimator (figure 7; KX-L1-A12) provides the attributable estimate of the state the intervention acts on: in detection it reaches an area under the ROC curve of 0.980, above the single-channel matched filters (0.917 strain-rate, 0.977 magnetization), and its independent re-derivation of the fused matched filter reproduces the KX-L1-A20 fused-precursor result (AUC 0.99166 against 0.99165) under a separate implementation. At the 1% false-positive rate a protection decision demands, the learned estimator reaches a true-positive rate of 0.729, above both single channels; the two-channel fusion ceiling is 0.852 at that operating point.



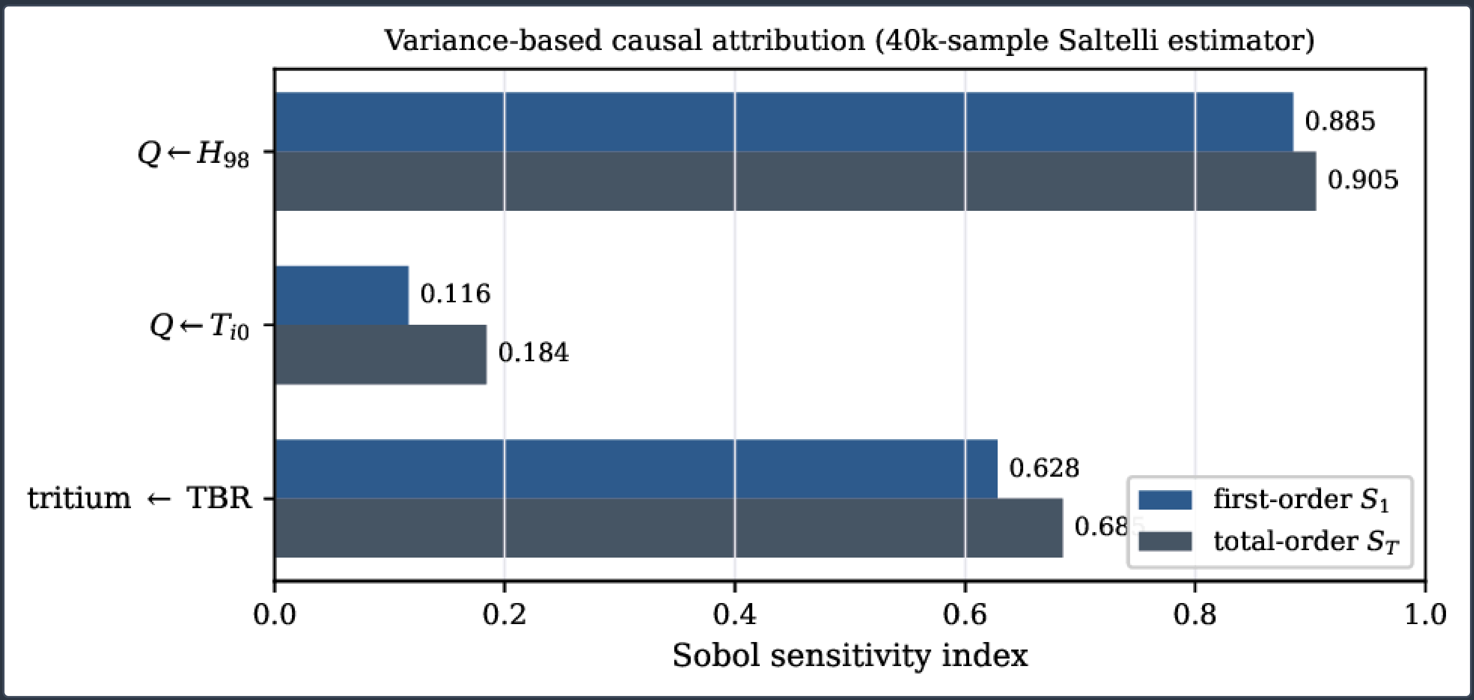
The physics-informed equilibrium operator (KX-L1-A11). **(a)** RMS ψ reconstruction error (0.139%) clears the 1% acceptance target and improves on the 1.79% prior attempt; the Grad–Shafranov operator residual (0.32%) confirms interior physics consistency. **(b)** The control-latency speedup: $0.87\,ms$ neural inference against a $2.5\,s$ free-boundary solve, a $2877\times$ acceleration.



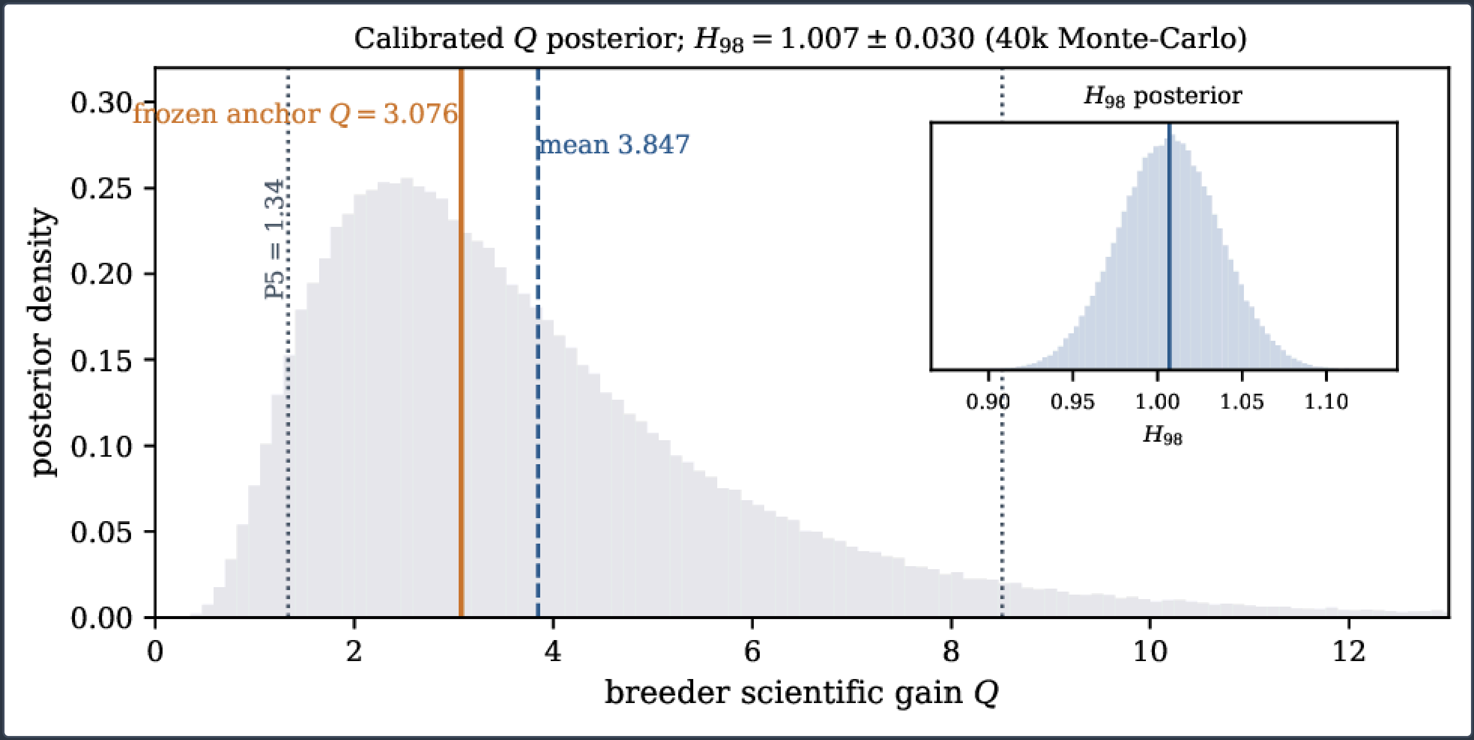
The attributable state estimator (KX-L1-A12). Detection area-under-ROC and true-positive rate at 1% false-positive rate for the single-channel matched filters, the learned graph-network estimator, and the two-channel fusion. The learned estimator beats each single channel and recovers most of the fusion gain; its explanations are read from the same differentiable mechanism.

THE POPULATION-LEVEL CAUSAL ATTRIBUTION

Figure 8 gives the variance-based attribution over the 40 000-sample ensemble (KX-L1-A5, Sobol refinement). The confinement factor H_{98} dominates the breeder-Q response, carrying a first-order Sobol index $S_1 = 0.885$ and a total index $S_T = 0.905$ — 88.5% of the variance first-order, with only a two-point gap to the total that bounds all interaction effects. The core ion temperature T_{i0} is a distant second ($S_1 = 0.116$, $S_T = 0.184$), and the tritium output is governed by the breeding ratio ($S_1 = 0.628$, $S_T = 0.685$). The ensemble is calibrated (figure 9): the anchor $Q = 3.076$ sits inside the posterior band, whose 5th and 95th percentiles are 1.34 and 8.51 around a mean of 3.847, and the confinement posterior is $H_{98} = 1.007 \pm 0.030$. The attribution tells the controller and the operator the same thing in one number: hold H_{98} and the operating point is held.



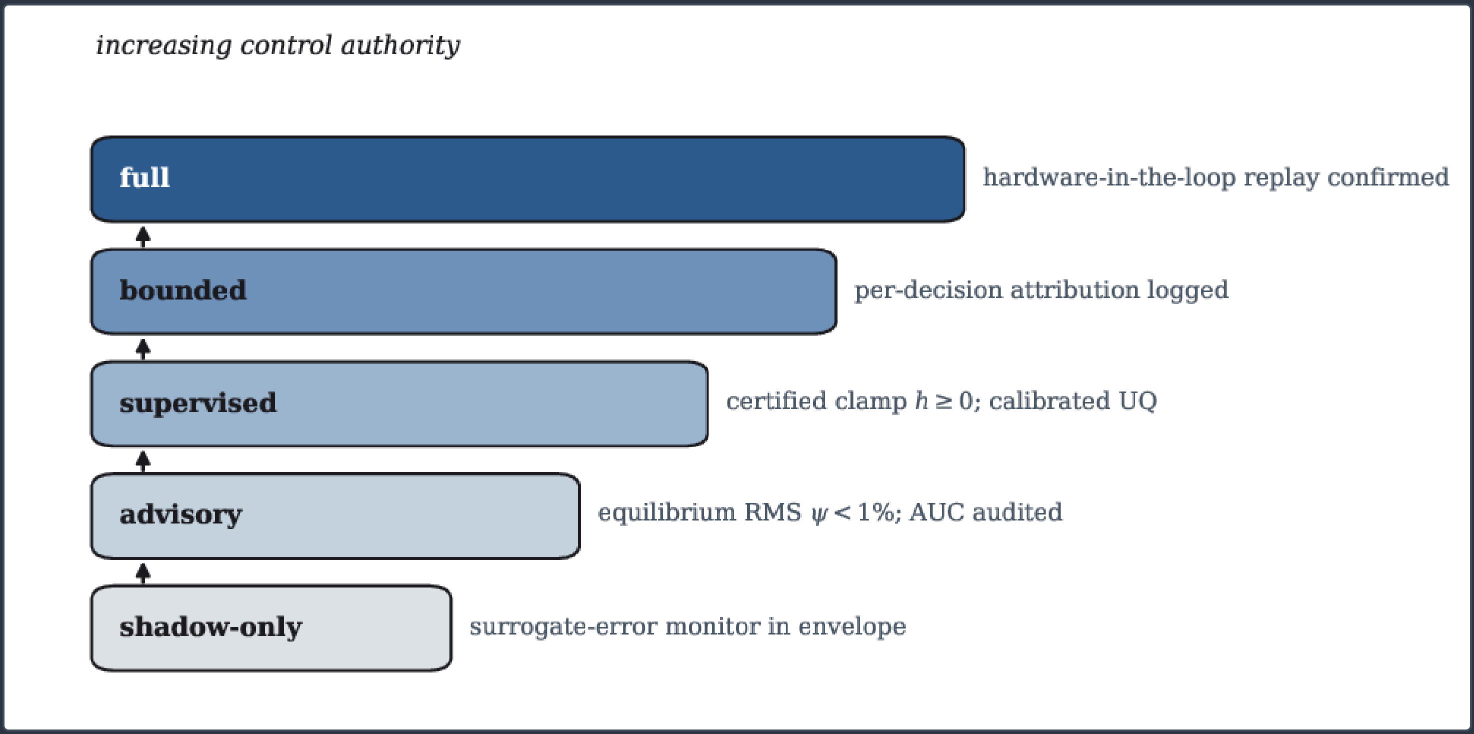
Population-level causal attribution by Sobol variance decomposition (KX-L1-A5; 40 000-sample Saltelli estimator). The confinement factor H_{98} carries 88.5% of the breeder-Q variance first-order ($S_1 = 0.885$, $S_T = 0.905$); T_{i0} is a distant second; the tritium output is governed by the breeding ratio. The first-order/total gap bounds the interaction effects.



The calibrated ensemble the attribution is computed on (KX-L1-A5). The breeder-Q posterior is right-skewed with 5th/95th percentiles $1.34/8.51$ and mean 3.847; the frozen anchor $Q = 3.076$ (orange) sits inside the band. Inset: the confinement posterior $H_{98} = 1.007 \pm 0.030$.

MATURITY-GATED AUTHORITY

The two forms of causal reasoning compose into a single auditable controller, and its authority is not binary. Each mechanism is admitted to a graded level of control authority — shadow-only, advisory, supervised, bounded, full — and rises only as it clears successive acceptance criteria (figure 10), the same maturity-gated principle that governs the certified clamp and the companion actuation-authority result [37] (official page). A mechanism that has passed only its surrogate-error monitor may inform but not command; full-plant authority requires the certified clamp, calibrated uncertainty, logged per-decision attribution, and a hardware-in-the-loop confirmation. Because the mapping from criteria to authority is explicit, the path from the demonstrated subsystems to a fully auditable autonomous controller is a defined, testable sequence.



(Schematic.) Verification-maturity gating. Each rung admits a higher level of control authority and is unlocked by a pre-registered acceptance criterion (the demonstrated clamp, equilibrium, state-estimator and attribution bars of table 1); the authority ladder itself is a schematic of the gating policy, not measured data.

§ 5

Discussion

WHY CAUSAL, AND WHY HERE

The claim is not that a causal model is more accurate than a predictive one; it is that a controller *acts*, and acting is intervening. Equation (5) is the whole argument: the distribution of the outcome under an action is generically not the distribution one would read off correlations, and a controller that confuses the two will be confidently wrong exactly where it matters [2,5]. The Kronos twin sidesteps the hardest part of causal inference — identification of interventional effects from confounded observational data — because its mechanisms are known physics operators, so the truncated factorisation of equation (4) is *evaluated*, not *estimated*. The do-calculus here is therefore not an add-on but a description of what the differentiable twin already computes: severing the incoming edges of the intervened node is what running the mutilated submodel forward does.

QUANTITATIVE COMPARISON TO PUBLISHED BENCHMARKS

Each demonstrated metric is competitive against the relevant literature. The certified clamp realises the forward-invariance guarantee of the control-barrier-function framework [6,7] on a fusion operating limit, with $0/100$ crossings against $50/100$ unclamped; this is the same guarantee, on the same β_N limit, that the companion safety-filter result establishes with a calibrated-uncertainty envelope [33] (official page). The equilibrium operator's 0.139% RMS ψ error is far inside the $\sim 1\%$ tolerance that real-time reconstruction targets [12,14,24], and its $2877\times$ speedup is consistent with the three-to-five orders of magnitude reported for neural-operator surrogates of expensive plasma solvers [18,19,21]. The state estimator's ROC (AUC 0.980) is competitive with the deep disruption-forecasting literature [16] while being, by construction, read from a differentiable mechanism so that its decisions are attributable through equations (11)–(14). The Sobol attribution restates, in variance-based causal terms, the physical result that confinement quality governs a compact ST's gain — the negative-triangularity quiescence that the companion turbulence-suppression result establishes at the operating point [39] (official page) is exactly why H_{98} is the variable to hold. The per-decision explainability construction is the integrated-gradient and Shapley machinery [8,9] applied where its completeness axiom (equation 13) is meaningful: to a physics mechanism rather than a black box, which is the interpretability posture a high-stakes decision demands [5].

RELATION TO THE CONTROL STACK

The causal layer is not free-standing; it inherits its calibrated uncertainty from the ensemble twin and its abstention behaviour from the conformal-prediction layer [38] (official page), and its differentiable mechanisms are the neural-operator flux and equilibrium surrogates demonstrated elsewhere [34,35] (official pages: flux surrogate, equilibrium). What this paper adds on top of that stack is the explicit statement that the composed system is a causal model: control actions are executable interventions, decisions are gradient-attributable, and the population governor is identified by variance decomposition. The runtime per-decision explainer that packages these attributions for the operator is built to this same demonstrated pattern on a staged schedule, on the NVIDIA GPU HPC campaign.

§ 6

Limitations

One honest gate bounds the claims. The interventions are validated against the twin and the attributions against the mechanisms they explain, all at reduced-order, design-point-anchored fidelity: the β_N dynamics of equation (6) are a 1-D surrogate, and the state-estimator ROC is computed on synthetic signals. The one confirming step is a hardware-in-the-loop audit that replays live decisions against as-built diagnostics and checks each logged attribution against the mechanism that produced it; it is the top rung of figure 10 and it reuses the demonstrated benches unchanged. Until that audit is complete, the runtime per-decision explainer is carried at advisory authority, not command.

§ 7

Conclusion

The Kronos controller reasons causally by construction. Its digital twin is a differentiable structural causal model, so a control action is an executable do-calculus intervention $\mathrm{do}(u)$ whose interventional distribution (equation 4) the twin computes, and every decision is gradient-attributable through equations (11)–(14). A certified intervention on β_N holds the safety barrier with $0/100$ crossings and $h_{\min} = +0.000$ against $50/100$ crossings without it (KX-L1-A13); the mechanisms the explanations are drawn from are demonstrated at RMS $\psi = 0.139\%$ and $2877\times$ (KX-L1-A11) and AUC 0.980 (KX-L1-A12); and a variance-based attribution assigns 88.5% of the breeder-Q variance to H_{98} ($S_1 = 0.885$, $S_T = 0.905$; KX-L1-A5) on a calibrated ensemble in which the anchor $Q = 3.076$ lies inside the $[1.34, 8.51]$ band. Certified interventions and attributable decisions are the working core of auditable autonomy, and the maturity-gated authority argument makes the route from these demonstrated subsystems to a regulator-facing controller a defined, testable sequence.

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 underlying every quantitative claim.

Reproducibility. The intervention, mechanism and attribution metrics are frozen anchors in the Kronos Physics De-Risking Register [31]: the certified clamp (KX-L1-A13), the equilibrium operator (KX-L1-A11), the graph-network state estimator (KX-L1-A12), the fused precursor (KX-L1-A20), and the global sensitivity and its Sobol refinement (KX-L1-A5). Each is regenerable from the archived analysis under the fixed seed `20260726`. The register is deposited at doi:10.5281/zenodo.22133057; this paper is archived at doi:10.5281/zenodo.22645825, and its official version is hosted at https://www.kronosfusionenergy.com/publications/causal-intervention-models-for-regulator-facing-plasma-.html.

Data availability tocsectionData availability The intervention, mechanism and attribution computations underlying figures 3–9 are archived with the deposit at doi:10.5281/zenodo.22645825, which references the Kronos 2026 series including the AI and quantum-control paper [32]. All quantitative values trace to the register [31] 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 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 Pearl, Causal diagrams for empirical research, *Biometrika* **82** 669 (1995). doi:10.1093/biomet/82.4.669

2 J Pearl, *Causality: Models, Reasoning, and Inference*, 2nd edn (Cambridge University Press, 2009).

3 J Peters, D Janzing and B Sch\"olkopf, *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, 2017).

4 B Sch\"olkopf, F Locatello, S Bauer, N R Ke, N Kalchbrenner, A Goyal and Y Bengio, Toward causal representation learning, *Proc. IEEE* **109** 612 (2021). doi:10.1109/JPROC.2021.3058954

5 C Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nat. Mach. Intell.* **1** 206 (2019). doi:10.1038/s42256-019-0048-x

6 A D Ames, X Xu, J W Grizzle and P Tabuada, Control barrier function based quadratic programs for safety critical systems, *IEEE Trans. Autom. Control* **62** 3861 (2017). doi:10.1109/TAC.2016.2638961

7 A D Ames, S Coogan, M Egerstedt, G Notomista, K Sreenath and P Tabuada, Control barrier functions: theory and applications, *Proc. 18th European Control Conf. (ECC)* 3420 (2019). doi:10.23919/ECC.2019.8796030

8 M Sundararajan, A Taly and Q Yan, Axiomatic attribution for deep networks, *Proc. 34th Int. Conf. on Machine Learning (ICML)*, PMLR **70** 3319 (2017). arXiv:1703.01365

9 S M Lundberg and S-I Lee, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems (NeurIPS)* **30** 4765 (2017). arXiv:1705.07874

10 I M Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.* **55** 271 (2001). doi:10.1016/S0378-4754(00)00270-6

11 A Saltelli, P Annoni, I Azzini, F Campolongo, M Ratto and S Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Comput. Phys. Commun.* **181** 259 (2010). doi:10.1016/j.cpc.2009.09.018

12 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

13 A Pironti and M Walker, Fusion, tokamaks, and plasma control: an introduction and tutorial, *IEEE Control Syst. Mag.* **25**(5) 30 (2005). doi:10.1109/MCS.2005.1512794

14 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

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

16 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

17 M G Kapteyn, J V R Pretorius and K E Willcox, A probabilistic graphical model foundation for enabling predictive digital twins at scale, *Nat. Comput. Sci.* **1** 337 (2021). doi:10.1038/s43588-021-00069-0

18 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

19 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

20 M Raissi, P Perdikaris and G E Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* **378** 686 (2019). doi:10.1016/j.jcp.2018.10.045

21 K L van de Plassche, J Citrin, C Bourdelle *et al*, Fast modeling of turbulent transport in fusion plasmas using neural networks, *Phys. Plasmas* **27** 022310 (2020). doi:10.1063/1.5134126

22 P Benner, S Gugercin and K Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, *SIAM Rev.* **57** 483 (2015). doi:10.1137/130932715

23 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

24 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

25 B Dudson *et al*, FreeGS: free-boundary Grad–Shafranov equilibrium solver (2020). doi:10.5281/zenodo.3745669

26 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

27 R E Kalman, A new approach to linear filtering and prediction problems, *J. Basic Eng.* **82** 35 (1960). doi:10.1115/1.3662552

28 G Evensen, The ensemble Kalman filter: theoretical formulation and practical implementation, *Ocean Dyn.* **53** 343 (2003). doi:10.1007/s10236-003-0036-9

29 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

30 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

31 P I Ford, *The Physics De-Risking Register: Gate-by-Gate Evidence, Independently Re-Run and Stamped*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22133057; official page https://www.kronosfusionenergy.com/publications/de-risking-register.

32 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; official page https://www.kronosfusionenergy.com/publications/ai-quantum-control.html.

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

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

35 P I Ford, *Sub-Millisecond Neural Equilibrium Reconstruction Robust to Sensor Dropout*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645801; official page https://www.kronosfusionenergy.com/publications/sub-millisecond-neural-equilibrium-reconstruction-robus.html.

36 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; official page https://www.kronosfusionenergy.com/publications/four-coupled-digital-twins-for-a-compact-spherical-toka.html.

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

38 P I Ford, *Calibrated Uncertainty, Conformal Prediction, and Abstention for Safety-Critical Fusion Control*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645813; official page https://www.kronosfusionenergy.com/publications/calibrated-uncertainty-conformal-prediction-and-absent.html.

39 P I Ford, *Negative-Triangularity Turbulence Suppression in a Compact Spherical-Tokamak Breeder: Nonlinear Gyrokinetics of Ion- and Electron-Scale Quiescence*, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645722; official page https://www.kronosfusionenergy.com/publications/negative-triangularity-confinement.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

Causal Intervention Models for Regulator-Facing Plasma Control: Do-Calculus Prediction and Per-Decision Explainability in a Compact Spherical-Tokamak Breeder

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

Zenodo · doi:10.5281/zenodo.22645825

Read on kronosfusionenergy.com

The Physics De-Risking Register