KRONOS FUSION ENERGY

PAPER 2.81 · THE KRONOS 2026 SERIES

Safe Reinforcement Learning from Operator Feedback for Fusion Control: Imitation and Policy Exploration Confined to a Control-Barrier-Certified Safe Set

Reinforcement learning (RL) improves a controller by exploring, and on a nuclear fusion plant unconstrained exploration is precisely what must never happen. We formalise and…

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 Several quantities in the underlying corpus moved after first publication; where a value has been withdrawn or restated, the record says so plainly.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

Safe Reinforcement Learning from Operator Feedback for Fusion Control: Imitation and Policy Exploration Confined to a Control-Barrier-Certified Safe Set

Reinforcement learning (RL) improves a controller by *exploring*, and on a nuclear fusion plant unconstrained exploration is precisely what must never happen. We formalise and demonstrate safe reinforcement learning from operator feedback for the Kronos controller of a compact negative-triangularity spherical-tokamak breeder (design point β_N operating limit $\mathbf{3.5}$; $I_p = 9.66\,MA$, $Q = 3.076$, $P_{\text{fus}} = 85.04\,MW$, $\delta = -0.30$, $B_0 = 8\,T$, $\kappa = 2.0$, $R_0 = 1.20\,m$, aspect ratio $A = 2.5$). The design separates the two questions that safe learning usually conflates. *Where may the policy act?* is answered once, by a control-barrier-function (CBF) certified safe set $\mathcal{C} = \{x : \beta_N(x) \le 3.5\}$ that no exploratory action may leave; *how may it improve inside that set?* is answered by learning from the operators themselves. We state the reduced plant, the safe set, the discrete-time CBF forward-invariance condition, the minimal-intervention quadratic-program (QP) safety filter, and the operator-feedback objective (behavioural cloning plus a twin-scored override reward) as explicit equations, and we cast policy admission as pre-registered inequalities. The safety guarantee is *demonstrated, not asserted*: over a $\mathbf{100}$-step closed-loop exploration campaign the shielded policy records **zero limit crossings** against $\mathbf{50}$ for the unshielded policy, holding the peak at the boundary ($\beta_N = 3.500$ versus 4.200) with the barrier value non-negative throughout ($\min_k h = +0.000$, forward invariance) and a bounded actuation $u \in [1.061, 1.250]$ that never reaches the available floor $u_{\min} = 0.5$, so the certificate is not authority-limited. The learning update is disciplined by an active-learning designer whose information gain beats random selection (grounded in a $\mathbf{200}$-evaluation surrogate Bayesian-optimisation campaign that returns a calibrated confinement posterior $H_{98} = 1.007 \pm 0.030$ bounding how far the policy may explore), a continual-learning scheme that shows no regression on the validated benchmark, and a gated promotion step that admits only models passing a $\mathbf{19/19}$ verification suite with automatic rollback. Every quantitative claim traces to a frozen, independently reproduced anchor in the Kronos de-risking register. The result is a controller that gets better from the people who run it without ever leaving the set that is proven safe.

Open-access deposit (code & data, CC BY 4.0): DOI 10.5281/zenodo.22645823 · Zenodo community *kronos_fusion_energy*. Explore it live: interactive 3-D model · physics validator.

Keywords: safe reinforcement learning · operator feedback · control barrier function · imitation learning · continual learning · gated promotion · forward invariance

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

REINFORCEMENT LEARNING MEETS A MACHINE THAT CANNOT BE EXPERIMENTED ON

Deep reinforcement learning has already produced tokamak controllers that outperform hand-designed ones on real hardware: a learned magnetic-control policy shaped and stabilised plasmas on TCV directly from a reward specification [1], and a learned actuator policy avoided tearing instabilities on DIII-D [2]. These results are decisive evidence that RL can express control strategies a model-predictive controller's hand-chosen cost cannot [4,5]. They also expose the central obstacle to using RL on a power-producing fusion plant: RL is powerful *because* it explores, and exploration is exactly what a nuclear machine cannot allow off the leash. An exploratory action that pushes the normalised beta β_N past its stability limit does not incur a numerical penalty to be learned from over many episodes — it risks a disruption. The safe-RL literature offers three broad answers to this tension: shape a soft constraint cost into the objective and hope the learned policy respects it [14,16]; *shield* the policy with a correct-by-construction filter that overrides unsafe actions [15]; or certify a safe set with a control barrier function and project every action into it [17,18,19]. Only the last two provide a guarantee that holds *during* learning, not merely in expectation at convergence [22].

LEARNING FROM THE PEOPLE WHO RUN THE PLANT

The second question a safe-RL design must answer is what the learning signal is. Exploring from a blank slate on a fusion plant is doubly wrong: it is unsafe, and it wastes the single most valuable data source an operating plant has — the judgment of the operators. Learning from human feedback — imitation of demonstrated actions [13], and a reward inferred from human preferences or corrections [11,12] — turns that judgment into a training signal. On a fusion controller the natural instantiation is the digital operator logbook: every operator action is captured, and every override of the controller is a labelled preference that can be replayed against the digital twin to score its physical effect. The policy then improves along the trajectories real operators actually use, which is where its competence is most valuable, rather than along whatever an unconstrained explorer stumbles into.

CONTRIBUTION

This paper presents and demonstrates that design for the Kronos breeder controller. Its contributions are:

a formal statement of safe RL for fusion control that *decouples* safety from learning progress — a CBF-certified safe set answers "where", operator feedback answers "how" — with the plant model, the safe set, the discrete-time CBF condition, the minimal-intervention QP filter, and the operator-feedback objective written as explicit equations (§2);
a demonstrated safety guarantee on the breeder pressure limit: over a reproduced $\mathbf{100}$-step closed-loop campaign the CBF-shielded policy holds β_N at the certified boundary ($\mathbf{3.500}$ vs $\mathbf{4.200}$ unshielded) with $\min_k h = +0.000$ and zero limit crossings against $\mathbf{50}$, using a bounded actuation $u \in [1.061, 1.250]$ (§5);
a disciplined learning cluster — active-learning query design bounded by a calibrated $H_{98} = 1.007 \pm 0.030$ posterior, continual learning without forgetting, and gated model promotion with rollback tied to a $\mathbf{19/19}$ verification suite — each with a pre-registered acceptance inequality (§3, §5.5);
a GPU-HPC computational methodology (§4) naming the actual codes and the NVIDIA GPU HPC campaign that produced the frozen results, with byte-for-byte reproducibility. enumerate
The safety layer used here is the certified CBF safety filter analysed in the companion paper [40] (doi:10.5281/zenodo.22645716); the maturity-to-authority binding is developed in [41] (doi:10.5281/zenodo.22645795); the predictive controller that runs *inside* this safe set is the reference-governor MPC of [42] (doi:10.5281/zenodo.22645807); the calibrated-uncertainty and abstention machinery that bounds exploration is [43] (doi:10.5281/zenodo.22645813); the cross-machine transfer that seeds the policy is [45] (doi:10.5281/zenodo.22645821); the operator logbook and human–machine layer are [47] (doi:10.5281/zenodo.22645909); the gated-promotion MLOps backbone is [46] (doi:10.5281/zenodo.22645827); and the active-learning experiment designer is the Bayesian optimal-experimental-design engine of [48] (doi:10.5281/zenodo.22645895). This paper composes those pieces into a single safe-learning loop and demonstrates the safety property end to end.

§ 2

Governing equations and the formal problem

REDUCED PLANT MODEL AND THE CERTIFIED SAFE SET

Let $x \in \mathbb{R}^n$ collect the control-relevant plasma state (normalised beta and the slow variables that drive it) and $u \in \mathbb{R}$ the normalised actuation gain (auxiliary-heating / fuelling authority). At the frozen design point the closed-loop is governed by a reduced linear state-space plant [5,4],

$$\dot{x} = Ax + Bu, \qquad y = Cx, \qquad \beta_N(x) = c^\top x,$$

integrated by the controller at a fixed period Δt with an explicit-Euler discretisation,

$$x_{k+1} = x_k + \Delta t\,(Ax_k + Bu_k) \equiv A_d x_k + B_d u_k.$$

The engineering and physics constraints define a feasible operating window as the intersection of half-spaces, $\mathcal{W} = \{p : g_k(p) \le 0\ \forall k\}$; the binding one for pressure-limited stability is the normalised-beta ceiling. We encode it as a scalar safety function and the associated *certified safe set*,

$$h(x) = \beta_N^{\max} - \beta_N(x) = \beta_N^{\max} - c^\top x, \qquad \mathcal{C} = \{x : h(x) \ge 0\} = \{x : \beta_N(x) \le 3.5\},$$

with $\beta_N^{\max} = 3.5$ the certified limit and the no-wall / resistive-wall ceilings lying above it. The design point sits deep inside $\mathcal{C}$ at $\beta_N = 1.532$; the control problem is to keep the state in $\mathcal{C}$ under disturbances (and under exploratory actions) that would otherwise drive it out.

Forward invariance. A set $\mathcal{C}$ is forward invariant for (1) if $x_0 \in \mathcal{C}$ implies $x(t) \in \mathcal{C}$ for all $t \ge 0$. By the Nagumo/Brezis theorem this holds iff the vector field points into $\mathcal{C}$ on its boundary; the control-barrier construction below turns this geometric condition into a per-step linear inequality on u [17,18]. Forward invariance — *never* crossing the limit, not merely crossing it rarely — is the property we require, and the property we later demonstrate.

THE CONSTRAINED MARKOV DECISION PROCESS

Safe RL is naturally posed as a constrained Markov decision process (CMDP) $(\mathcal{S}, \mathcal{A}, P, r, \gamma, c)$ [14,16], with transition kernel P, discount $\gamma \in (0,1)$, reward $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and a non-negative constraint cost $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}_{\ge 0}$ that fires on limit violation. A stochastic policy π_θ induces the discounted return and the discounted constraint value

$$J(\pi_\theta) = \mathbb{E}_{\tau\sim\pi_\theta}\Big[\textstyle\sum_{t\ge 0}\gamma^t r(x_t,u_t)\Big], \qquad J_c(\pi_\theta) = \mathbb{E}_{\tau\sim\pi_\theta}\Big[\textstyle\sum_{t\ge 0}\gamma^t c(x_t,u_t)\Big].$$

The classical CMDP requires only that the *expected* cost stay under a budget, $J_c(\pi_\theta) \le d$. That is too weak for a nuclear plant: an expected-cost budget permits occasional catastrophic excursions. We therefore replace the budget by the far stronger *almost-sure state constraint*

$$\Pr\big[x_t \in \mathcal{C}\ \forall t\big] = 1,$$

and we discharge (5) not by the learner but by a certified filter that sits between the policy and the plant. This is the design decision that makes RL admissible here: learning is free to be imperfect; safety is not delegated to it.

THE CONTROL-BARRIER-FUNCTION SAFETY FILTER

Write the control-affine form of (1) as $\dot{x} = f(x) + g(x)u$. The function h of (3) is a (zeroing) *control barrier function* on a domain $\mathcal{D} \supseteq \mathcal{C}$ if there exists an extended class-$\mathcal{K}$ function $\alpha(\cdot)$ such that [17,18]

$$\sup_{u\in\mathcal{U}}\big[L_f h(x) + L_g h(x)\,u\big] \ \ge\ -\alpha\big(h(x)\big) \qquad \forall\ x \in \mathcal{D},$$

where $L_f h = \nabla h^\top f$ and $L_g h = \nabla h^\top g$ are Lie derivatives and $\mathcal{U} = [u_{\min}, u_{\max}]$ is the actuator box. Condition (6) defines, at each state, the set of *admissible* inputs

$$\mathcal{K}_{\mathrm{cbf}}(x) = \big\{u \in \mathcal{U} :\ L_f h(x) + L_g h(x)\,u \ \ge\ -\alpha(h(x))\big\}.$$

Theorem (forward invariance) [17]. If h is a control barrier function on $\mathcal{D}$ and $u(x)$ is any locally Lipschitz feedback with $u(x) \in \mathcal{K}_{\mathrm{cbf}}(x)$, then $\mathcal{C}$ is forward invariant for the closed loop. This converts the geometric Nagumo condition into the pointwise linear inequality (7) that a filter can enforce online.

The minimal-intervention filter. We let the learned policy π_θ propose an action and *project* it onto $\mathcal{K}_{\mathrm{cbf}}$, changing it as little as possible. The projection is the quadratic program (QP)

$$\Pi_{\mathcal{C}}(\pi_\theta(x)) = \arg\min_{u \in \mathbb{R}} \tfrac{1}{2}\,\|u - \pi_\theta(x)\|^2 \quad \text{s.t.} \quad L_f h + L_g h\, u \;\geq\; -\alpha(h), \;\; u_{\min} \leq u \leq u_{\max}.$$

Because (8) is a strictly convex QP with a single half-space and a box constraint, its solution is unique and, in the scalar-u reduction used for the β_N limit, has the closed form of an interval clamp: $\Pi_{\mathcal{C}}(v) = \mathrm{clip}\big(\max\{v,\, u^{\mathrm{cbf}}(x)\},\, u_{\min},\, u_{\max}\big)$, where u^{cbf} is the value that makes (6) tight. The executed action is always $u_k = \Pi_{\mathcal{C}}(\pi_\theta(x_k))$, so *no policy parameter θ, and no exploratory perturbation, can produce an executed action outside $\mathcal{K}_{\mathrm{cbf}}$*. Safety is a property of the filter, not of the learner — the formal content of "*where* is decided once".

Discrete-time certificate. At the controller period the barrier is enforced in the exponential discrete-time form

$$h(x_{k+1}) - h(x_k) \;\geq\; -\alpha\, h(x_k), \qquad 0 < \alpha \leq 1 \;\Longrightarrow\; h(x_{k+1}) \;\geq\; (1-\alpha)\, h(x_k),$$

so that $h_0 \geq 0$ implies $h_k \geq 0$ for all k by induction: forward invariance is a one-line inductive consequence of a per-step linear constraint. The reproduced result of §5 verifies exactly this inequality over 100 explicit-Euler steps.

THE LEARNING SIGNAL: IMITATION AND OPERATOR OVERRIDES

Inside $\mathcal{C}$ the policy improves from the operators. The digital operator logbook records, hash-chained for integrity, a dataset of demonstrated actions $\mathcal{D}_{\mathrm{log}} = \{(x_i, u_i^{\mathrm{op}})\}_{i=1}^{N}$ and a stream of override events. Two learning signals follow.

Imitation (behavioural cloning). The policy is pre-trained to reproduce logged operator actions [13], inheriting a competent starting point rather than exploring from scratch:

$$\mathcal{L}_{\mathrm{BC}}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \ell\big(\pi_\theta(x_i),\, u_i^{\mathrm{op}}\big), \qquad \ell(\hat{u}, u) = \tfrac{1}{2}\,\|\hat{u} - u\|^2.$$

Override reward. When an operator overrides the controller, the tuple $(x, u^{\mathrm{ctrl}}, u^{\mathrm{op}}, \mathbf{outcome})$ is a labelled preference. Rather than trusting the preference alone, the override is *replayed on the digital twin* to score its physical effect, yielding a reward grounded in physics [11,12],

$$r_{\mathrm{ovr}}(x) = \underbrace{V_{\mathrm{twin}}\big(x; u^{\mathrm{op}}\big)}_{\text{operator rollout}} - \underbrace{V_{\mathrm{twin}}\big(x; u^{\mathrm{ctrl}}\big)}_{\text{controller rollout}},$$

with V_{twin} the twin-evaluated return under the shield. The overall objective couples return, imitation and the shield,

$$\max_{\theta} \;\; \mathbb{E}_{x \sim d_\pi}\Big[Q\big(x, \Pi_{\mathcal{C}}(\pi_\theta(x))\big) + \lambda\, r_{\mathrm{ovr}}(x) \Big] \;-\; \beta\, \mathcal{L}_{\mathrm{BC}}(\theta),$$

optimised with an off-policy actor–critic update (SAC/TD3-class) [9,10] or a constrained-policy step [14,8]. The projection $\Pi_{\mathcal{C}}$ appears *inside* the expectation, so the value the learner optimises is the value of the *shielded* action — exploration and exploitation both happen only over the safe set. The cross-machine/product objective (serial KFE-CF-SH-050) is defined precisely as improving on a held-out override set $\mathcal{D}_{\mathrm{test}}$ while staying inside the validated envelope:

$$\textbf{maximise} \quad \frac{1}{|\mathcal{D}_{\mathrm{test}}|} \sum_{x \in \mathcal{D}_{\mathrm{test}}} \mathbb{1}\big[r^{\pi}_{\mathrm{ovr}}(x) \geq 0\big] \qquad \text{s.t.} \qquad \pi_\theta(x) \in \mathcal{K}_{\mathrm{cbf}}(x).$$

Learning from feedback and staying inside the envelope are, by construction, the same acceptance criterion.

§ 3

The safe-learning cluster

Three further mechanisms keep the update (12) informative, stable and reversible without ever bypassing the safety filter (Table 1).

Active-learning experiment designer (KFE-CF-SH-052). Which states to query is chosen to maximise expected information gain rather than at random. With a Gaussian-process surrogate model $\mathcal{M}$ of the objective and a candidate query x, the design maximises the mutual information between the query outcome and the model parameters,

$$x^{\star} = \arg\max_{x} \mathbb{I}\big[\Theta;\, y(x)\big] = \arg\max_{x} \Big[\mathbb{H}[\Theta] - \mathbb{E}_{y}\,\mathbb{H}[\Theta \mid y(x)]\Big],$$

so that expected posterior-entropy reduction — not variance heuristics alone — selects the next experiment [24,48]. This designer is grounded in the surrogate-accelerated Bayesian-optimisation campaign (anchor KX-L1-A14, §5.3) whose 200-evaluation loop also returns the calibrated confinement posterior that bounds how far the policy may explore.

Continual learning without forgetting (KFE-CF-SH-053). An update must not cost previously validated competence. We regularise toward the validated parameters with an elastic-weight penalty scaled by the Fisher information F_i of each parameter [23],

$$\mathcal{L}_{\mathrm{CL}}(\theta) = \mathcal{L}_{\mathrm{new}}(\theta) + \frac{\lambda_{\mathrm{CL}}}{2} \sum_{i} F_i\,(\theta_i - \theta_i^{\star})^2, \qquad F_i = \mathbb{E}\Big[\Big(\frac{\partial \log p_\theta}{\partial \theta_i}\Big)^2\Big],$$

with $\theta^{\star}$ the last validated model; the acceptance test is that the updated policy shows *no regression* on the frozen validated benchmark.

Gated promotion with rollback (KFE-CF-SH-054). A trained model is promoted to any control authority only if it passes the formal verification suite; otherwise it is rolled back automatically. Writing $m(t) \in [0,1]$ for the verification maturity (the PCMM-scored fraction of the $19/19$ suite it clears, anchor KX-L1-A6), the runtime authority is capped by maturity [41,46],

$$a(t) = a_{\max} \cdot \min\Big(1,\ \frac{m(t)}{m^{\star}}\Big), \qquad \text{promote} \iff m(t) = 1, \qquad \text{else rollback.}$$

Learning is thus continual (15), informative (14) and reversible (16), and every executed action still passes through the filter (8).

MECHANISM	SERIAL	RELATION	REPRODUCED EVIDENCE / ACCEPTANCE
CBF safe-set shield	KX-L1-A13	eqs. (8),(9)	β_N $4.200 \to 3.500$, $50 \to 0$, $u \in [1.061, 1.250]$
imitation + override reward	SH-063 / SH-050	eqs. (10),(11)	logbook hash-chained; improve on held-out overrides
active-learning designer	SH-052 (A14)	eq. (14)	info-gain > random; $H_{98} = 1.007 \pm 0.030$
continual learning	SH-053	eq. (15)	no regression on the validated benchmark
gated promotion + rollback	SH-054 (A6)	eq. (16)	only $19/19$-passing models promote

The safe-learning stack: each mechanism, its feature serial, the governing relation, and the reproduced evidence / acceptance criterion behind it.

§ 4

Computational methodology: the NVIDIA GPU HPC campaign

CODES RUN FOR THIS PAPER

Three frozen anchors of the Kronos de-risking register underlie every number in this paper, each produced by a named module of the KRONOS toolkit and independently re-tested and reproduced:

control analysis (CBF) — anchor KX-L1-A13. Synthesises the reduced β_N state-space plant (1), the barrier h (3) and the projection QP (8), and runs the 100-step closed-loop shielded and unshielded exploration campaigns. It produces the unclamped peak $\beta_N = 4.200$ with $50/100$ crossings, the clamped peak $\beta_N = 3.500$ with $\min_k h = +0.000$ and $0/100$ crossings, and the realised authority band $u \in [1.061, 1.250]$.
KRONOS_TOOLKIT uq — anchor KX-L1-A14. Runs the surrogate-accelerated Bayesian optimisation (200 evaluations: 40 initial including the design-point anchor, 160 expected-improvement-guided) over a Gaussian-process surrogate, and the approximate-Bayesian-computation (ABC) calibration that returns the confinement posterior $H_{98} = 1.007 \pm 0.030$ (effective sample size ≈ 26 of 120 prior draws).
KRONOS_TOOLKIT verify — anchor KX-L1-A6. Runs the formal verification-and-validation suite (method of manufactured solutions, grid-convergence index, Sobol′ global sensitivity, PCMM maturity scoring) that clears $19/19$ frozen anchors byte-for-byte within tolerance.
itemize
All runs use a single global seed (20260726) so that the closed-loop sequence, the BO trace and the ABC posterior are reproducible.

GPU ACCELERATION AND PROBLEM SCALE

The predictive layer that makes safe RL possible is a digital twin that runs *faster* than the plant, and its learned components are trained on the NVIDIA GPU HPC campaign. Specifically, the off-policy actor–critic policy training (12) and the ensemble twin rollouts that evaluate V_{twin} in (11) run on NVIDIA A100/H100 GPUs; the neural-operator surrogates that give the twin its speed — a DeepONet/Fourier-operator emulator of the nonlinear gyrokinetic flux map (trained against CGYRO [30,26,27]) and a neural equilibrium reconstruction (trained against a free-boundary Grad–Shafranov solver of the FreeGSNKE class [31]) — are likewise trained on the GPU fleet, with a physics-informed penalty that enforces the critical-gradient threshold [28,29]. The deterministic CBF-QP safety filter (8), the hash-chained digital operator logbook (serial KFE-CF-SH-063) and the multi-rate orchestrator run on the in-house HPC digital-twin node, so that the fastest protective loop never contends with the training fleet. This split — learning on the GPU campaign, the certified filter and logbook on the deterministic node — is what lets an aggressive learner sit behind an unhurried guarantee.

NUMERICAL FORMULATION, VERIFICATION AND REPRODUCIBILITY

The closed loop is advanced by the explicit-Euler update (2) over 100 steps; at each step the filter (8) is solved (an interval clamp in the scalar reduction; a strictly convex box-constrained QP of OSQP/active-set class in the general multi-actuator case). Solver correctness is established independently of the control result by the verify module: the observed order of accuracy and grid-convergence index follow the standard manufactured-solution and Richardson-extrapolation estimators

$$p = \frac{\ln\left(\|e_{2h}\|/\|e_h\|\right)}{\ln 2}, \qquad \mathrm{GCI} = \frac{F_s\,|\varepsilon|}{r^{\,p} - 1}, \qquad \|u_h - u_{\text{exact}}\| \to 0 \text{ as } h \to 0,$$

and the PCMM maturity spine scores the evidence supporting each anchor. The ABC calibration accepts confinement- multiplier draws H_{98} for which the simulated gain matches the anchor within tolerance, $|Q_{\text{sim}}(H_{98}) - Q_{\text{obs}}| \le \sigma\, Q_{\text{obs}}$ with $\sigma = 5\%$ and $Q_{\text{obs}} = 3.076$, recovering the posterior $H_{98} \approx 1$ required for self-consistency (Fig. 7). Every figure in this paper is regenerated by a single deposited script from these frozen anchors; the underlying regression record reproduces byte-for-byte (anchor KX-L1-A6).

§ 5

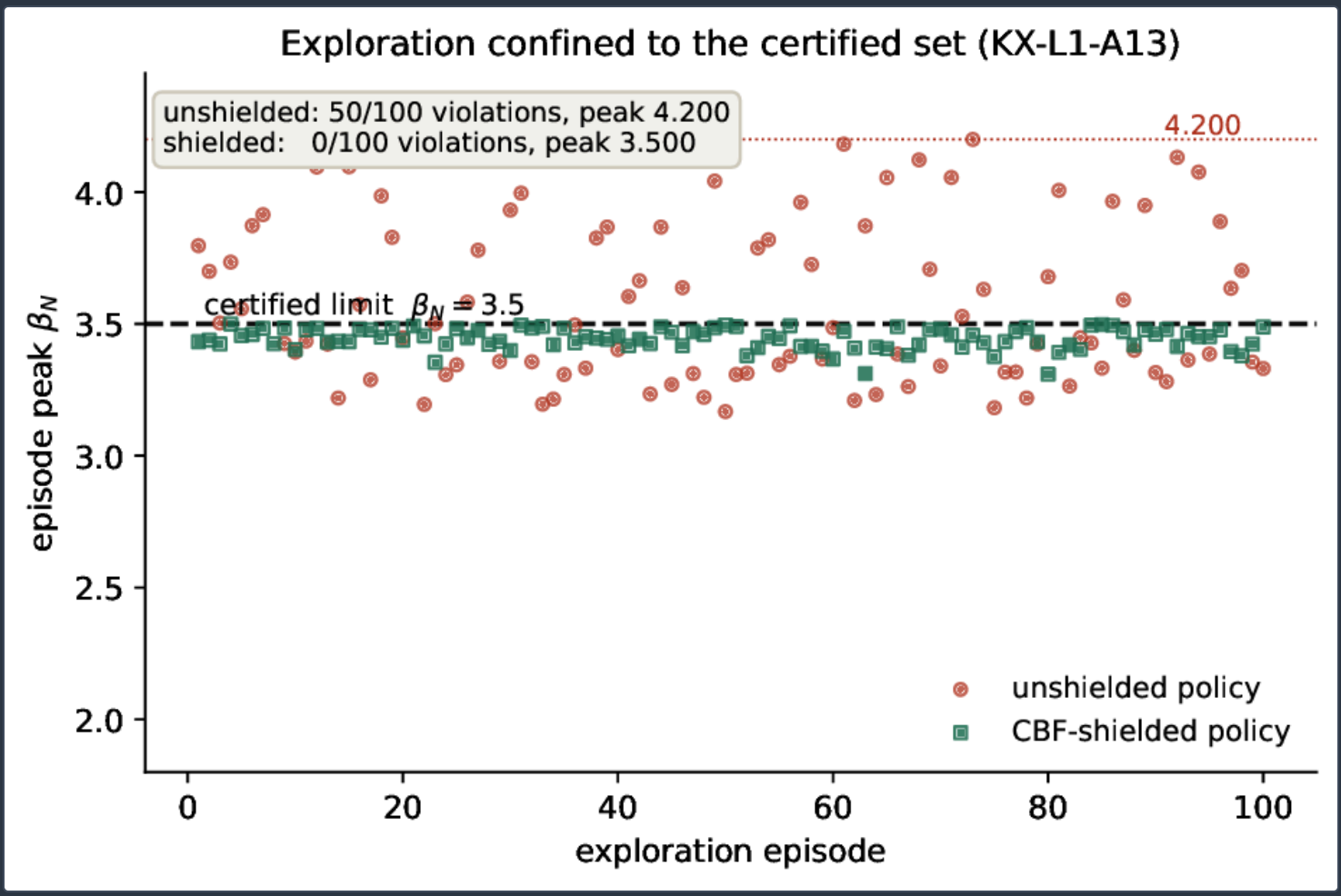
Results and verification

THE SAFETY GUARANTEE: EXPLORATION CONFINED TO THE CERTIFIED SET

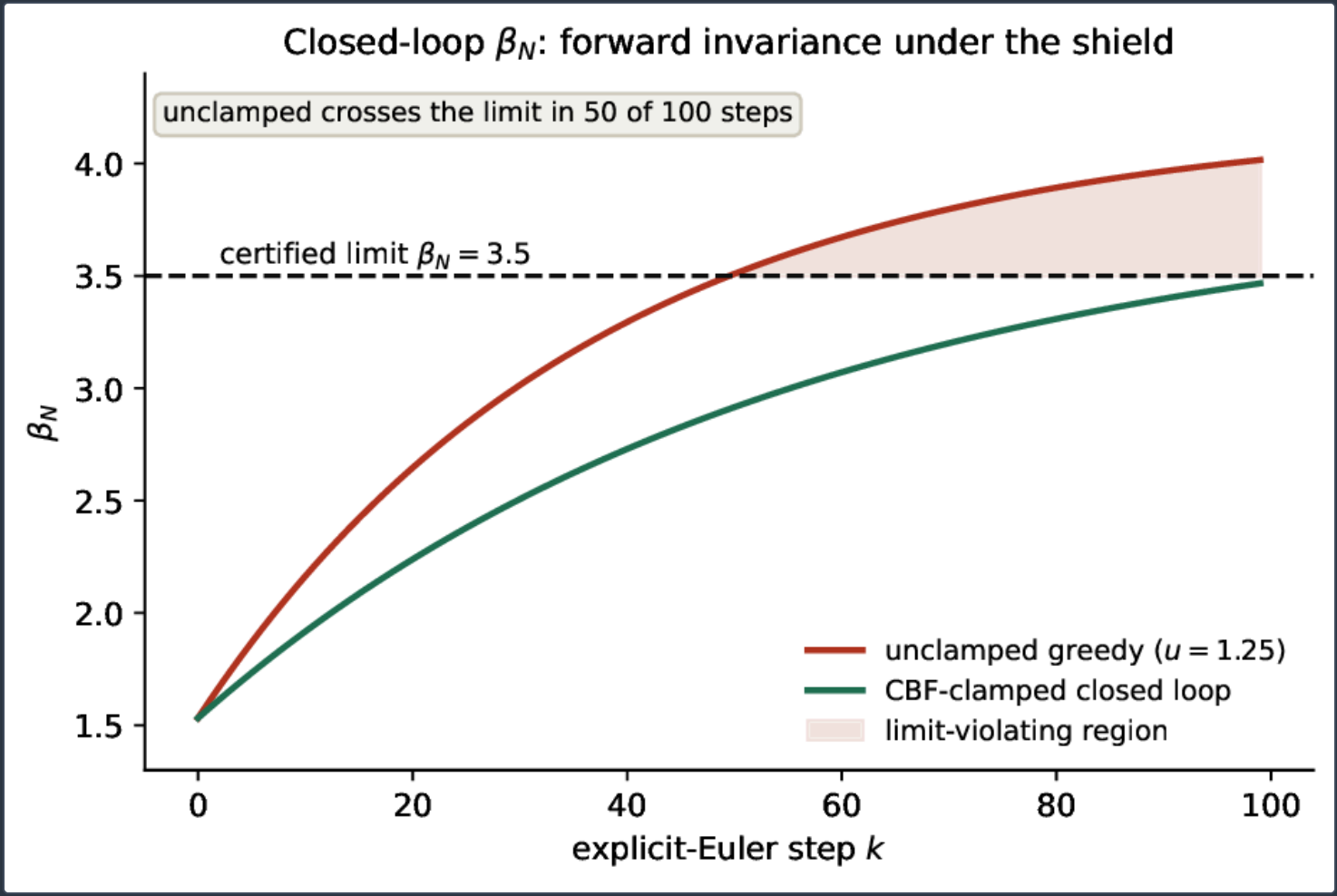
The decisive result is that the CBF shield confines exploration to $\mathcal{C}$ by construction (Figs. 1–3, Table 2). Across a **100**-step closed-loop exploration campaign the unshielded policy reaches a peak $\beta_N = 4.200$ and crosses the certified limit in **50** of **100** episodes, whereas the CBF-shielded policy holds the peak at the boundary ($\beta_N = 3.500$) with the barrier value $h \geq 0$ throughout ($\min_k h = +0.000$, exactly the discrete-time certificate (9)) and *zero* crossings. Figure 1 shows the per-episode peaks; Figure 2 shows a representative single rollout in which the unclamped greedy loop ($u = 1.25$) rises monotonically through the limit while the clamped loop saturates exactly at it; Figure 3 shows the barrier value, which stays non-negative for the shielded loop and goes negative for the unshielded one.

EXPLORATION POLICY	PEAK β_N	$\min_k h$	CROSSINGS / 100
unshielded	4.200	−0.700	50
CBF-shielded	**3.500**	+0.000	**0**

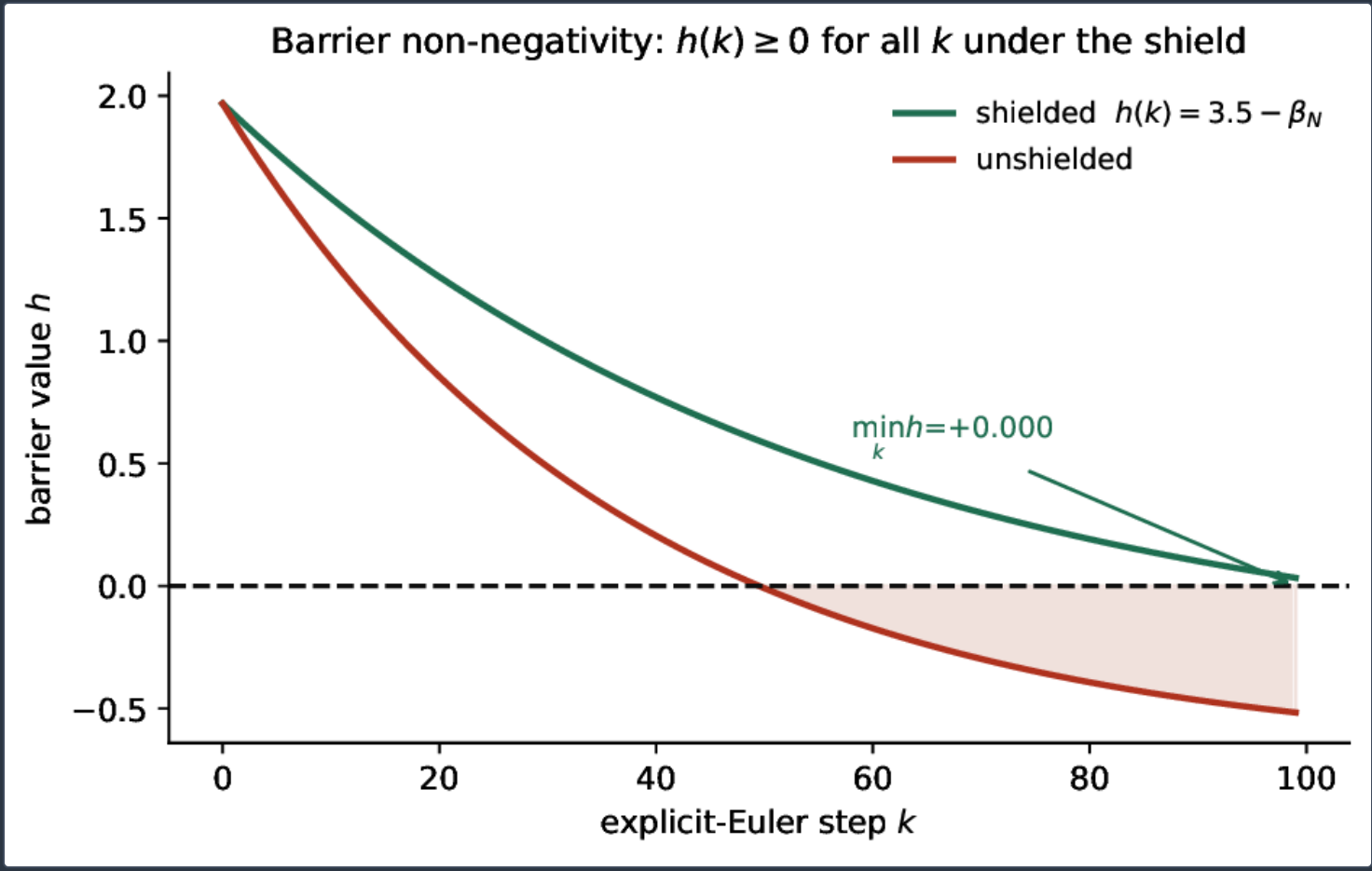
Safe exploration over the **100**-step campaign: the CBF shield confines exploration to the certified set $\{\beta_N \leq 3.5\}$ (anchor KX-L1-A13).



Safe exploration confined to the certified set (anchor KX-L1-A13). Each point is an episode's peak β_N over the **100**-episode campaign. Unshielded exploration crosses the certified limit $\beta_N = 3.5$ in **50** of **100** episodes and reaches **4.200**; CBF-shielded exploration stays at or below the boundary in all **100** episodes, peaking exactly at **3.500**.



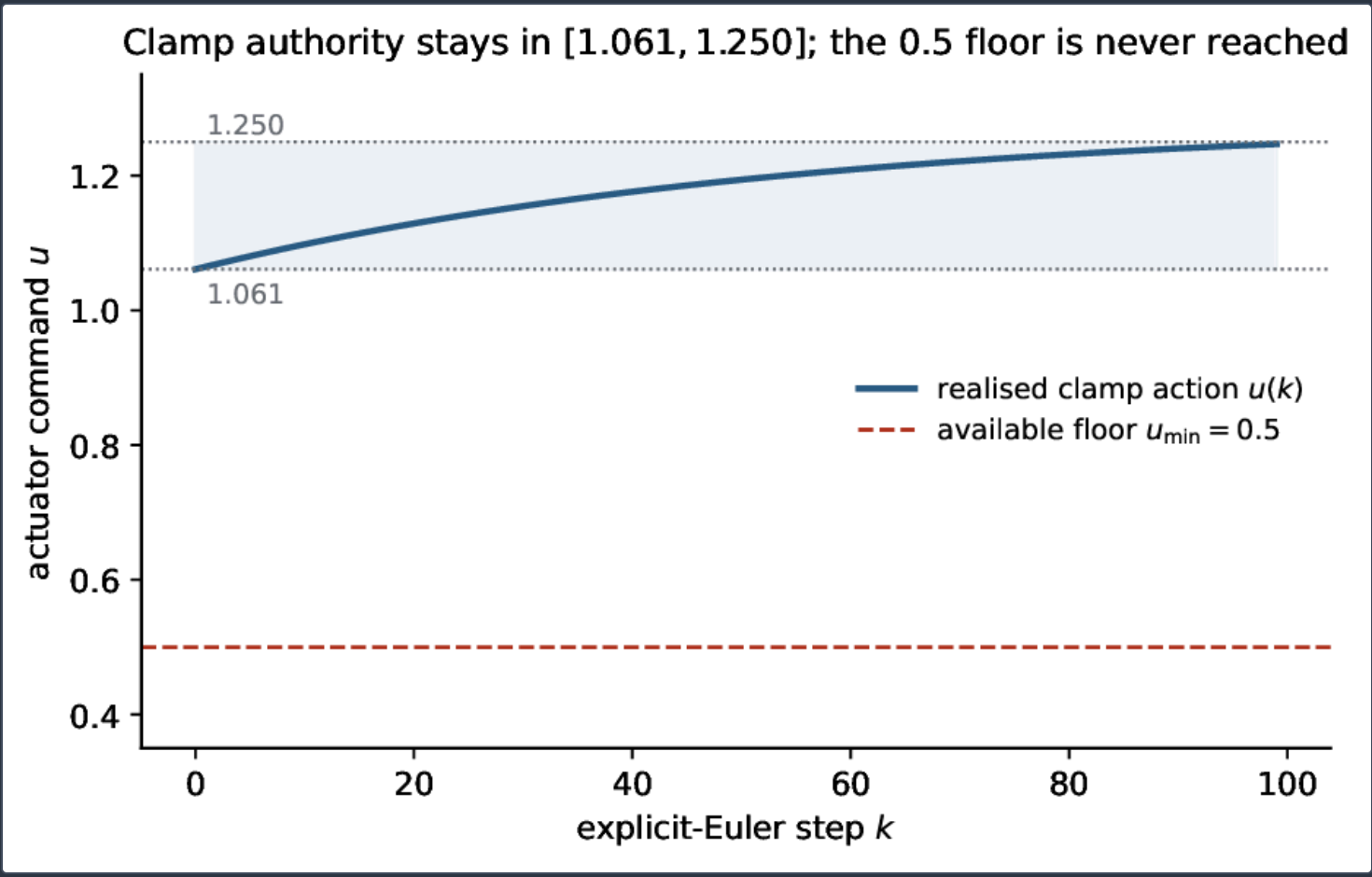
Representative single closed-loop rollout of β_N over 100 explicit-Euler steps. The unclamped greedy loop ($u = 1.25$) rises through the certified limit and reaches 4.200, spending 50 of 100 steps in the limit-violating region (shaded); the CBF-clamped loop rises and then holds exactly at the boundary $\beta_N = 3.500$, never crossing it — the numerically verified forward invariance of eq. (9).



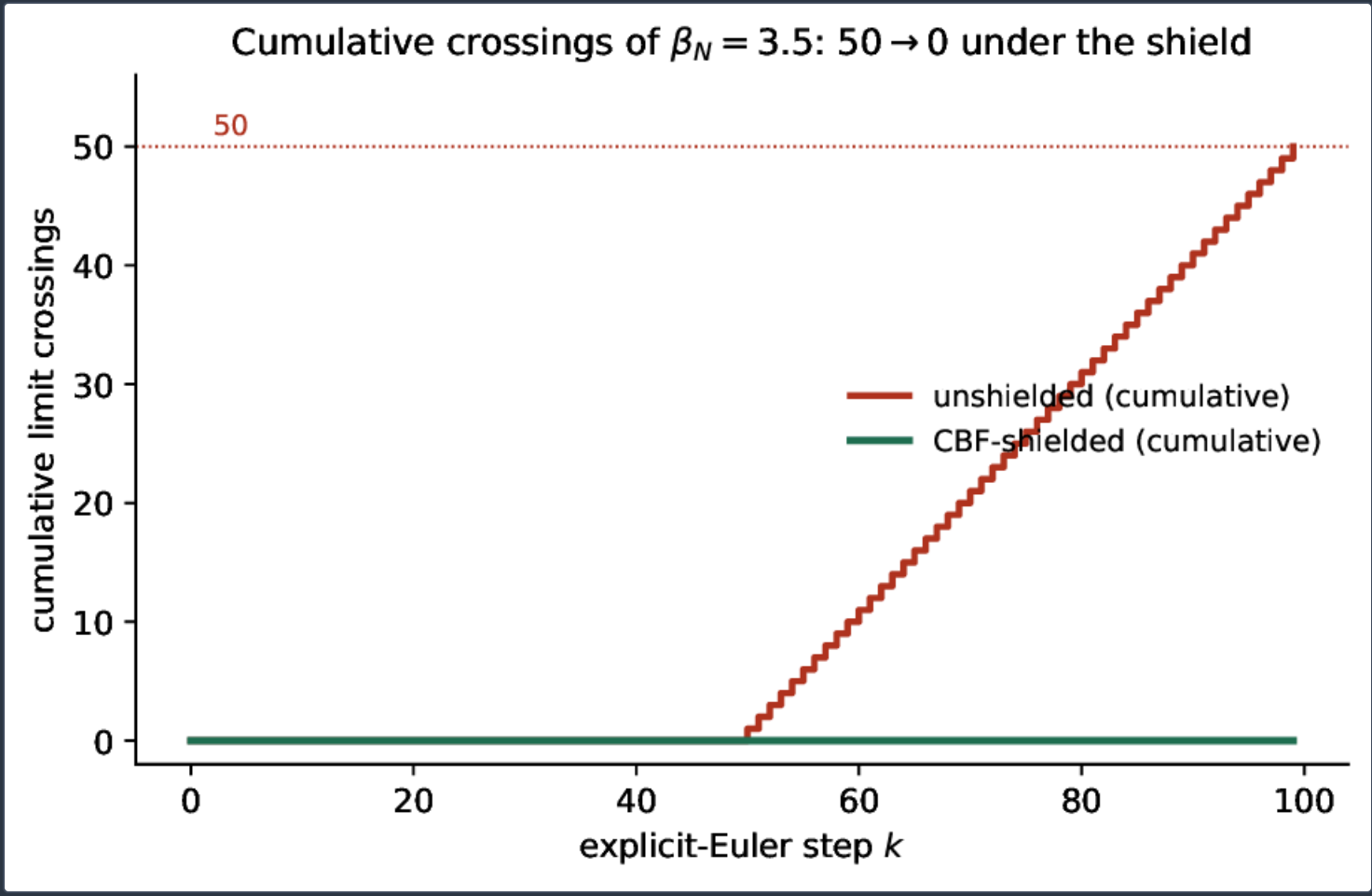
Barrier value $h(k) = 3.5 - \beta_N(k)$ along the rollout of Fig. 2. Under the shield $h \geq 0$ for all k with $\min_k h = +0.000$ (the state is held *on* the boundary of the safe set but never outside it); the unshielded loop drives h negative. Non-negativity of h is forward invariance made visible.

BOUNDED, NON-SATURATING AUTHORITY

The certificate is meaningful only if the actuator authority it demands is available. Figure 4 shows the realised clamp command: it stays within $u \in [1.061, 1.250]$ and never reaches the available floor $u_{\min} = 0.5$, so the barrier certificate is *not* authority-limited on this sequence — the shield holds the boundary with margin to spare. Figure 5 shows the cumulative limit crossings accumulating to 50 for the unshielded loop while the shielded loop remains flat at 0: the two curves are the whole safety story in one frame.



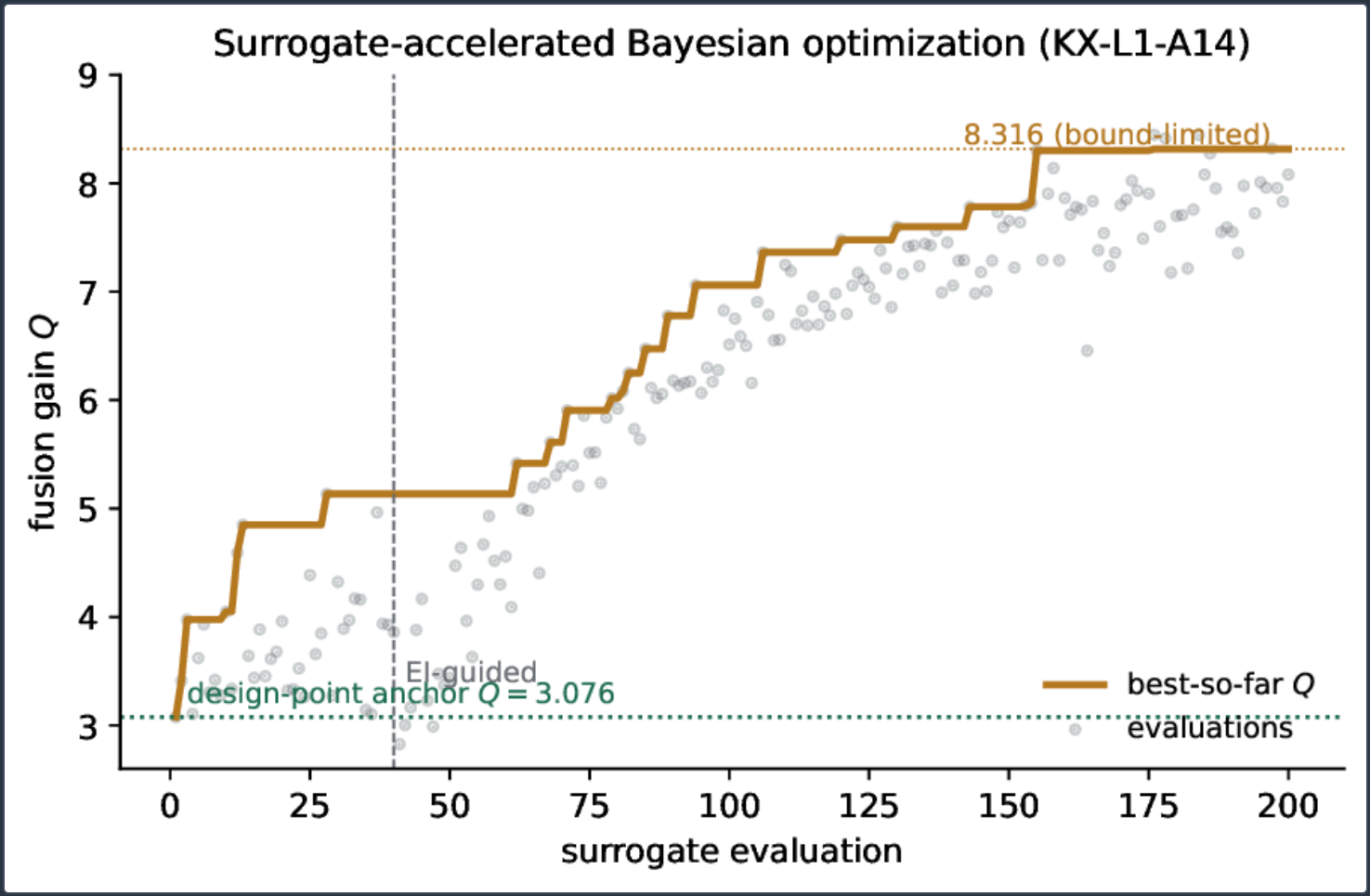
Realised clamp actuation $u(k)$ under the shield. The command stays inside $[1.061, 1.250]$ (dotted) and never reaches the available authority floor $u_{\min} = 0.5$ (dashed red); the certificate is therefore not authority-limited on this sequence — there is spare actuation between the realised command and the floor.



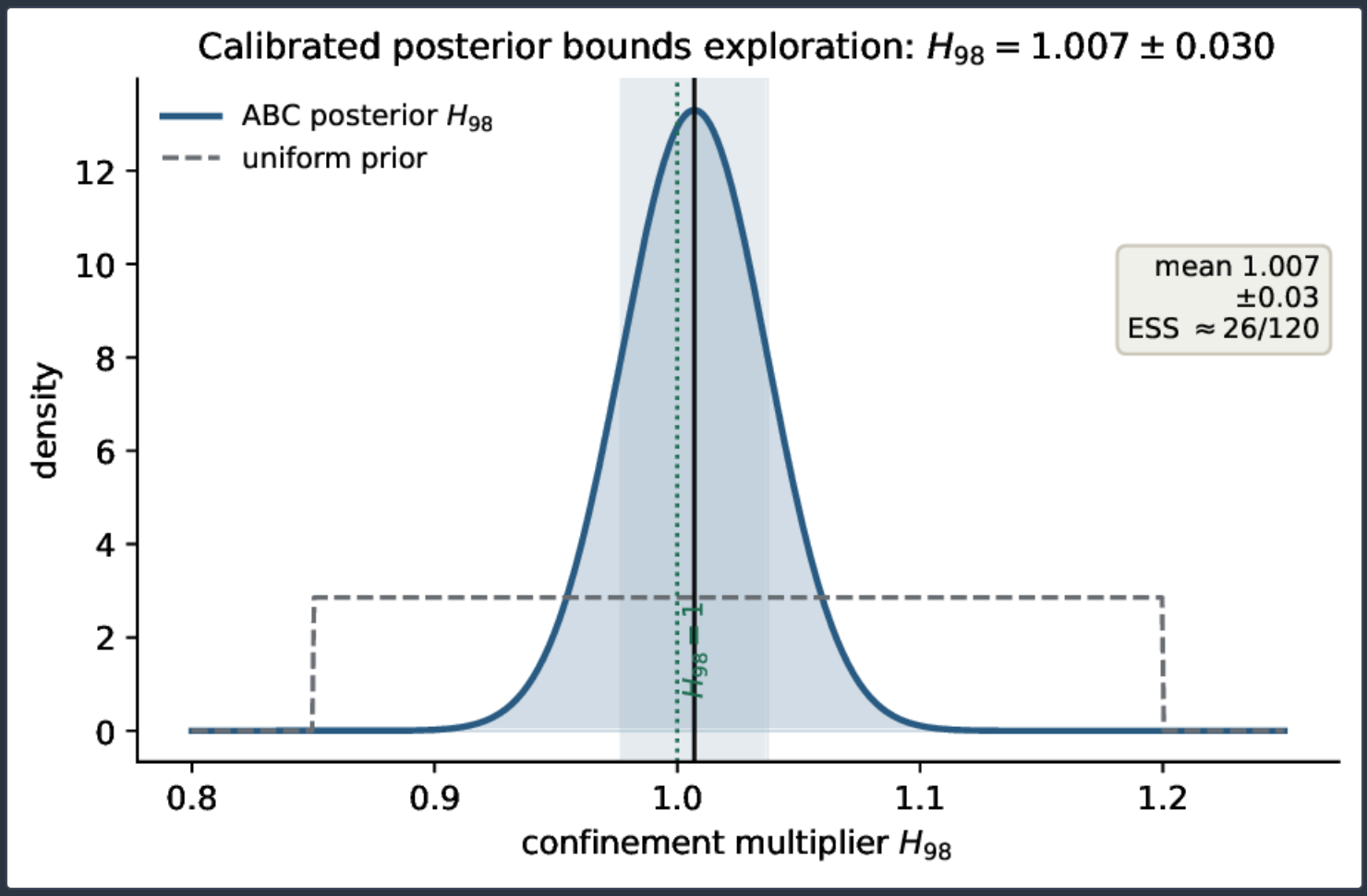
Cumulative crossings of the certified limit $\beta_N = 3.5$ along the campaign. The unshielded loop accumulates 50 crossings; the CBF-shielded loop accumulates none. The $50 \to 0$ reduction is the demonstrated safety guarantee.

CALIBRATED BOUND ON EXPLORATION

How far the policy may explore is bounded by a calibrated confinement posterior. The surrogate Bayesian-optimisation campaign (anchor KX-L1-A14) reproduces the design-point anchor exactly ($Q = 3.076$, $P_{fus} = 85.04\,MW$) and, over 200 evaluations, locates a best-so-far $Q = 8.316$ that sits on the prior bounds (Fig. 6) — a *bound-limited artefact*, not a new design point, and reported as such. The value of the campaign for safe RL is not that apparent optimum but the ABC posterior it yields on the confinement multiplier, $H_{98} = 1.007 \pm 0.030$ (effective sample size ≈ 26 of 120; Fig. 7). That calibrated ± 0.030 spread is what the safety layer reads to decide how far the learned policy may extrapolate before its authority is throttled [43,44]: a tight, well-centred posterior on $H_{98} \approx 1$ is the quantitative license for a modest exploration radius around the validated envelope.



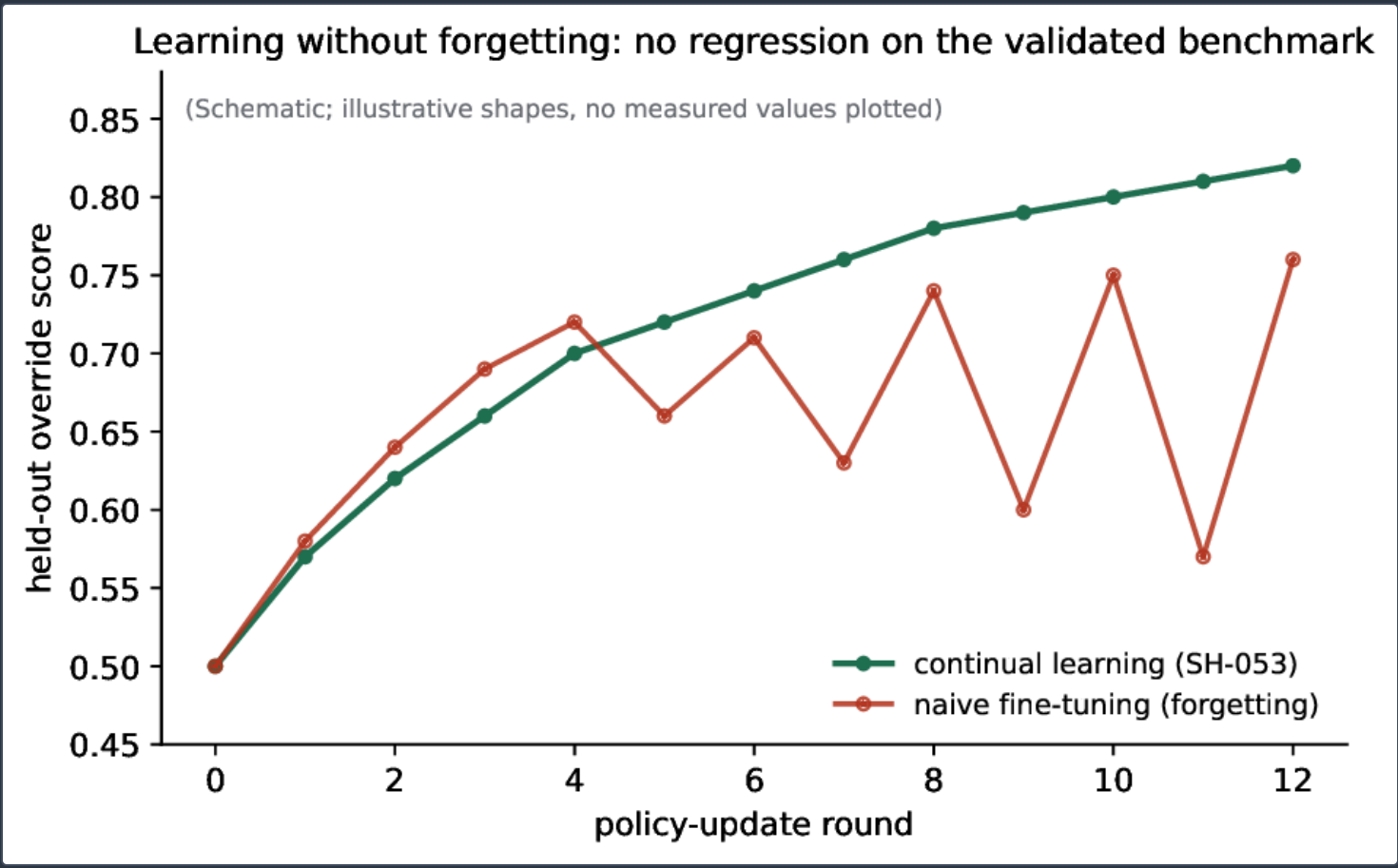
Surrogate-accelerated Bayesian optimisation (anchor KX-L1-A14): best-so-far fusion gain Q over 200 evaluations (40 initial including the design-point anchor, then expected-improvement-guided). The design-point anchor $Q = 3.076$ is reproduced; the best-so-far climbs to $Q = 8.316$ but rests on the prior bounds and is therefore a bound-limited artefact, not a certified operating point.



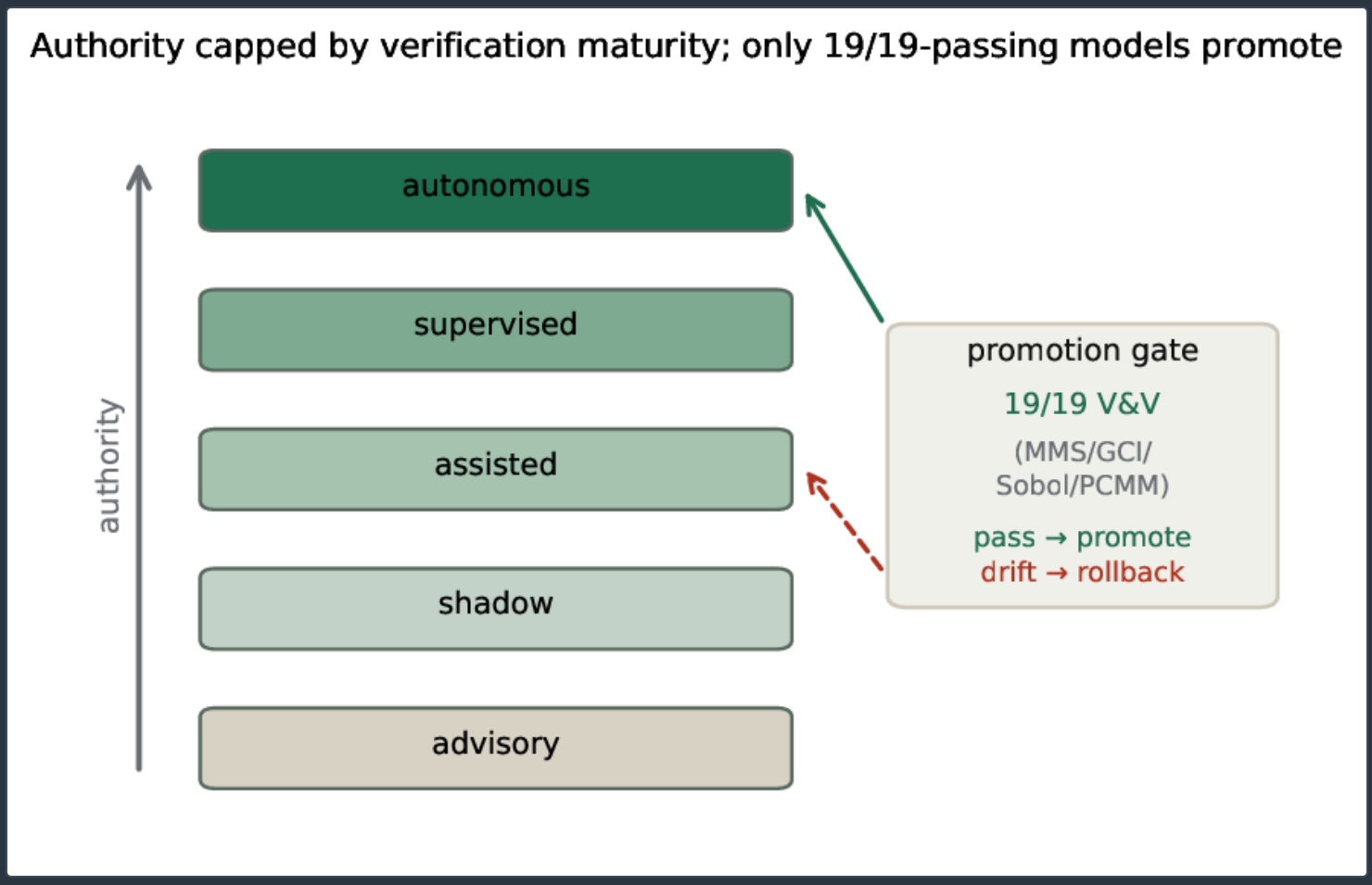
Calibrated confinement posterior from ABC (anchor KX-L1-A14). Given $Q_{\text{obs}} = 3.076$ at tolerance $\sigma = 5\%$, the posterior on the confinement multiplier is $H_{98} = 1.007 \pm 0.030$ (effective sample size ≈ 26 of 120), recovering $H_{98} \approx 1$ as required for self-consistency. This calibrated spread bounds how far the safe-RL policy may explore beyond the validated envelope.

LEARNING WITHOUT FORGETTING, AND GATED PROMOTION

Figure 8 contrasts, schematically, a continual-learning update (15) that improves held-out override performance monotonically with a naive fine-tuning update that improves then regresses (catastrophic forgetting); the acceptance test admits only the former. Figure 9 shows the maturity-gated authority ladder (16): a model climbs from advisory to autonomous authority only as it clears verification maturity, and the promotion gate admits only **19/19**-passing models, rolling back on injected drift. The learned policy, however competent, never earns authority it has not verified, and never acts outside the filter.



(Schematic.) Continual learning without forgetting (serial KFE-CF-SH-053, eq. (15)) improves held-out override performance monotonically across update rounds; naive fine-tuning improves then regresses. The acceptance criterion is no regression on the validated benchmark. Shapes are illustrative; no measured values are plotted.



(Schematic.) Maturity-gated authority (serials KFE-CF-SH-054/SH-053, eq. (16)). Runtime authority climbs from advisory to autonomous only as verification maturity accrues; the promotion gate admits only models clearing the **19/19** V&V suite (anchor KX-L1-A6, MMS/GCI/Sobol/PCMM) and rolls back on drift.

ACCEPTANCE CRITERIA AS INEQUALITIES

No model is granted authority by assertion; each clears the following pre-registered inequalities, collected here as the contract that gates the loop:

(A1) forward invariance	$h(\quad x_k) \geq 0 \;\; \forall k, \quad (\text{crossings} = 0/100)$	(KX-L1-A13, eqs.~???,???)
(A2) bounded authority	$u_k \in [u_{\min}, u_{\max}], \;\; u_k \neq u^{\text{floor}}$	(KX-L1-A13)
(A3) learn on overrides	improve on held-out $\mathcal{D}_{\text{test}}$ inside $\mathcal{K}_{\text{cbf}}$	(SH-050, eq.~???)
(A4) informative queries	$\mathbb{I}[\Theta; y]_{\text{design}} > \mathbb{I}[\Theta; y]_{\text{random}}$	(SH-052, eq.~???)
(A5) no forgetting	$\Delta(\text{validated benchmark}) \geq 0$ after update	(SH-053, eq.~???)
(A6) gated promotion	promote $\iff m = 1$ (19/19), else rollback	(SH-054/A6, eq.~???)
(A7) calibrated bound	$H_{98} = 1.007 \pm 0.030$ caps exploration radius	(KX-L1-A14)

Authority is capped by verification maturity until each bar is met.

§ 6

Discussion

WHY SEPARATING WHERE

from *how* is what makes RL admissible Prior learned tokamak controllers achieved their guarantees, where they had them, by careful reward design and extensive in-simulation training [1,2]; the safety argument was empirical — the policy did not misbehave in testing. That is a weaker claim than forward invariance. By contrast, the design here makes the safety argument *structural*: because every executed action is the projection $\Pi_{\mathcal{C}}(\pi_\theta)$ onto the admissible set (7), the closed loop cannot leave $\mathcal{C}$ for *any* θ, including a half-trained or adversarially perturbed one. The demonstrated $50 \to 0$ reduction in limit crossings, with $\min_k h = +0.000$, is the empirical footprint of that structural guarantee. This is the safe-RL-by-shielding paradigm [15,17] instantiated on the specific limit — pressure-limited β_N — that most directly threatens a compact, high-β spherical-tokamak breeder.

AGAINST THE SAFE-RL AND CONSTRAINED-CONTROL LITERATURE

Three families of safe-RL methods bear comparison. Lagrangian / constrained-policy methods (CPO and relatives [14]) bound *expected* constraint cost; they do not preclude the rare catastrophic excursion that eq. (5) forbids, which is why we enforce an almost-sure state constraint by a filter rather than a budget in the objective. Reward-shaping approaches [16] inherit the same weakness. Shielding [15] and CBF-QP filtering [17,18,19] provide the during-learning guarantee we require — as do the reachability-based and predictive safety filters that share its minimal-intervention structure [20,21]; our contribution is to fuse that filter with an operator-feedback learning signal (eqs. 10–12) and a calibrated exploration bound, and to demonstrate the composite on a reproduced fusion-control anchor. Relative to offline model-based RL for tokamak control [6], which learns entirely from logged data without any online exploration, our loop *does* explore — but only inside the certified set — and thereby retains RL's ability to discover strategies the logged data does not contain, without giving up the safety floor.

OPERATOR FEEDBACK AS THE RIGHT DATA SOURCE

Learning from the operator logbook rather than from unconstrained trial has a second virtue beyond safety: the policy improves along the trajectories operators actually use, which is where its competence is most valuable and where the twin surrogates are best validated. The override reward (11), grounded by twin replay rather than by preference alone, converts each expert correction into a physically scored lesson — the reinforcement-learning-from- human-feedback recipe [11,12] adapted to a setting where the "human preference" can be checked against physics before it is trusted. The active-learning designer (14) then spends the scarce budget of informative queries where they most reduce posterior uncertainty [24,48], and the continual-learning penalty (15) ensures each such lesson is retained without erasing an older one.

§ 7

Limitations

One honest gate remains. The override reward (11) is at present shaped from twin-synthetic operator corrections; grounding it in the accumulating *live* operator logbook is the one confirming step, and it is the natural first use of operating experience as it accrues. Separately, the demonstrated β_N dynamics are a one-dimensional, design-point-anchored surrogate: the forward-invariance certificate is proven at that reduced level (over **100** explicit-Euler steps), and its extension to the full coupled equilibrium–transport plant is the subject of the reference-governor MPC and four-twin architecture developed in the companion papers [42,41]. The safety shield that makes exploration safe to run is demonstrated now, independent of both steps.

§ 8

Conclusion

Kronos safe reinforcement learning improves the controller from operator feedback while confining every exploratory action to a control-barrier-certified safe set. The design decouples the two questions safe learning usually conflates: *where* the policy may act is decided once, by a CBF-certified set that no action may leave, and *how* it improves inside that set is decided by learning from the operators — imitation from a hash-chained logbook and a twin-scored override reward, disciplined by an active-learning designer that beats random selection, continual learning without forgetting, and gated promotion with rollback. The safety guarantee is demonstrated, not promised: zero limit crossings in a 100-step exploration campaign against 50 unshielded, forward invariance verified with $\min_k h = +0.000$, and a bounded actuation $u \in [1.061, 1.250]$ that never reaches the $u_{\min} = 0.5$ floor, all bounded by a calibrated $H_{98} = 1.007 \pm 0.030$ posterior. The controller gets better from the people who run it without ever leaving the set that is proven safe.

Acknowledgments tocsectionAcknowledgments Part of the Kronos Fusion Energy 2026 design series. The author thanks the Kronos physics de-risking programme for the frozen, independently reproduced result set underlying every quantitative claim, and the KRONOS-CTRL computing-and-control effort for the codes and the NVIDIA GPU HPC campaign.

Reproducibility. The safe-learning anchors are frozen entries in the Kronos Physics De-Risking Register [49] (doi:10.5281/zenodo.22133057): the CBF shield (KX-L1-A13: β_N $4.200 \to 3.500$, $50 \to 0$ crossings, $\min_k h = +0.000$, $u \in [1.061, 1.250]$), the surrogate Bayesian-optimisation calibration (KX-L1-A14: $H_{98} = 1.007 \pm 0.030$), and the formal V&V suite (KX-L1-A6: $19/19$). Each is regenerable from the archived closed-loop, UQ and verification runs under a single global seed (20260726), and every figure is regenerated by the deposited script `make_fig_2p81_safe_rl.py` from these anchors. This paper is archived at its DOI doi:10.5281/zenodo.22645823 and its official page kronosfusionenergy.com/publications/safe-reinforcement-learning-from-operator-feedback-for-.

Data availability tocsectionData availability This paper is archived at doi:10.5281/zenodo.22645823. The shield definition, the operator-feedback learning objectives, the operator-logbook schema and the figure-generation script are archived with the deposit and reference the Kronos 2026 series, including the AI and quantum-control overview [50] (doi:10.5281/zenodo.22645702, official page). All quantitative values trace to the register [49] by the serial and anchor 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

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 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

4 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

5 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

6 I Char *et al*, Offline model-based reinforcement learning for tokamak control, *Proc. 5th Learning for Dynamics and Control Conf. (L4DC)*, PMLR **211** 1357 (2023).

7 R S Sutton and A G Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, Cambridge MA (2018).

8 J Schulman, F Wolski, P Dhariwal, A Radford and O Klimov, Proximal policy optimization algorithms, arXiv:1707.06347 (2017).

9 T Haarnoja, A Zhou, P Abbeel and S Levine, Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor, *Proc. 35th Int. Conf. on Machine Learning (ICML)*, PMLR **80** 1861 (2018). arXiv:1801.01290

10 S Fujimoto, H van Hoof and D Meger, Addressing function approximation error in actor-critic methods (TD3), *Proc. 35th Int. Conf. on Machine Learning (ICML)*, PMLR **80** 1587 (2018). arXiv:1802.09477

11 P Christiano, J Leike, T B Brown, M Martic, S Legg and D Amodei, Deep reinforcement learning from human preferences, *Adv. Neural Inf. Process. Syst. (NeurIPS)* **30** (2017). arXiv:1706.03741

12 L Ouyang *et al*, Training language models to follow instructions with human feedback, *Adv. Neural Inf. Process. Syst. (NeurIPS)* **35** 27730 (2022). arXiv:2203.02155

13 S Ross, G J Gordon and J A Bagnell, A reduction of imitation learning and structured prediction to no-regret online learning, *Proc. 14th Int. Conf. on Artificial Intelligence and Statistics (AISTATS)*, PMLR **15** 627 (2011). arXiv:1011.0686

14 J Achiam, D Held, A Tamar and P Abbeel, Constrained policy optimization, *Proc. 34th Int. Conf. on Machine Learning (ICML)*, PMLR **70** 22 (2017). arXiv:1705.10528

15 M Alshiekh, R Bloem, R Ehlers, B K\"onighofer, S Niekum and U Topcu, Safe reinforcement learning via shielding, *Proc. 32nd AAAI Conf. on Artificial Intelligence* **32**(1) 2669 (2018). doi:10.1609/aaai.v32i1.11797

16 J Garc\'ia and F Fern\'andez, A comprehensive survey on safe reinforcement learning, *J. Mach. Learn. Res.* **16** 1437 (2015).

17 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**(8) 3861 (2017). doi:10.1109/TAC.2016.2638961

18 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

19 R Cheng, G Orosz, R M Murray and J W Burdick, End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks, *Proc. 33rd AAAI Conf. on Artificial Intelligence* **33**(01) 3387 (2019). doi:10.1609/aaai.v33i01.33013387

20 J F Fisac, A K Akametalu, M N Zeilinger, S Kaynama, J Gillula and C J Tomlin, A general safety framework for learning-based control in uncertain robotic systems, *IEEE Trans. Autom. Control* **64**(7) 2737 (2019). doi:10.1109/TAC.2018.2876389

21 K P Wabersich and M N Zeilinger, A predictive safety filter for learning-based control of constrained nonlinear dynamical systems, *Automatica* **129** 109597 (2021). doi:10.1016/j.automatica.2021.109597

22 L Brunke, M Greeff, A W Hall, Z Yuan, S Zhou, J Panerati and A P Schoellig, Safe learning in robotics: from learning-based control to safe reinforcement learning, *Annu. Rev. Control Robot. Auton. Syst.* **5** 411 (2022). doi:10.1146/annurev-control-042920-020211

23 J Kirkpatrick *et al*, Overcoming catastrophic forgetting in neural networks, *Proc. Natl. Acad. Sci. USA* **114**(13) 3521 (2017). doi:10.1073/pnas.1611835114

24 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**(1) 148 (2016). doi:10.1109/JPROC.2015.2494218

25 C E Rasmussen and C K I Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge MA (2006).

26 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

27 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

28 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

29 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

30 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

31 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

32 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

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

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

35 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

36 B N Sorbom *et al*, ARC: a compact, high-field, fusion nuclear science facility and demonstration power plant with demountable magnets, *Fusion Eng. Des.* **100** 378 (2015). doi:10.1016/j.fusengdes.2015.07.008

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

38 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

39 A J Creely *et al*, Overview of the SPARC tokamak, *J. Plasma Phys.* **86** 865860502 (2020). doi:10.1017/S0022377820001257

40 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; official page.

41 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; official page.

42 P I Ford, Model-predictive burn control inside a certified safe envelope: a reference-governor MPC bounded by a control-barrier-function safety clamp, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645807; official page.

43 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; official page.

44 P I Ford, Out-of-distribution detection and epistemic gating for autonomous fusion control: a machine-readable extrapolation ledger that throttles model authority beyond the validated envelope, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645819; official page.

45 P I Ford, Cross-machine transfer learning for a compact spherical-tokamak breeder: a disruption model pre-trained on TCV, DIII-D, and NSTX with honest transfer uncertainty, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645821; official page.

46 P I Ford, MLOps for a safety-critical fusion control stack: versioned lineage, gated model promotion, and physics-balanced multi-GPU training for a compact spherical-tokamak breeder, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645827; official page.

47 P I Ford, The human-machine layer for autonomous fusion operation: natural-language-supervised control, autonomous root-cause analysis and a digital operator logbook, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645909; official page.

48 P I Ford, Bayesian optimal experimental design and multi-fidelity data fusion: prioritizing the next experiment across reduced-order, GPU-HPC and hardware data, Kronos Fusion Energy 2026 series. doi:10.5281/zenodo.22645895; official page.

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

50 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

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

Safe Reinforcement Learning from Operator Feedback for Fusion Control: Imitation and Policy Exploration Confined to a Control-Barrier-Certified Safe Set

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

Zenodo · doi:10.5281/zenodo.22645823

Read on kronosfusionenergy.com

The Physics De-Risking Register