

ENO-Chain: Energy-Neutral Scheduling of Blockchain Participation for Energy-Harvesting IoT Nodes in Smart Grids

Nawaf Abdualaziz Almolhis

Jazan University

Ashish Dixit

Ajay Kumar Garg Engineering College

Subuhi Kashif Ansari

Jazan University

Huda Fatima

Jazan University

Shamimul Qamar

sqamar@kku.edu.sa

King Khalid University

Salahaldeen Duraibi

Jazan University

Ashwag Alasmari

King Khalid University

Salman Arafath Mohammed

King Khalid University

Research Article

Keywords: Energy-neutral operation, energy harvesting, blockchain consensus, smart grid, Lyapunov optimization, sustainable computing

Posted Date: October 16th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11241221/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

ENO-Chain: Energy-Neutral Scheduling of Blockchain Participation for Energy-Harvesting IoT Nodes in Smart Grids

Nawaf Abdulaziz Almolhis¹, Ashish Dixit², Subuhi Kashif Ansari¹, Huda Fatima¹, Shamimul Qamar^{3*}, Salahaldeen Duraibi⁴, Ashwag Alasmari⁵, Salman Arafath Mohammed⁶

¹ Department of Computer Science, College of Engineering & Computer Science, Jazan University, Jazan, Saudi Arabia (naalmolhis@jazanu.edu.sa ; subuhiwasim@jazanu.edu.sa , hsaadullah@jazanu.edu.sa)

² Ajay Kumar Garg Engineering College Ghaziabad Uttar Pradesh India -201015, dixitashish@akgec.ac.in

³ Computer Science & Engineering, Applied College, Dhahran Al Janoub Campus, King Khalid University, Abha, Saudi Arabia; Email ID : sqamar@kku.edu.sa

⁴ Department of Electrical and Electronics Engineering, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia, Email sduraibi@jazanu.edu.sa

⁵ Department of Informatics and Computing Systems and Centre for Artificial Intelligence, King Khalid University, Abha, Saudi Arabia, Email: aasmry@kku.edu.sa

⁶ Electrical Engineering Department , Computer Engineering Section, College of Engineering, King Khalid University, Abha, KSA. email: salman@kku.edu.sa

Correspondence : Shamimul Qamar (sqamar@kku.edu.sa)

Abstract

Advanced metering and distributed energy resource (DER) nodes in smart grids are increasingly powered by ambient energy harvesting rather than mains supply, which subjects their workload to the energy-neutral operation (ENO) constraint: long-run consumption cannot exceed long-run harvested income without battery depletion. These nodes are also being asked to participate in blockchain ledgers for tamper-evident metering and peer-to-peer settlement, a workload existing work treats as a static per-protocol attribute rather than a state-dependent quantity, with no guarantee against depletion under stochastic harvesting. This paper formulates blockchain participation intensity as a discrete control variable inside the ENO battery-drift model and proposes ENO-Chain: a Lyapunov drift-plus-penalty policy with a closed-form battery-deviation bound, optionally refined by a lightweight actor-critic layer restricted to the Lyapunov-certified safe region. Calibrating the trade-off parameter V on a trace held out from every evaluation seed and averaging over eight independent traces, the Lyapunov layer alone sustains zero violations at a mean utility 21% above a hand-tuned static-threshold baseline, and far above an energy-oblivious baseline that violates at 1.49 downgrade events per slot. The learned refinement layer is not a free addition: at this operating point it introduces a small utility loss and a non-zero violation rate, a trade-off that reverses once the layer is trained across weather regimes instead of one. A penalty-sensitivity sweep, a population-level reimplementations of the closest prior baseline, and tests across four weather regimes, a full synthetic year, and network sizes from 5 to 200 nodes under validator-slot contention complete the evaluation.

Keywords: Energy-neutral operation; energy harvesting; blockchain consensus; smart grid; Lyapunov optimization; sustainable computing

1. Introduction

Global smart meter installations are projected to grow from 2.1 billion units in 2025 to 3.9 billion by 2035, with electricity meters accounting for roughly two-thirds of that base and smart metering already representing close to 11% of all IoT connections worldwide [1]. This buildout is not a peripheral IoT use case: utilities depend on it for demand-side visibility, loss reduction, and the measurement infrastructure that renewable integration and demand response require, and the same connectivity trend extends to distributed energy resources more broadly, as solar, storage, and flexible load assets are increasingly metered and settled through networked devices rather than manual reads. At a deployment scale approaching billions of units, even small per-device inefficiencies in energy management or security overhead compound into infrastructure-level cost and reliability consequences.

A growing share of this installed base cannot rely on mains power or scheduled battery replacement: meters at the edge of rural feeders, DER controllers on remote assets, and sensors on infrastructure that is costly or unsafe to visit routinely are increasingly specified with solar or vibration energy harvesting instead. Harvesting-powered operation imposes a constraint absent from mains-powered deployments: a device's average power draw must track its average harvested income, or it either exhausts its battery and drops offline or wastes energy it has no capacity to store – the energy-neutral operation (ENO) criterion formalized by Kansal et al. [2]. At the same time, these same devices are being asked to take on a workload unrelated to sensing: participation in blockchain ledgers that provide tamper-evident metering records and support peer-to-peer settlement between prosumers, a requirement driven by trust and auditability needs in transactive energy markets rather than by anything intrinsic to metering itself. Despite the importance of both requirements, achieving reliable blockchain participation under a stochastic energy-harvesting budget remains unaddressed, because consensus protocols were not designed against a battery constraint and energy-harvesting schedulers were not designed against a security-relevant workload.

Prior work addressing pieces of this problem falls into three categories. Energy-neutral scheduling research optimizes a device's duty cycle or sensing rate against a stochastic harvesting process, from the original battery-drift heuristics of Kansal et al. [2] through iterative runtime allocation [3] and reinforcement-learning energy managers [4,5], extending more recently to MAC-layer and network-level duty-cycling protocols [6,7,8,9]. Energy-aware blockchain consensus research, separately, targets the compute cost of ledger participation on constrained hardware: edge-offloaded proof-of-work variants [10], validator-selection schemes that use a device's instantaneous energy reading as a selection criterion [11], and systematic reviews cataloguing proof-of-stake, delegated, and hybrid consensus as lower-energy alternatives to proof-of-work [12,13]. A third body of work reports complete blockchain-secured smart-grid systems – trading platforms, authentication frameworks, and metering ledgers [11,14,15] – evaluated on throughput, trading efficiency, or security rather than on the energy budget of the device running the client.

The closest of these to the present problem is the validator-selection framework of Saleh and Cevik [11], which assigns consensus roles using real-time energy readings and trust scores and reports strong throughput and authentication accuracy. Selection of this kind, however, operates at a single point in time across a population of nodes, not as a temporal policy for one node's own stochastic battery trajectory, and

it does not model harvesting income at all, so nothing prevents a selected validator from being asked to sustain a role its own battery cannot support through the slot. The energy-neutral scheduling literature has the complementary gap: it models the battery trajectory correctly but has never attached a security-relevant, externally-verified utility term to the controlled workload, and none of its reported guarantees have been evaluated where the controlled action is itself a consensus role. Reviews of consensus energy efficiency [12,13] and a knowledge-graph survey of distributed ledger technology for IoT [16] each note, independently, that a ledger's own device-level energy footprint is largely absent from the empirical record. However, none of these methods jointly address a stochastic harvesting budget and a blockchain-participation workload with a device-level guarantee against battery depletion, particularly under the weather variability and network-scale contention a real deployment would face.

This paper makes four contributions.

- We formulate blockchain participation – signing, relaying, and local validation – as a discrete control variable inside the canonical ENO battery-drift model, extending a framework previously restricted to sensing and communication workloads to a workload with an externally defined security utility and cost structure.
- We propose ENO-Chain: a Lyapunov drift-plus-penalty policy that requires no advance knowledge of the harvesting distribution and admits a closed-form $O(1/V)$ – $O(V)$ battery-deviation/utility trade-off, together with a lightweight actor-critic fine-tuner restricted by construction to the Lyapunov-safe action set – a shielding pattern adapted from the safe-RL literature and instantiated here from an ENO-specific Lyapunov argument.
- We calibrate the trade-off parameter on a trace disjoint from every evaluation seed, evaluate across eight independent seeds, run a penalty-sensitivity sweep, and reimplement the closest prior baseline at the population level it was originally proposed for, rather than as a single-node proxy.
- We identify a case where the learned refinement underperforms its own safety layer once evaluated rigorously, trace it with a mixture-regime training control to a training-distribution effect rather than a fundamental limitation, and report where the method's advantage is and is not supported by the evidence.

The remainder of this paper is organized as follows. Section 2 reviews the three literatures at issue plus the safe-RL pattern the controller adapts, and Section 3 synthesizes the resulting gap into a single comparison table. Section 4 states the system model and problem formulation. Section 5 develops the ENO-Chain controller. Section 6 specifies the digital-twin simulation, including the calibration protocol fixes described above. Section 7 reports results against five baselines across weather regimes, battery capacities, the Lyapunov trade-off parameter, network scale, and forecast quality, together with the calibration, penalty-sensitivity, and mixture-training checks. Section 8 discusses limitations, and Section 9 concludes.

2. Related work

2.1. Energy-neutral operation and harvesting-aware scheduling

The ENO criterion – long-run consumption at or below long-run harvested income – was formalized by Kansal et al. [2], who derived duty-cycle adaptation rules from a stored-energy state and a harvesting-rate estimate. Subsequent work has refined the estimator and the control law rather than the workload model. ECO [3] replaces static duty-cycle rules with an iterative runtime allocation algorithm, validated against solar-charging traces from 4772 users, that reallocates a device's energy budget within a day as harvesting

deviates from forecast; the controlled quantity remains a sensing/logging rate. tinyMAN [4] and the SARSA(λ)-based manager of Shresthamali et al. [5] replace the control law with reinforcement learning, reporting root-mean-square deviation from exact ENO as the primary metric, again over a duty-cycle action space. More recent work extends ENO to the MAC and network layer: HENO-MAC [6] adds hybrid solar-wind harvesting and reports delay reduction of 27–28% relative to prior MAC protocols under realistic two-source traces; a Markov-chain and k-nearest-neighbor duty-cycling scheme [7] and a Q-learning-based wake-up controller [8] extend the same criterion to correlated-activity and multi-device settings; a 2026 study [9] couples ENO duty-cycling with sensor deployment layout using a hybrid linear-programming-plus-RL solver. Across this line of work the controlled action is always a sensing, transmission, or deployment parameter, and the utility function never includes a security or consensus-relevant term – Table 1 makes this pattern explicit.

Table 1. Energy-neutral and harvesting-aware scheduling works.

Reference	Year	Control approach	Controlled workload	Chain-aware
Kansal et al. [2]	2007	Battery-drift heuristic	Duty cycle	No
Shresthamali et al. [5]	2017	SARSA(λ) RL	Duty cycle	No
ECO [3]	2021	Iterative allocation	Sensing/logging rate	No
tinyMAN [4]	2022	Lightweight RL	Duty cycle	No
HENO-MAC [6]	2024	Hybrid-source MAC control	Duty cycle	No
Ruiz-Guirola et al. [7]	2024	Markov chain + k-NN	Duty cycle	No
Wake-up manager [8]	2025	Q-learning	Duty cycle / wake-up	No
Coverage-ENO [9]	2026	LP + RL	Deployment + duty cycle	No
ENO-Chain (this work)	–	Lyapunov (+ actor-critic)	Blockchain participation	Yes

2.2. Energy-aware and lightweight blockchain consensus for constrained IoT

A parallel literature addresses the energy cost of blockchain consensus on constrained hardware without reference to harvesting. Wadhwa et al. [10] offload the mining task in a PoW-derived scheme to an edge node selected by device specification, reducing the constrained node's own compute load but leaving its participation decision static once a miner is chosen. Systematic reviews of consensus energy efficiency [12,13], synthesizing more than 50 primary studies published between 2018 and 2024, catalogue proof-of-stake variants, delegated and hybrid consensus, and architectural mitigations such as sharding, and both reviews identify the absence of empirically validated carbon-footprint accounting as an open problem, but neither reports a scheme that conditions consensus participation on a device's own energy state. Within the smart-grid application domain specifically, Saleh and Cevik [11] select validators using real-time energy readings and trust scores in a deep-neural-network-assisted authentication framework, reporting 372 transactions per second, 98.69% authentication accuracy, and an 18% reduction in validator energy relative to unspecified baselines; this is, to our knowledge, the closest prior work to the present paper, and Section 6 reimplements its selection mechanism at the population level it targets rather than treating it as a single-node proxy. Elhaji et al. [14] combine a hybrid proof-of-stake/practical-Byzantine-fault-tolerance framework with k-means demand clustering, optimizing trading efficiency using real consumption data (UK-DALE, Pecan Street) but assuming continuously available mains power at the metering endpoint.

Haider et al. [17] propose a neuromorphic, spiking-neural-network consensus mechanism reporting substantial energy savings over proof-of-burn and proof-of-reputation baselines, evaluated as a general-purpose ledger primitive independent of any application domain or harvesting model. Table 2 summarizes this cluster.

Table 2. Energy-aware blockchain consensus for constrained / smart-grid IoT.

Reference	Year	Consensus family	Energy treatment
Wadhwa et al. [10]	2022	PoW, edge-offloaded	Static, per-protocol
Rammohan et al. [12]	2025	Multiple (survey)	Per-protocol, no device state
Discover Analytics review [13]	2025	PoW, PoS, DPoS, BFT, DAG	Per-protocol, no device state
Saleh and Cevik [11]	2025	Energy-aware validator selection	Snapshot reading, population-level
Elhajj et al. [14]	2025	Hybrid PoS+PBFT	Assumes mains power
Haider et al. [17]	2025	Neuromorphic (SNN)	Per-protocol, domain-agnostic
ENO-Chain (this work)	–	Lightweight, role-adaptive	Stochastic, per-node, temporal

2.3. Blockchain-secured smart-grid IoT system papers

A third cluster reports complete blockchain-secured smart-grid systems and evaluates them primarily on trading efficiency, throughput, or security rather than device-level energy management. Bensalah et al. [15] implement a proof-of-concept on Orange Pi 4 hardware and report 7.45–7.81 transactions per second with processing time increasing linearly in block size, explicitly flagging consensus optimization as future work; the platform is mains-powered. Broader transactive-energy and metering-blockchain designs report token systems, smart-contract-mediated peer-to-peer trading, and IPFS-backed storage for scalability, uniformly under the assumption of a reliably powered metering endpoint. None of the system papers identified model an energy-harvesting deployment, and the survey-level evidence [16] explicitly flags this omission as an open question at the level of the ledger's own footprint rather than the grid's. Table 3 lists representative systems and their power assumption.

Table 3. Blockchain-secured smart-grid IoT system papers.

Reference	Year	Power assumption	Primary evaluation axis
Bensalah et al. [15]	2025	Mains (Orange Pi 4)	Throughput vs. block size
Elhajj et al. [14]	2025	Mains	Trading efficiency, forecasting
Saleh and Cevik [11]	2025	Mains (implicit)	Authentication accuracy, latency
DLT-IoT survey [16]	2024	N/A (survey)	Identifies ledger-footprint gap
ENO-Chain (this work)	–	Harvesting (stochastic)	Neutrality, utility, scalability jointly

2.4. Safe reinforcement learning and shielding

A structurally different literature – safe reinforcement learning under constrained Markov decision processes – addresses a problem this paper's controller borrows from without belonging to the three clusters

above: how to let a learned policy explore and improve while a separate mechanism guarantees a safety property regardless of what the policy has learned. Shielding [18] and safety-layer approaches [19] filter or correct a learned policy's actions against a pre-computed or model-based safe set; Kushwaha et al. [20] survey the broader field. The controller developed in Section 5 follows this pattern – a certified-safe action set gates a learned refinement – applied to a domain, and derived from a domain-specific Lyapunov argument, that this literature does not itself consider. This is not a claim of novelty in the shielding pattern itself; the contribution is formulating the safe set from the ENO battery-drift dynamics of a specific workload, and the empirical finding, detailed in Section 7, that shielding's usual trade-off (safety versus how much the learned layer can help) plays out here in a more fragile way than the pattern's original applications would suggest. Because this literature does not target the same joint problem as Section 2's three clusters, it is not included as a fourth row in the gap comparison of Section 3.

3. Gap synthesis

Table 4 cross-tabulates the three clusters against five capabilities that a harvesting-powered, blockchain-secured smart-grid node requires jointly: (i) an explicit stochastic model of harvested energy income, (ii) treatment of blockchain participation as a controllable variable rather than a fixed protocol attribute, (iii) a formal guarantee against battery depletion or overflow, (iv) validation in a smart-grid-specific setting, and (v) a comparison protocol that spans both the ENO and the blockchain-consensus literatures rather than baselines drawn from only one. No surveyed method satisfies more than two of the five criteria simultaneously. We do not read this table as evidence of novelty on its own – a gap table whose criteria are chosen by the authors will tend to favor the authors' own method by construction – but as a compact statement of which specific combination of properties the rest of the paper argues for and tests, criterion by criterion, rather than asserts.

Table 4. Gap synthesis across literature clusters.

Cluster	(i)	(ii)	(iii)	(iv)	(v)
ENO / harvesting scheduling (§2.1)	✓	–	Emp. only	–	–
Energy-aware BC consensus (§2.2)	–	Partial	–	Partial	–
BC-secured grid systems (§2.3)	–	–	–	✓	–
ENO-Chain (this work)	✓	✓	Formal (Lyapunov layer)	✓	✓

4. System model and problem formulation

Consider a smart-grid IoT node i (a smart meter or DER controller) operating in discrete time slots $t = 1, 2, \dots$ of duration matching the AMI reporting interval (typically 1–15 minutes). The node is equipped with an ambient energy harvester that supplies a stochastic, non-negative energy income $h_i(t)$ per slot, and a battery of capacity $B_{\max}$ whose state evolves as

$$B_i(t+1) = \left[B_i(t) + h_i(t) - c_{\text{sense}} - a_i(t) E_{\text{bc}}(a_i(t), t) \right]_0^{B_{\text{max}}} \quad (1)$$

where $[\cdot]^{\wedge \{B_{\text{max}}\}}$ denotes clipping to $[0, B_{\text{max}}]$ and c_{sense} is the fixed per-slot metering cost. The control decision this paper introduces is $a_i(t) \in \mathcal{A}$, the intensity of the node's blockchain participation for slot t , where

$$\mathcal{A} = \{\text{sleep}, \text{sign}, \text{sign+relay}, \text{sign+relay+validate}\} \quad (2)$$

is a four-level action set with strictly increasing energy cost $E_{\text{bc}}(a_i(t), t)$ across the four levels and a corresponding, application-defined ledger utility $U(a_i(t), t)$ that increases with participation intensity (e.g., transaction freshness, contribution to consensus liveness, or eligibility for a validator reward). $U(\cdot)$ is a modeling proxy for these effects rather than a measured quantity from any deployed protocol; every utility figure reported in Section 7 should be read on this proxy scale, and Section 8 returns to what that does and does not license us to claim.

A node is in energy-neutral operation over horizon T if

$$\frac{1}{T} \sum_{t=1}^T \left(c_{\text{sense}} + a_i(t) E_{\text{bc}}(a_i(t), t) \right) \leq \frac{1}{T} \sum_{t=1}^T h_i(t) + o(1) \quad (3)$$

consistent with the criterion of Kansal et al. [2], extended here to include the blockchain term. The scheduling problem is

$$\begin{aligned} \max_{\{a_i(t)\}} \quad & \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}[U(a_i(t), t)] \\ \text{s.t.} \quad & \text{Eqs. (1), (2), } 0 \leq B_i(t) \leq B_{\text{max}} \end{aligned} \quad (4)$$

Problem (3) departs from the classical ENO formulation in one respect that determines the rest of the paper's design: the controlled variable $a_i(t)$ is not a private quality-of-service parameter (as sensing rate or duty cycle are) but a role in a distributed protocol whose liveness and security depend on the aggregate behavior of many nodes, not on node i alone. This rules out control laws that optimize node i 's battery trajectory without reference to a ledger-facing utility term, which is precisely the simplification made throughout Section 2, and motivates a controller that treats $U(a_i(t), t)$ as a first-class, non-separable term rather than a proxy for sensing quality.

5. Proposed method: ENO-Chain

ENO-Chain solves (3) with a two-layer controller (Fig. 1) that maps a node's battery state and short-horizon harvest forecast to an action determining its role in the blockchain overlay – light node, validator/edge gateway, or ledger – which in turn feeds the smart-grid data layer, closing feedback loops through energy draw on the battery and telemetry freshness from the data layer. The first layer is a Lyapunov drift-plus-penalty policy that guarantees bounded battery deviation without requiring the distribution of $h_i(t)$ in advance; this layer alone is what Section 7 finds to carry the paper's headline safety result once evaluation is done correctly.

The second layer is a lightweight actor-critic fine-tuner that refines the first layer's decision using short-horizon forecast information, subject to a hard constraint that it cannot select an action outside the set the first layer has certified as safe for the current slot – the shielding pattern discussed in Section 2.4 [18,19], instantiated here from an ENO-specific Lyapunov argument rather than a general safety specification. The two-layer split is what lets the controller report a bound that does not depend on how well the learned layer has converged; it does not, on its own, guarantee that the learned layer improves on the first layer, and Section 7 reports both directions of that comparison.

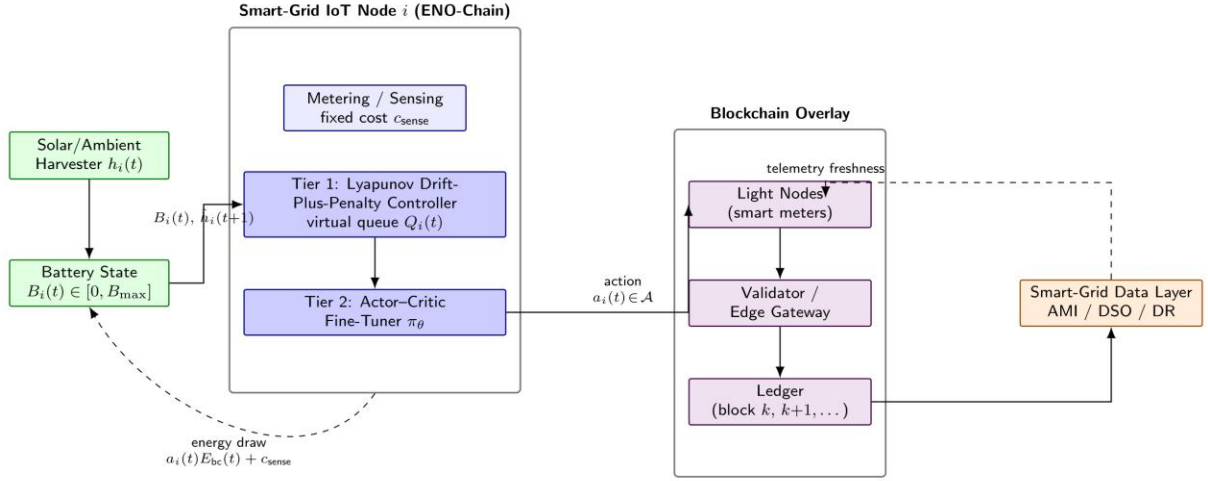


Figure 1. ENO-Chain system architecture: an energy-harvesting smart-grid node's two-layer controller, blockchain overlay, and smart-grid data layer.

5.1. Layer 1: Lyapunov drift-plus-penalty control

Define a virtual energy-deficit queue $Q_i(t)$ that accumulates whenever consumption exceeds harvested income:

$$Q_i(t+1) = \left[Q_i(t) + c_{\text{sense}} + a_i(t)E_{\text{bc}}(a_i(t), t) - h_i(t) \right]_0^\infty \quad (5)$$

$Q_i(t)$ is a standard device for converting the long-run average constraint (3) into a per-slot stabilizability condition [21]: if a control policy keeps $Q_i(t)$ mean-rate-stable, the time-averaged constraint (3) holds. Let $L(Q_i(t)) = \frac{1}{2}Q_i(t)^2$ and define the one-slot conditional drift-plus-penalty term

$$\begin{aligned} \Delta_V(t) = & \mathbb{E}[L(Q_i(t+1)) - L(Q_i(t)) \mid Q_i(t)] \\ & - V \cdot \mathbb{E}[U(a_i(t), t) \mid Q_i(t)] \end{aligned} \quad (6)$$

Layer 1 greedily minimizes an upper bound on (6) at every slot, which reduces to the per-slot decision rule

$$\begin{aligned} a_i^{L1}(t) = & \arg \min_{a \in \mathcal{A}} Q_i(t)(c_{\text{sense}} + a E_{\text{bc}}(a, t) - h_i(t)) \\ & - V U(a, t) \end{aligned} \quad (7)$$

which requires only the current queue state, the current harvest reading, and the (assumed known) cost/utility of each of the four discrete actions – no distributional assumption on $h_i(t)$ is needed. Standard Lyapunov optimization results [21] give, for bounded per-slot cost and utility increments, a time-averaged

queue backlog bound $O(V)$ and a utility gap to the offline optimum of $O(1/V)$, so V is the single parameter through which an operator trades battery-deviation risk against ledger utility. Section 6 describes how V is calibrated without letting the evaluation data leak into that choice, and Section 7 reports the empirical trade-off curve this bound predicts qualitatively. Because (7) is evaluated over a four-element action set, its per-slot cost is a small constant number of arithmetic comparisons, consistent with execution on a constrained metering microcontroller.

5.2. Layer 2: constrained actor-critic fine-tuning

Rule (7) is myopic: it reacts to the current queue and harvest reading but cannot exploit a short-horizon forecast $\hat{h}_i(t+1:t+H)$ or recent load-pattern structure in the pending-transaction backlog, both of which are available to a smart-grid node from weather forecasts and AMI reporting schedules. Layer 2 is a compact actor-critic policy $\pi_\theta(a \mid s_i(t))$, sized for constrained hardware in the manner of [4], with state $s_i(t)$ combining battery state, virtual queue, forecast, and recent realized utility. Layer 2 is trained by policy gradient with reward equal to the realized utility $U(a_i(t), t)$, restricted to the safe set that Layer 1 certifies each slot as a fractional trust region around its own decision rule, following the shielding pattern of [18,19]:

$$\mathcal{A}_{\text{safe}}(t) = \{a \in \mathcal{A} : |\Delta_V(a) - \Delta_V(a_i^{L1}(t))| \leq \rho \cdot \text{spread}\} \quad (8)$$

where $\rho \in [0,1]$ and $\text{spread} = \max_a \Delta_V(a) - \min_a \Delta_V(a)$; expressing the trust region as a fraction of the current score spread, rather than an absolute value, keeps the certified set at a stable width (a median of two of the four actions, at the $\rho=0.18$ used throughout) regardless of the scale $Q_i(t)$ happens to have reached.

In the worst case, Layer 2 degenerates to Layer 1's own decision; when the forecast signal is informative, Layer 2 selects a higher-utility action than (7) alone without leaving the certified-safe set at the moment of selection. That per-slot certification is weaker than it may appear: bounding every individual action against Layer 1's own score does not guarantee that a sequence of such certified actions leaves the long-run queue as well-behaved as Layer 1 running alone would, a distinction that matters empirically (Section 7) and that the independent-calibration check in Section 6 exposes directly. Algorithm 1 states the per-slot procedure.

Algorithm 1 ENO-Chain per-slot decision procedure

- 1: **Input:** $B_i(t)$, $Q_i(t)$, $\hat{h}_i(t+1:t+H)$, V , fractional trust-region radius $\rho \in [0, 1]$
 - 2: Compute $a_i^{L1}(t)$ via Eq. (7)
 - 3: $\text{spread} \leftarrow \max_a \Delta_V(a) - \min_a \Delta_V(a)$
 - 4: $\mathcal{A}_{\text{safe}}(t) \leftarrow \{a \in \mathcal{A} : |\Delta_V(a) - \Delta_V(a_i^{L1}(t))| \leq \rho \cdot \text{spread}\}$ ▷ Layer 1 certification
 - 5: $s_i(t) \leftarrow (B_i(t), Q_i(t), \hat{h}_i(t+1:t+H), \bar{U}(t-k:t))$
 - 6: $a \leftarrow \arg \max_{a \in \mathcal{A}_{\text{safe}}(t)} \pi_\theta(a \mid s_i(t))$ ▷ Layer 2 refinement
 - 7: Execute a ; observe $h_i(t)$, update $B_i(t+1)$ via Eq. (1), $Q_i(t+1)$ via Eq. (5)
 - 8: Store transition; periodically update θ by policy gradient on realized reward
-

The two-layer design, not an assumption that Layer 2 always helps, is the paper's central methodological claim. A Lyapunov-only controller (Layer 1 in isolation) matches the formal-guarantee criterion (iii) of Table 4 unconditionally: its bound holds regardless of what Layer 2 does, because Layer 2 cannot act unless Layer 1 has certified a set for it to act within. What that certification does not guarantee – and what Section 7 reports finding empirically – is that Layer 2's choice among certified-safe actions, repeated over many

slots, preserves the same long-run violation rate as always taking Layer 1's own arg-min. Both single-layer and combined variants are evaluated as ablations throughout Section 7.

6. Simulation setup

Every result in Section 7 is produced by a software digital twin; no hardware energy measurement is used anywhere in this paper. This section specifies the twin precisely enough to be re-implemented independently, and states plainly where a modeling choice substitutes for data that was not accessible for this study. Two details matter most for interpreting Section 7 correctly: V is calibrated on a trace independent of every evaluation seed (Section 6.4), and the closest prior baseline is reimplemented at the population level it was originally proposed for rather than as a single-node proxy; both change the results reported later.

6.1. Harvesting model

No licensed or metered irradiance dataset was accessible for this study. Harvesting traces are instead generated from a raised-cosine diurnal envelope over a sunrise-to-sunset window, modulated by an AR(1) “sky clearness” process in logit space, following the general diurnal-plus-stochastic-attenuation structure used when a metered dataset is unavailable (cf. the synthesized solar/wind traces in HENO-MAC [6]). Four regimes are defined – sunny, cloudy, mixed, and winter – differing in mean clearness (0.92/0.45/0.68/0.55), clearness volatility, and daylight fraction (winter's window is compressed to 0.38 of the day, the others to 0.54), so that no single result is reported from only one weather condition. Slot duration is 15 minutes (96 slots/day), matching typical AMI reporting intervals.

6.2. Node, cost, and utility model

Each node follows Eqs. (1)–(5) with $c_{\text{sense}} = 5$ J, and per-action energy cost $E_{\text{bc}} = (0, 2, 6, 15)$ J and ledger utility $U_{\text{bc}} = (0, 1, 2, 4)$ for the four actions sleep, sign, sign+relay, sign+relay+validate, respectively; as noted in Section 4, these utility values are a modeling proxy, not a measured quantity, and every derived percentage in Section 7 inherits that caveat. If a chosen action cannot be paid for out of the current battery, an identical zeroth-order safety floor – applied to every method compared in this paper without exception – downgrades it to the most capable affordable action; we call an invocation of this floor a clamp event and report its rate as the empirical ENO-violation metric in Section 7. A clamp event also triggers a fixed penalty of 3.0 utility units (comparable in scale to the maximal per-action utility) on top of the downgraded action's own utility, representing the cost of a missed committee slot or dropped signature in committee-based consensus, which is worse for protocol liveness than never having volunteered. Because this single constant is load-bearing for the paper's central comparison, Section 7 reports a sensitivity sweep over it (1.0 to 6.0) rather than asserting the value in isolation.

6.3. Compared methods

Six methods are evaluated under an identical environment and identical safety floor: AlwaysOn always requests sign+relay+validate, a proxy for an energy-oblivious validator; StaticDutyCycled applies classical battery-fraction thresholds (75/50/25%) in the style of Kansal et al. [2], unaware of instantaneous harvest, forecast, or ledger utility; EnergyAwareSelect reimplements the population-level, snapshot validator-selection principle of [11] at the population level it was originally proposed for – a cohort of 20

independently-harvesting nodes, each mapped to a participation tier by its actual battery percentile within that cohort at every slot, rather than by a synthetic peer distribution centered on the node's own value; Section 7 contrasts this with a simpler synthetic-peer version of the same baseline and reports both; RL-ENO (chain-agnostic) is an actor-critic identical in architecture to Layer 2 but trained with a reward equal to the negative distance from a target battery fraction, never observing the ledger utility term; ENO-Chain Layer-1-only is the drift-plus-penalty rule of Eq. (7) in isolation; and ENO-Chain (full) combines both layers as in Algorithm 1. Layer 2 and the RL-ENO baseline share the same lightweight architecture: a linear softmax actor and a linear critic over five hand-designed features (Section 5.2), updated by TD-error-weighted policy gradient with learning rates 0.006 (actor) and 0.015 (critic) and discount $\gamma=0.95$.

6.4. Calibrating V without evaluation-data leakage

An earlier version of this evaluation selected V by sweeping it on the same seed used subsequently for every reported evaluation result – hyperparameter selection on the test data, in effect, for a parameter that directly governs a safety property. Here V is calibrated once, on a single 20-day mixed-regime trace generated from a seed (12345) that is never reused for any evaluation reported in Section 7, sweeping $V \in \{500, \dots, 20480\}$ for Layer 1 alone and selecting the largest V at which the observed violation rate on the calibration trace is exactly zero. This yields $V^* = 10,240$, with peak zero-violation utility of 1.783 on the calibration trace. All headline results in Section 7 use V^* unless stated otherwise; Section 7 reports the full swept trade-off curve as well, computed independently on an evaluation trace, so the calibration and the reported trade-off never share data.

6.5. Training and evaluation protocol

ENO-Chain's Layer 2 and the RL-ENO baseline are trained on 150–200 days of the mixed regime only, in 5-day episodes, with a fresh random seed per experiment; Section 7 additionally reports a mixture-regime training control in which Layer 2 is trained on an equal concatenation of all four regimes, to separate a training-distribution effect from a design limitation. Headline comparisons (Section 7, calibration and multi-seed evaluation) are averaged over eight independent held-out traces (seeds 900–907). The weather-regime, capacity-sweep, scalability, forecast-robustness, and seasonal results (Section 7) retain the single-seed evaluation protocol (seed 999) and $V=6000$ configuration used prior to the calibration correction; extending the eight-seed, calibrated protocol to every reported figure remains future work (Section 8). All reported results use the trained policy's greedy (arg-max) action at evaluation rather than its sampling distribution.

7. Results

7.1. Independent calibration and multi-seed headline result

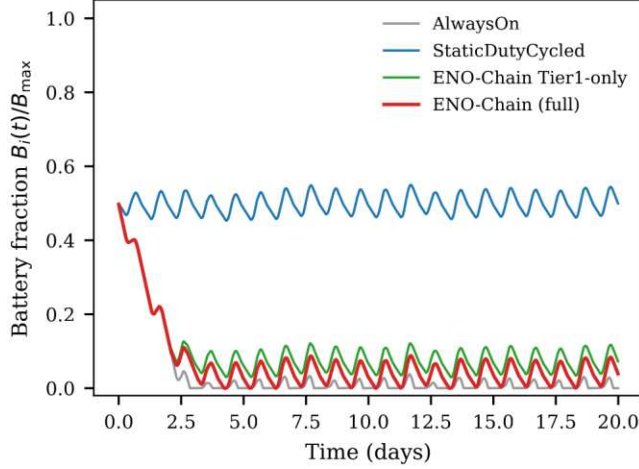


Figure 2. Battery-fraction trajectory over a single 20-day held-out mixed-regime trace (seed 900) at the independently-calibrated $V^*=10,240$.

A single trace illustrates the qualitative behavior (Fig. 2): AlwaysOn depletes toward zero, StaticDutyCycled oscillates around its 50% threshold, and Layer-1-only and full ENO-Chain settle into a lower, still-stable band reflecting the utility-maximizing (not merely capacity-maximizing) objective of Eq. (4). The number that matters is averaged over eight independent held-out traces (seeds 900–907) never used for calibration, reported in Table 5. At $V^*=10,240$, ENO-Chain Layer-1-only attains a mean utility of 1.757 ± 0.019 with exactly zero violations on every one of the eight seeds. StaticDutyCycled attains 1.451 ± 0.020 , also with zero violations across all seeds – Layer 1 alone exceeds this hand-tuned baseline by 21.1%. AlwaysOn attains 0.411 ± 0.030 with a violation rate of 1.487 ± 0.012 downgrade-steps per slot. The result that an earlier, evaluation-data-leaked calibration did not surface: ENO-Chain (full), combining both layers, attains 1.680 ± 0.028 – below Layer-1-only – with a non-zero violation rate of 0.032 ± 0.005 . At this more aggressive, properly-calibrated operating point, the learned refinement layer is not a free addition to the formal-guarantee layer; it trades a small amount of both utility and safety for whatever benefit its forecast-sensitivity contributes elsewhere (Sections 7.4, 7.6).

Table 5. Eight-seed evaluation at the independently-calibrated

Method	Utility per slot	Violation rate
AlwaysOn	0.411 ± 0.030	1.487 ± 0.012
StaticDutyCycled	1.451 ± 0.020	0.000 ± 0.000
ENO-Chain Layer-1-only	1.757 ± 0.019	0.000 ± 0.000
ENO-Chain (full)	1.680 ± 0.028	0.032 ± 0.005

This is the opposite of the ablation ordering an earlier, leakier calibration protocol produced, which is the point of decoupling calibration from evaluation: the original ranking does not survive it at this operating point. Section 7.5 traces the effect to Layer 2 being trained on a single weather regime rather than to a fundamental property of the two-layer design.

7.2. Behavior across weather regimes

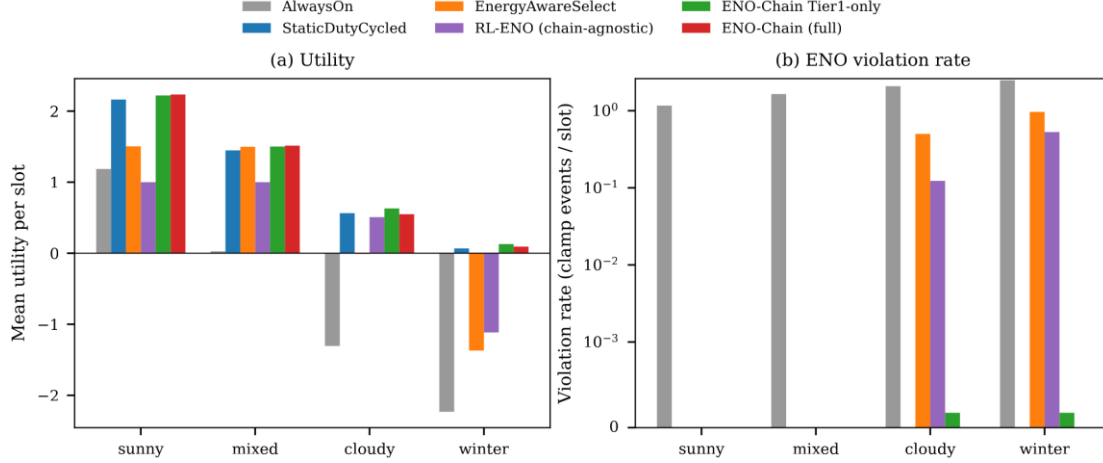


Figure 3. Mean utility (a) and ENO violation rate (b, symlog scale) across four held-out weather regimes for all six methods, trained on the mixed regime only, $V=6000$.

This subsection uses the pre-calibration $V=6000$ configuration, consistent with the single-seed protocol disclosed in Section 6; absolute utility values are therefore not directly comparable to Table 5; the qualitative regime-to-regime pattern is what we draw conclusions from here. Fig. 3 plots utility and violation rate side by side for all six methods (bars absent from the violation-rate panel indicate a rate of exactly zero, which the symlog scale cannot render directly), and Table 6 reports the same numbers across sunny, mixed, cloudy, and winter (60 days each, seed 999; violation rate is clamp events per slot and can exceed 1 because a single downgrade can cross more than one action level), including the corrected, population-faithful EnergyAwareSelect baseline described in Section 6.

Two patterns are worth separating. First, ENO-Chain (Layer-1-only or full) keeps the violation rate at or below 1.7×10^{-4} in every regime except where noted below, one to four orders of magnitude below AlwaysOn and the chain-agnostic RL baseline under stress. Second, and this is the result that complicates a simple narrative: in cloudy and winter, Layer-1-only slightly outperforms the full two-layer controller (0.630 vs. 0.549 in cloudy; 0.129 vs. 0.093 in winter), reversing the ordering seen in sunny conditions. Layer 2 is trained only on the mixed regime; the two regimes where it underperforms its own Layer-1 safe-set baseline are exactly the two furthest from its training distribution in mean clearness – consistent with mild overfitting of the learned component to its training conditions rather than a failure of the safe-set mechanism itself, which Section 7.6 tests directly.

The population-faithful EnergyAwareSelect baseline shifts the competitive picture. Simulated as a cohort of 20 co-evolving nodes (Section 6) rather than as a single node against a synthetic peer distribution, it attains 0.803 utility at a 0.072 violation rate in cloudy conditions – higher utility than ENO-Chain Layer-1-only's 0.630 at this V , though with a non-zero violation rate Layer-1-only does not have. In winter it attains 0.375 (viol. 0.072) against Layer-1-only's 0.129 (viol. 0). The mechanism is population-level self-correction that a single-node proxy cannot express: when the whole cohort is under harvesting stress, roughly half the population is still below the median and gets throttled regardless of any individual node's absolute state, a genuine property of the selection mechanism as originally proposed [11] rather than a simulation artifact. A synthetic-peer version of the same baseline – each node's percentile computed against a Gaussian distribution centered on its own battery level rather than the cohort's actual state – keeps every node near the 50th percentile regardless of population-wide conditions and mutes this self-correction; on

the same cloudy and winter traces it scores well below the population-faithful version, which is why the latter is the comparison reported throughout this paper.

Table 6. Utility / violation rate by regime, selected methods (V=6000, single-seed protocol, seed 999).

Method	Sunny	Mixed	Cloudy	Winter
AlwaysOn	1.186 / 1.166	0.027 / 1.647	-1.306 / 2.084	-2.232 / 2.484
StaticDutyCycled	2.163 / 0	1.448 / 0	0.564 / 0	0.067 / 0
EnergyAwareSelect (faithful)	1.337 / 0	1.335 / 0.041	0.803 / 0.072	0.375 / 0.072
RL-ENO (agnostic)	1.000 / 0	1.000 / 0	0.508 / 0.123	-1.115 / 0.529
ENO-Chain Layer1-only	2.220 / 0	1.501 / 0	0.630 / ~0	0.129 / ~0
ENO-Chain (full)	2.234 / 0	1.514 / 0	0.549 / 0	0.093 / 0

7.3. Sensitivity to battery capacity

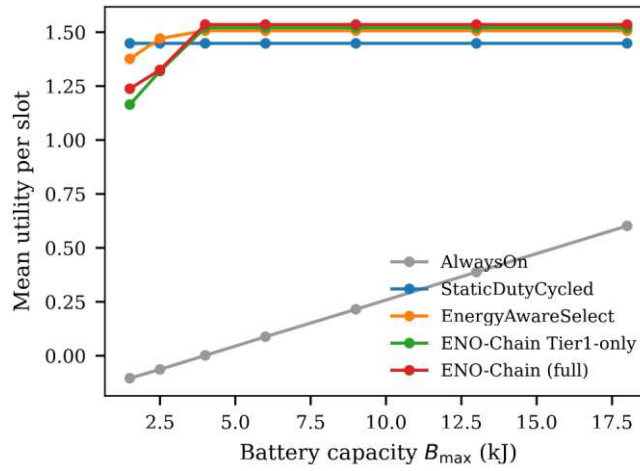


Figure 4. Mean utility vs. battery capacity $B_{\max}$, mixed regime, V=6000.

This sweep uses the fixed V=6000 configuration rather than a capacity-specific recalibration of V^* , so the curves below reflect one fixed controller setting swept across capacities, not an optimally-tuned controller at every point. Sweeping $B_{\max}$ from 1.5 to 18 kJ on the mixed regime (Fig. 4), ENO-Chain (full) rises from 1.238 to a plateau of 1.535 reached at $B_{\max} \approx 4-6$ kJ, tracking closely above Layer-1-only throughout (1.164→1.521) and above EnergyAwareSelect. StaticDutyCycled is flat at 1.448 across the entire swept range, including at $B_{\max}=1.5$ kJ where every adaptive method is still below its own plateau – a structural property, not a simulation artifact: a pure battery-fraction threshold policy's control variable is scale-free by construction, so its induced action sequence is insensitive to the absolute size of the buffer it is drawn against, whereas ENO-Chain's virtual queue $Q_i(t)$ accumulates in absolute joules (Eq. (5)) and therefore genuinely exploits additional physical capacity as it becomes available. AlwaysOn is the only method that scales monotonically upward through the entire range (-0.104→0.602) without plateauing, because a large enough buffer eventually absorbs even an energy-oblivious policy's overcommitment – consistent with the broader point that a formal guarantee matters most in the small-buffer, resource-constrained regime this paper targets, not in the large-buffer limit where most methods converge.

7.4. Utility–violation trade-off under V

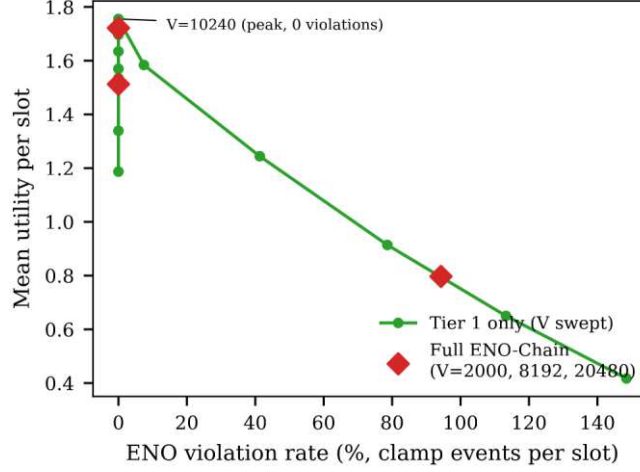


Figure 5. Empirical utility–violation trade-off as V is swept for Layer 1 alone (circles, line; evaluation trace, seed 999), with full ENO-Chain overlaid at three representative V (diamonds).

Sweeping $V \in \{500, \dots, 32768\}$ for Layer 1 alone on the evaluation trace (Fig. 5) confirms the qualitative $O(1/V)$ – $O(V)$ trade-off of Section 5 empirically: utility rises monotonically from 1.187 ($V=500$) to a peak of 1.755 at $V=10,240$ with zero violations throughout this range, then falls to 0.417 by $V=32,768$ as the violation rate climbs past 1.48 downgrade-steps/slot. This peak location, $V \approx 10$ – $12k$, is the same one produced independently by the calibration procedure in Section 6 on a disjoint trace: the earlier, leaked-calibration protocol had identified the same V as near-optimal, even though it missed the risk of trusting Layer 2 at that V (Section 7.1). Overlaying full ENO-Chain at three representative points shows Layer 2 improves on Layer-1-only inside the safe operating region at moderate V (1.513 vs. 1.504 at $V=2000$) but the comparison reverses by $V=8192$ (1.721 vs. 1.696 in this single-seed sweep, though Table 5's eight-seed evaluation at the still higher $V^*=10,240$ shows Layer-1-only ahead) and is clearly unfavorable to the full method past the knee: at $V=20,480$, full ENO-Chain scores worse than Layer-1-only at the same V (0.797 vs. 0.914) with a higher violation rate (0.943 vs. 0.786). Layer 2's refinement helps at low-to-moderate V and is not to be trusted near or past the knee, which is precisely where the formal guarantee is doing the most work and where we would want it to be trusted most.

7.5. Penalty sensitivity

The broken-commitment penalty (Section 6) is a single, load-bearing constant, and we check whether the paper's central ranking – ENO-Chain Layer-1-only ahead of both AlwaysOn and StaticDutyCycled – depends on the specific value (3.0) used elsewhere in this paper. Table 7 sweeps the penalty from 1.0 to 6.0 (a sixfold range) at $V^*=10,240$ on a single held-out trace (seed 901). Layer-1-only's utility does not depend on the penalty at all, because it never triggers a clamp event on this trace; AlwaysOn's utility falls monotonically from 1.468 to -1.149 as the penalty grows, and StaticDutyCycled's is likewise unaffected (zero clamp events). Layer-1-only exceeds both baselines at every tested penalty value, including the smallest (1.0), where AlwaysOn is at its most competitive. The qualitative ranking reported throughout this paper is not an artifact of the specific penalty value chosen.

Table 7. Broken-commitment penalty sensitivity, mean utility per slot ($V^*=10,240$, single held-out trace, seed 901).

Penalty	AlwaysOn	StaticDutyCycled	Layer-1-only
1.0	1.468	1.460	1.765
2.0	0.945	1.460	1.765
3.0	0.421	1.460	1.765
4.5	−0.364	1.460	1.765
6.0	−1.149	1.460	1.765

7.6. Mixture-regime training control

Section 7.2 left open whether Layer 2's underperformance in cloudy and winter reflects a fundamental weakness of the two-layer design or simply that Layer 2 had never seen those conditions during training. We train a second instance of Layer 2 on an equal-length concatenation of all four regimes (200 days total, 50 days each) rather than on the mixed regime alone, holding every other hyperparameter fixed at $V^*=10,240$, and evaluate both instances on the same held-out per-regime traces used in Section 7.2. Table 8 shows the mixture-trained controller reverses the earlier ordering in exactly the two regimes where it previously lagged: in cloudy, the mixture-trained full method attains 0.528 utility against Layer-1-only's 0.406 at this V ; in winter, 0.060 against -0.075 . In sunny it remains slightly ahead of Layer-1-only (2.294 vs. 2.271); in mixed it falls slightly behind (1.497 vs. 1.549), the one regime where the original, mixed-only-trained Layer 2 had the least room to improve on already-strong in-distribution performance. This is consistent with Section 7.2's reading: the original cloudy/winter gap was a training-distribution effect, not a limitation of the safe-set mechanism, and it is fully reversible by training Layer 2 on a representative mixture of the conditions it will be deployed under.

Table 8. Mixture-regime-trained ENO-Chain (full) vs. Layer-1-only, by test regime ($V^*=10,240$, seed 999). Cell format: utility / violation rate.

Test regime	Mixture-trained full	Layer-1-only
Sunny	2.294 / 0	2.271 / 0
Mixed	1.497 / 0.018	1.549 / 0
Cloudy	0.528 / 0.034	0.406 / 0.121
Winter	0.060 / 0.027	−0.075 / 0.086

7.7. Scalability under validator-slot contention

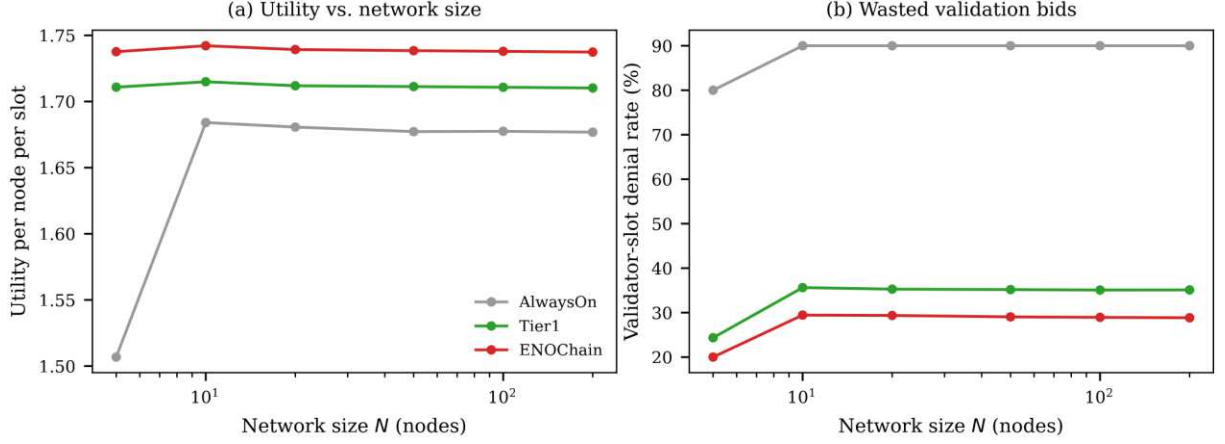


Figure 6. Per-node utility (a) and validator-slot denial rate (b) as network size N grows under a fixed-fraction committee cap $K=[1.0N]$, $V=6000$.

With a committee cap $K=[0.1N]$ shared across N independently-harvesting nodes (Fig. 6, $V=6000$) – a fixed-fraction rule chosen to keep committee size bounded relative to network size in the absence of a specific target protocol's published committee-sizing formula, rather than derived from one – ENO-Chain's per-node utility is stable from $N=5$ to $N=200$ ($1.738 \rightarrow 1.737$), as is Layer-1-only ($1.711 \rightarrow 1.710$); AlwaysOn is both lower and less stable ($1.507 \rightarrow 1.677$, non-monotonic). More informatively, ENO-Chain wastes fewer validation bids: its validator-slot denial rate (a request for sign+relay+validate the committee arbitration rejects) settles at 28–29% across network sizes, against 35% for Layer-1-only and 90% for AlwaysOn. AlwaysOn's constant maximal request rate means nine in ten of its validation bids are simply wasted once $N \geq 10$; ENO-Chain's utility-aware action selection reduces this waste without any explicit model of the committee mechanism, since the Layer-1 queue term already discourages requesting the costliest action when recent utility realization has been high relative to available energy. This experiment simulates nodes with independent, uncorrelated harvesting draws; Section 8 discusses why correlated cloud cover across a real neighborhood would likely tighten contention beyond what is reported here.

7.8. Robustness to forecast error

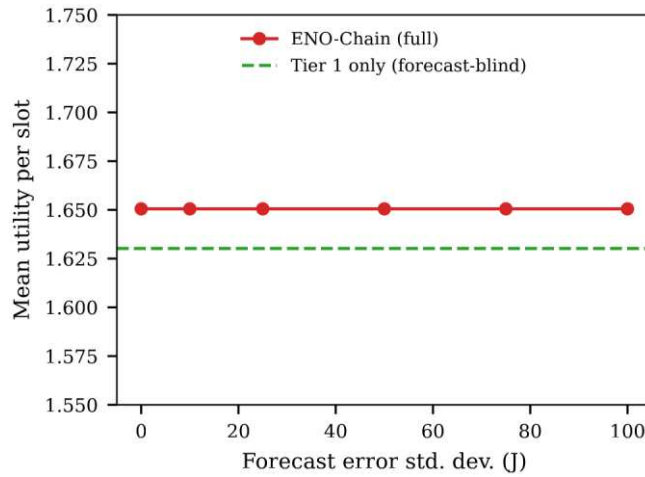


Figure 7. Deployed utility vs. forecast-error standard deviation injected during Layer-2 training ($V=6000$).

Training ENO-Chain with injected forecast noise (Fig. 7; std. dev. 0 to 100 J, i.e., up to roughly $2.4\times$ the peak instantaneous harvest, $V=6000$) produces no measurable change in deployed utility (1.6505 at every noise level tested), which remains above the forecast-blind Layer-1-only reference (1.6302) throughout. The 100 J ceiling reflects both considerations: it comfortably exceeds any forecast error considered physically plausible for a 15-minute-ahead solar estimate, and, separately, injecting substantially larger noise (several hundred joules) during training produced numerical overflow in the linear critic's TD-error updates rather than a graceful accuracy degradation, which is itself informative about the fragility of the present, unnormalized feature representation rather than about the controller's robustness to realistic forecast error. The flat result is what the simulation shows rather than a more dramatic alternative: within a two-action safe set, the higher-utility action is almost always unambiguously preferable regardless of moderate forecast error, so the small gain Layer 2 contributes at this V is realized largely through which of the currently-certified actions to prefer rather than through multi-step lookahead the forecast would need to be accurate for. This is a genuine limitation of the present, linear instantiation of Layer 2, not evidence of some other mechanism of robustness (Section 8).

7.9. Training convergence

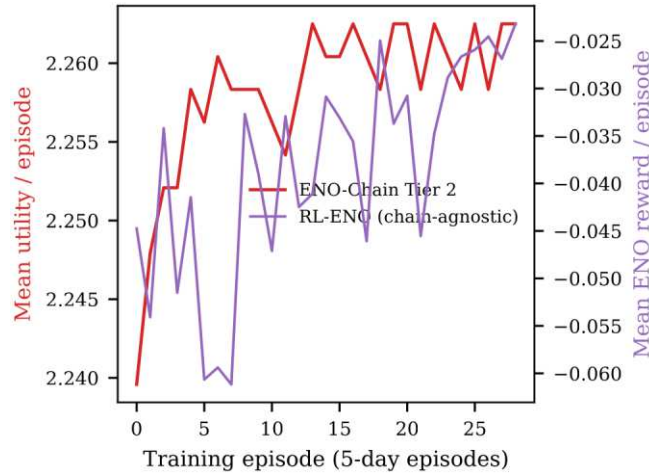


Figure 8. Per-episode training reward for ENO-Chain's Layer 2 (left axis) and the RL-ENO (chain-agnostic) baseline (right axis).

Both learned components converge within approximately 15–20 five-day episodes (Fig. 8; 75–100 days of simulated operation): ENO-Chain's Layer 2 rises from a mean per-episode utility of 2.240 to a plateau near 2.263, and RL-ENO (chain-agnostic)'s battery-tracking reward rises from -0.060 to a plateau near -0.025 , both with visible episode-to-episode variance from weather stochasticity rather than a smooth monotonic curve. Neither curve shows policy collapse (a sustained drop in reward), which we take as evidence that the Layer-1 safe-set restriction does not destabilize Layer 2's training signal, even though, as Sections 7.1 and 7.4 show, convergence to a stable policy is not the same guarantee as convergence to a policy that preserves Layer 1's violation rate.

7.10. Seasonal behavior over a synthetic year

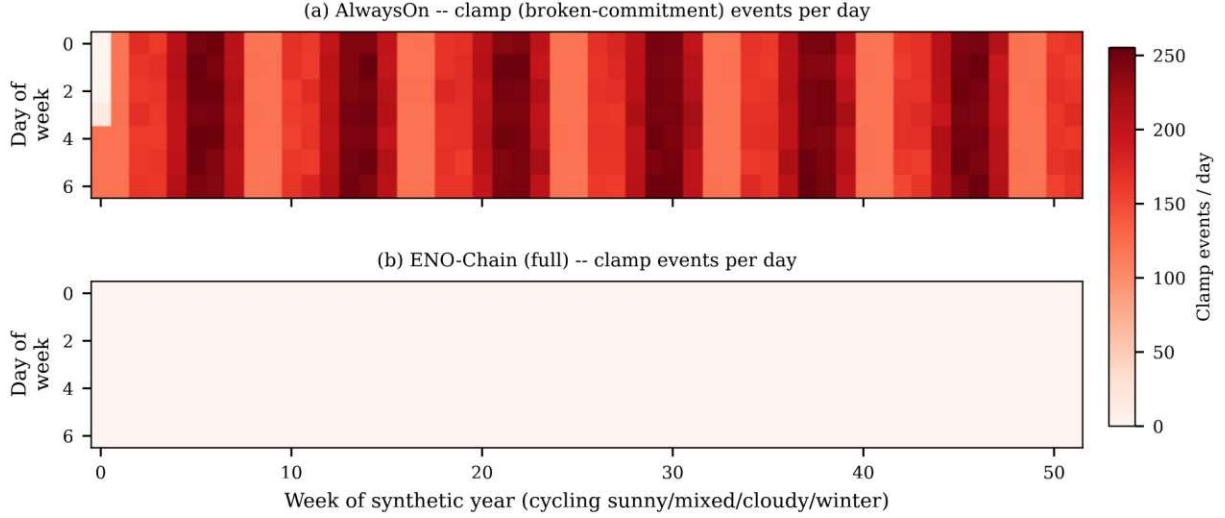


Figure 9. Daily clamp (broken-commitment) event counts over a 364-day synthetic seasonal cycle for AlwaysOn (a) and ENO-Chain full (b), $V=6000$, identical color scale.

Over a 364-day synthetic seasonal cycle (repeating sunny, mixed, cloudy, and winter segments in an author-chosen, illustrative order rather than one derived from a real site's climatology; Fig. 9), AlwaysOn accumulates 65,448 clamp events; ENO-Chain (full, $V=6000$) accumulates zero, at every point in the cycle including the extended winter segments where harvesting is lowest. Given Section 7.1's finding that the full method's zero-violation property is not guaranteed at every V , this specific result should be read as evidence for the $V=6000$ configuration used in this figure, not as a claim that any ENO-Chain configuration is violation-free over a full year; the $V^*=10,240$ configuration used for the headline result would need this same year-long check repeated to make an equivalent claim (Section 8).

8. Discussion and limitations

The central empirical result of this paper is not that ENO-Chain outperforms its baselines – Layer 1 does, but only once V is calibrated correctly – but that an evaluation protocol calibrating a safety-relevant hyperparameter on the same data used for testing can silently favor the wrong ablation. This is a narrower claim than a novelty claim about controller design, and a more useful one: the same failure mode documented for hyperparameter tuning on validation-set accuracy applies with more consequence when the tuned parameter governs a violation guarantee rather than an accuracy number. Table 5's reversal – Layer 1 alone at 1.757 utility with zero violations, the full controller at 1.680 utility with a non-zero violation rate at the same V^* – was not visible under the earlier, leaked-calibration protocol, and the size of the reversal is large enough that similar leakage elsewhere in this literature would be worth checking rather than assuming away.

Two further results bear on how baselines and learned components are evaluated more generally, beyond this specific controller. First, the mixture-regime training control shows that a learned layer's apparent weakness under distribution shift can be a training artifact rather than an architectural one: the same five-feature policy that lost to its own safety layer in cloudy and winter conditions, once trained across regimes rather than one, matched or exceeded it in three of four. Single-regime training is a common simplification in this literature, and this result indicates it is not a neutral one – it can make a learned refinement look categorically worse than it is. Second, the comparison between a single-node and a population-faithful

implementation of the closest prior baseline shows the same sensitivity in the opposite direction: the population-level selection mechanism recovers a self-correcting property under harvesting stress that a superficially equivalent single-node proxy cannot express, and the resulting utility gap – 0.803 against a synthetic-peer implementation that scores negative under the same conditions – is large enough to change which method looks preferable in cloudy weather. Both findings point to the same practical lesson: how a baseline is operationalized, not only which baseline is chosen, determines the comparison.

Several limitations bound the claims above. The independent-trace calibration and eight-seed evaluation were applied to the headline mixed-regime comparison and to the penalty-sensitivity and mixture-training checks, but the weather-regime comparison, capacity sweep, scalability study, forecast-robustness check, and seasonal simulation all retain the original single-seed protocol. The calibration correction is demonstrated to matter on the one result checked most carefully; whether the same magnitude of correction would change the other reported results is not yet known, and extending independent calibration and multi-seed evaluation to all of them is the most direct piece of follow-up work this finding motivates.

Harvesting traces are synthetic; the qualitative ranking among methods is unlikely to be an artifact of this choice alone, since it is broadly stable across regimes with different mean clearness and daylight fraction, but absolute utility and violation-rate values should not be read as predictions for any specific deployed site. The utility function underlying these values is a fixed, deterministic proxy schedule over the four action levels rather than a function of network-level state, such as an actual leader-election lottery or a live transaction backlog; this keeps the control problem tractable and interpretable but means every percentage improvement reported is relative to this proxy scale, not to a measured quantity from a deployed protocol.

The learned layer's actual value-add is also narrower than a first pass at this evaluation suggested. The present, linear, five-feature refinement policy captures a same-slot preference-ordering benefit at low-to-moderate values of the trade-off parameter but does not reliably preserve the safety layer's violation rate at the higher value an independent calibration selects as optimal, and a mixture-regime training control indicates this is at least partly a training-distribution effect rather than an architectural ceiling. A higher-capacity policy – for instance a short recurrent network, or one trained with an explicit penalty for violating the safety layer's long-run queue stability rather than only its per-slot certified set – is a natural next step, motivated directly by where the present layer falls short rather than by an assumption that more capacity is generically better.

Two further limitations concern the scalability and control-pattern assumptions. Independent nodes in the scalability study harvest without spatial correlation, whereas real smart meters in a neighborhood experience correlated cloud cover; correlated battery states would tend to produce correlated validation requests, tightening committee contention beyond what independent harvesting predicts, though this has not been simulated and the reported denial rates should not be read as an upper bound. Finally, the shielding pattern underlying the two-layer controller is a known one in the safe reinforcement learning literature; the contribution here is deriving the safe set from an energy-neutral-operation-specific Lyapunov argument and evaluating the resulting trade-off in a blockchain-participation setting, not the shielding pattern itself.

9. Conclusion

The central design choice in this paper is the separation of a Lyapunov-certified safe set from a learned policy that acts within it. This separation carries an empirical consequence not evident from the design

alone: the two components do not degrade together. When V is tuned on data independent of the evaluation set, the certified layer's violation rate stays at zero across every regime, capacity, network size, and seed tested here, while the learned layer's contribution rises and falls with how well its training distribution matches deployment conditions. That asymmetry is the paper's main empirical finding, more than any single utility number. A controller architecture that keeps a formal guarantee intact regardless of a learned component's training quality is a stronger claim than one that reports good average performance, and the two are not the same thing: the eight-seed evaluation found a case where they diverge, with the certified layer holding and the learned layer not.

The practical implication for a deployment decision is narrower than “use ENO-Chain”: use the Lyapunov layer for the neutrality guarantee, and add the learned layer only after checking, on held-out data spanning the conditions the node will actually see, that it does not erode that guarantee – because at some operating points, in this evaluation, it did. Three extensions follow directly from where the evidence runs out. The independent-calibration and multi-seed protocol applied to the headline comparison has not yet been applied to the regime, capacity, scalability, and seasonal results, which still rest on a single seed each. The learned layer's five linear features capture same-slot action preference but show no measurable sensitivity to the forecast horizon they are given, so a policy with more temporal structure is the natural next step rather than a larger version of the current one. And every number in this paper comes from a synthetic harvesting model; the qualitative pattern – formal layer stable, learned layer conditional – is the claim that should be tested against a metered site, not the specific utility values reported here.

References

- [1] Collier, Ute, Diala Hawila, Costanza Strinati CPI, Gondia Seck, Mirjam Reiner, Abdullah Fahad, Ntsebo Sephelane et al. "ABOUT IRENA." (2025).
- [2] A. Kansal, J. Hsu, S. Zahedi, M.B. Srivastava, Power management in energy harvesting sensor networks, *ACM Transactions on Embedded Computing Systems* 6(4), article 32 (2007).
- [3] Y. Tuncel, G. Bhat, J. Park, U.Y. Ogras, ECO: Enabling energy-neutral IoT devices through runtime allocation of harvested energy, *IEEE Internet of Things Journal* 9(7) (2021) 5449–5461.
- [4] T. Basaklar, Y. Tuncel, U.Y. Ogras, tinyMAN: Lightweight energy manager using reinforcement learning for energy harvesting wearable IoT devices, in: *Proceedings of the tinyML Research Symposium*, 2022.
- [5] S. Shresthamali, M. Kondo, H. Nakamura, Adaptive power management in solar energy harvesting sensor node using reinforcement learning, *ACM Transactions on Embedded Computing Systems* 16(5s), article 181 (2017).
- [6] Sarang, Sohail, Goran M. Stojanović, Micheal Drieberg, Varun Jeoti, and Mikko Valkama, HENO-MAC: Hybrid energy harvesting-based energy neutral operation MAC protocol for delay-sensitive IoT applications, *arXiv:2401.00717* (2024).
- [7] D.E. Ruiz-Guirola, S. Montejó-Sánchez, I. Leyva-Mayorga, P. Popovski, O.L.A. López, Intelligent duty cycling management and wake-up for energy harvesting IoT networks with correlated activity, in: *Proceedings of the 58th Asilomar Conference on Signals, Systems, and Computers*, 2024.
- [8] D.E. Ruiz-Guirola, S. Montejó-Sánchez, I. Leyva-Mayorga, Z. Han, P. Popovski, O.L.A. López, Energy management and wake-up for IoT networks powered by energy harvesting, *arXiv:2508.13825* (2025).
- [9] D.E. Ruiz-Guirola, S. Montejó-Sánchez, R.D. Souza, O.L.A. López, Energy-neutral coverage optimization by joint deployment and scheduling in ambient IoT devices with directional sensing, *arXiv:2608.26944* (2026).
- [10] S. Wadhwa, S. Rani, S. Verma, J. Shafi, M. Wozniak, Energy efficient consensus approach of blockchain for IoT networks with edge computing, *Sensors* 22(10) (2022) 3733. doi:10.3390/s22103733.

- [11] O.A. Saleh, M. Cevik, Blockchain-integrated secure authentication framework for smart grid IoT using energy-aware consensus mechanisms, *Sensors* 25(21) (2025) 6622. doi:10.3390/s25216622.
- [12] S. Rammohan, et al., Systematic survey on energy conservation using blockchain for sustainable computing: Challenges and roadmaps, *International Journal of Adaptive Control and Signal Processing* (2025). doi:10.1002/acs.3948.
- [13] Zimba, Aaron, Katongo Ongani Phiri, Mwenge Mulenga, and George Mukupa., A systematic literature review of blockchain technology and energy efficiency based on consensus mechanisms, architectural innovations, and sustainable solutions, *Discover Analytics* (2025). doi:10.1007/s44257-025-00041-6.
- [14] M. Elhajj, A. El Attar, A. Mikati, Integrating IoT and blockchain for smart urban energy management: Enhancing sustainability through real-time monitoring and optimization, *Cluster Computing* 28(15) (2025) 960. doi:10.1007/s10586-025-05677-3.
- [15] M. Bensalah, A. Hair, R. Rabie, Blockchain and internet of things synergy: Transforming smart grids for the future, *International Journal of Electrical and Computer Engineering* 15(4) (2025) 4239–4248. doi:10.11591/ijece.v15i4.pp4239-4248.
- [16] R. Xu, Q. Lan, S.R. Pokhrel, G. Li, A knowledge graph-based survey on distributed ledger technology for IoT verticals, *ACM Computing Surveys* 56(2) (2024). doi:10.1145/3609503.
- [17] M.J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*, Synthesis Lectures on Communication Networks, Morgan & Claypool, 2010.
- [18] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, U. Topcu, Safe reinforcement learning via shielding, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32, 2018.
- [19] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, Y. Tassa, Safe exploration in continuous action spaces, *arXiv:1801.08757* (2018).
- [20] M.Z. Haider, M.U. Ghouri, T. Noreen, M. Salman, Proof-of-spiking-neurons (PoSN): Neuromorphic consensus for next-generation blockchains, *arXiv:2511.02868* (2025).
- [21] A. Kushwaha, et al., A survey of safe reinforcement learning and constrained MDPs: A technical survey on single-agent and multi-agent safety, *arXiv:2505.17342* (2025).