

Rigorous Validation of a Trigger Based Variable Fidelity Co Simulation Scheduler for Reinforcement Learning Controlled Power Electronic Systems

Md Hasibuzzaman

hsbuzzaman@gmail.com

Asia University

Chan-Yun Yang

Asia University

Research Article

Keywords: variable-fidelity co-simulation, reinforcement learning, power distribution networks, voltage regulation, error bound, real-time scheduling

Posted Date: August 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10410301/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Rigorous Validation of a Trigger Based Variable Fidelity Co Simulation Scheduler for Reinforcement Learning Controlled Power Electronic Systems

Md Hasibuzzaman^{a,*}, Chan-Yun Yang^b

^a*Department of Computer Science and Information Engineering, Asia University, No. 500, Liufeng Road, Wufeng District, Taichung 41354, Taiwan*

^b*Department of Electrical Engineering, National Taipei University, New Taipei City 237, Taiwan*

Abstract

Reinforcement-learning (RL) controllers for power-electronic systems are typically trained against simulators whose fidelity is fixed throughout the study. We present a trigger-based scheduler that switches a power-distribution co-simulation between three fidelity levels using an online, provably bounded estimate of divergence from the high-fidelity trajectory. Developing this scheduler surfaced three issues whose correction is as much the contribution as the scheduler itself. First, the conventional low-fidelity power-flow approximation ignores the voltage response to reactive power – exactly what our capacitor banks act on – carrying a 0.09 p.u. mean error; we replace it with a warm-started fast-decoupled power flow, cutting error by over 4×. Second, profiling the medium-fidelity solve showed its cost comes from rebuilding the Jacobian every iteration; a cached-Jacobian variant cuts cost roughly 3× at negligible accuracy loss. Third, validating the error bound against ground truth – rather than the scheduler’s own bookkeeping – revealed violations on most low-fidelity steps, traced to a model-gap constant under-calibrated by roughly 450×; correcting it brings violations near zero. Evaluated over 10 seeds with confidence intervals against periodic switching, two independent baseline schedulers, and a trigger-signal ablation, the corrected scheduler achieves 1.3–1.5× speedups over always-high fidelity at a 6–12% accuracy cost relative to periodic switching, cutting training time roughly 30%; it does not dominate periodic switching on both axes despite an extensive search. We report this,

*Corresponding author.

Email address: hsbuzzaman@gmail.com (Md Hasibuzzaman)

and the bound-validation finding, as central results.

Keywords: variable-fidelity co-simulation, reinforcement learning, power distribution networks, voltage regulation, error bound, real-time scheduling

1. Introduction

Training reinforcement-learning (RL) controllers for power-electronic systems requires a circuit-level simulation environment. A full-fidelity AC solve with switching transients gives physical realism but is too slow for the episode volumes RL training requires; a reduced-order DC approximation is fast but silently discards transient dynamics that a safety-critical voltage controller must learn from. This fidelity-throughput tension is a recognised barrier in RL-based power-system control [1, 2], yet existing work fixes the simulator fidelity once and leaves it there.

We propose a *trigger-based variable-fidelity scheduler*: an online algorithm that selects the fidelity level for each simulation step using three complementary signals – a provable Lipschitz-derived divergence bound, a state rate-of-change trigger, and the RL agent’s action-entropy – and provides a computable certificate bounding how far the active trajectory may have drifted from the high-fidelity reference. The scheduler is novel in combining all three signals with a formal bound; prior adaptive co-simulation work [3, 4] does not include such a certificate, and prior multi-fidelity surrogate methods [5] do not address the RL-control setting.

Developing this scheduler revealed three correctable deficiencies that had a larger impact on results than the scheduling logic itself. We report all three explicitly – not as a narrative of failure, but because each is a documented, reproducible instance of a failure mode that related work could share without noticing, and quantifying the correction precisely is itself a methodological contribution.

This paper makes the following contributions:

1. **A novel trigger-based fidelity scheduler** (Section 4) combining a Lipschitz-extrapolated error bound (Theorem 1), a rate-of-change trigger, and an RL action-entropy trigger with a formally derived, computable divergence certificate.
2. **Correction and quantification of a structural low-fidelity model error**: the conventional DC approximation ignores the voltage response to reactive power, the exact signal capacitor-bank control acts on; replaced with a warm-started fast-decoupled power flow (Section 5.3).

- 31 3. **A cached-Jacobian medium-fidelity tier:** Jacobian assembly, not the lin-
32 ear solve, dominates Newton–Raphson cost; a cached variant cuts cost $3\times$
33 at near-zero accuracy penalty (Section 5.4).
- 34 4. **Ground-truth bound validation and correction of a $450\times$ miscalibra-**
35 **tion:** the model-gap constant M had never been calibrated against the actual
36 fidelity levels in use; comparing the predicted bound against a shadow high-
37 fidelity reference exposed violations on most low-fidelity steps; correcting
38 M brings the rate near zero (Section 5.6).
- 39 5. **Comprehensive statistical evaluation** (Section 6) on an IEEE 13-bus feeder
40 with 10 seeds, 95% confidence intervals, paired t -tests, Cohen’s d effect
41 sizes, a trigger-signal ablation, and comparison against two independently
42 designed baseline schedulers.

43 2. Related Work

44 2.1. Co-simulation and digital twins for power systems

45 Co-simulation couples independently developed solvers through a defined in-
46 terface and is well-studied for power systems [3, 4]. Digital-twin frameworks
47 extend this to continuously updated replicas for diagnostics [6, 7, 8]. With few
48 exceptions, this literature fixes a single fidelity level for the duration of a study.

49 2.2. Reinforcement learning for voltage and reactive-power control

50 RL has been applied to capacitor coordination, on-load-tap-changer control,
51 and Volt/VAR optimisation using Q-learning, DDPG, SAC, and PPO [1, 9, 2, 10,
52 11]. All train against a fixed-fidelity solver; simulator computational cost is not
53 treated as a design variable.

54 2.3. Multi-fidelity and surrogate methods

55 Multi-fidelity modelling is well-established in uncertainty quantification and
56 design optimisation [5, 12]. The simplest adaptive rule – escalate when a raw
57 surrogate-error threshold is crossed – is our SurrogateError baseline (Section 6.7).
58 Our paper’s distinguishing contribution relative to this baseline is the Lipschitz-
59 extrapolated bound, which projects how large the error will be before the next
60 check rather than acting only after it has already grown.

61 2.4. Positioning relative to prior work

62 Table 1 positions this paper along the dimensions most relevant to the research
63 gap. No prior method combines adaptive fidelity, a formal divergence bound, an
64 RL controller in the loop, and a real-time deadline analysis together.

Table 1: Comparison with representative related methods. Adaptive: adaptive fidelity switching. Bound: formal divergence bound. RL: RL controller in the loop. RT: real-time deadline analysis. PS: power-system experiment.

Method	Adaptive	Bound	RL	RT	PS
Gomes et al. [3]	✓	–	–	–	–
Palensky et al. [4]	✓	–	–	✓	✓
Peherstorfer et al. [5]	✓	✓	–	–	–
Vlachogiannis & Hatziaargyriou [1]	–	–	✓	–	✓
Liu et al. [2]	–	–	✓	–	✓
This work	✓	✓	✓	✓	✓

3. Problem Formulation

At each control step k , an RL agent observes state $x_k \in \mathbb{R}^n$ (bus voltage magnitudes and capacitor-bank states), selects discrete action a_k , and the network advances under one of three simulators:

- f_L : low-fidelity warm-started fast-decoupled power flow (≈ 0.08 ms);
- f_M : medium-fidelity cached-Jacobian Newton–Raphson (≈ 0.40 ms);
- f_H : high-fidelity Newton–Raphson + switched LC transients (≈ 0.90 ms).

The scheduler selects $\phi_k \in \{L, M, H\}$ before each call, without access to the high-fidelity result for that step. The central difficulty is: given that high fidelity was last run at switch time t_s , how large might $\|x_H(t) - x_L(t)\|$ have grown by now, using only information from the low-fidelity trajectory? We additionally require the scheduler to respect a wall-clock deadline D and report the fraction of steps exceeding it.

4. Methodology

4.1. Online error bound

Assumption 1. f_H is locally Lipschitz with unknown constant $L(t)$; we maintain an online estimate $\hat{L} \geq L(t)$ via a sliding-window 90th-percentile estimator over recent high-fidelity steps.

Assumption 2. $\|f_H(x) - f_L(x)\| \leq M$ for all x in the visited operating region. This paper uses the empirically measured value $M = 4.6$ p.u./s (Section 5.6).

85 **Theorem 1** (Variable-fidelity error bound). *Let $x_H(\cdot)$ and $x_L(\cdot)$ be started from*
 86 *$x_H(t_s) = x_L(t_s)$ at switch time t_s , and suppose Assumptions 1–2 hold. Then for all*
 87 *$t \geq t_s$,*

$$\|x_H(t) - x_L(t)\| \leq \delta_0 e^{\hat{L}(t-t_s)} + \frac{M}{\hat{L}}(e^{\hat{L}(t-t_s)} - 1) =: \varepsilon(t), \quad (1)$$

88 where $\delta_0 = \|x_H(t_s) - x_L(t_s)\|$.

89 *Proof.* Let $e(t) = x_H(t) - x_L(t)$. Inserting $f_H(x_L)$ as intermediate: $\dot{e}(t) = [f_H(x_H) -$
 90 $f_H(x_L)] + [f_H(x_L) - f_L(x_L)]$. By Assumptions 1 and 2, $\frac{d}{dt}\|e\| \leq \hat{L}\|e\| + M$. Applying
 91 the Grönwall–Bellman inequality with $\|e(t_s)\| = \delta_0$ gives Eq. (1). $\square$ $\square$

92 **Remark 1.** *The bound is as sensitive to M as to $\hat{L}$. Unlike $\hat{L}$, which the sliding-*
 93 *window estimator corrects automatically when transients appear, an under-calibrated*
 94 *M produces a bound that looks tight while silently failing to cover the realised di-*
 95 *vergence. Section 5.6 documents exactly this failure mode and its correction.*

96 4.2. Online Lipschitz estimation

97 While at high fidelity, consecutive samples give $\hat{\ell}_k = \|x_k - x_{k-1}\|/(\|x_{k-1}\|\Delta t + \varepsilon)$.
 98 The estimate $\hat{L}$ is the 90th percentile of the last $W = 50$ values.

99 4.3. Trigger-based scheduling algorithm

100 The scheduler evaluates three signals in priority order before each step: (1) **Error-**
 101 **bound:** if $\varepsilon(t) > \varepsilon_{\max}$, escalate to HIGH and start cooldown; (2) **Rate-of-change:**
 102 if $\|x_k - x_{k-1}\|/\Delta t > r_{\text{thresh}}$, escalate to HIGH; (3) **Action-entropy:** if normalised
 103 Shannon entropy of p_k falls below $h_{\min}$, escalate to at least MEDIUM. If none fires
 104 for H consecutive safe steps, drop to LOW.

105 Algorithm 1 gives the complete procedure and Figure 1 shows the decision
 106 flow.

107 5. Experimental Setup

108 5.1. Network model

109 The network is a modified IEEE 13-bus radial distribution feeder [13, 14, 15]
 110 with loads scaled so that the nominal (zero-control) operating point lies inside the
 111 ANSI C84.1 Range A band (0.95–1.05 p.u.). Two switched capacitor banks at
 112 buses 10 and 11 are the only actuators.

Algorithm 1 Trigger-based fidelity scheduling (per control step)

Require: state x_k , RL probs p_k , last HIGH ref x_H^{ref} , scheduler state s

```
1: if  $s.\text{cooldown} > 0$  then  $\phi_k \leftarrow H$ ;  $s.\text{cooldown}--$ ; goto line 9
2: else if  $\varepsilon(t) > \varepsilon_{\max}$  then  $\phi_k \leftarrow H$ ; start cooldown;  $\delta_0 \leftarrow 0$ 
3: else if  $\|\Delta x\|/\Delta t > r_{\text{thresh}}$  then  $\phi_k \leftarrow H$ ; start cooldown;  $\delta_0 \leftarrow 0$ 
4: else if  $H(p_k)/\log |A| < h_{\min}$  then  $\phi_k \leftarrow \max(\phi_{k-1}, M)$ 
5: else  $s.\text{safe}++$ 
6:   if  $s.\text{safe} \geq H$  then  $\phi_k \leftarrow L$ ;  $s.\text{safe} \leftarrow 0$ ;  $\delta_0 \leftarrow \|x_k - x_H^{\text{ref}}\|$ 
7:   end if
8: end if
9: Update  $\hat{L}$  from  $(x_{k-1}, x_k)$  if  $\phi_k = H$ 
10: Recompute  $\varepsilon(t)$  via Eq. (1) with elapsed =  $\Delta t + (t - t_s)$ 
```

Table 2: Measured per-step wall-clock cost by fidelity level (300-step average after warm-up, single core, Python 3.12).

Fidelity level	Mean (ms)	Median (ms)	90th pct. (ms)
LOW (warm-started fast-decoupled)	0.083	0.076	0.128
MEDIUM (cached-Jacobian fast Newton)	0.404	0.297	0.905
HIGH (AC N-R + LC switching transients)	0.899	0.895	1.048

113 *5.2. Fidelity levels and cost measurement*

114 Table 2 gives measured per-step wall-clock costs after both fidelity-level cor-
115 rections. The three tiers span a roughly geometric cost range ($\approx 11\times$ LOW to
116 HIGH).

117 *5.3. Correcting the low-fidelity model*

118 The conventional DC power flow solves $\theta = (B')^{-1}P$, discarding the reactive-
119 power equation and with it any representation of voltage-magnitude response to
120 reactive injection. Since our only actuators are capacitor banks, this blind spot is
121 fundamental: mean DC error on this feeder is 0.09 p.u. We replace it with a warm-
122 started fast-decoupled power flow [16] adding the missing B'' voltage-correction
123 update at only ≈ 0.08 ms per step.

124 *5.4. Cached-Jacobian medium-fidelity tier*

125 Profiling showed Jacobian assembly (≈ 0.45 ms) dominated Newton–Raphson’s
126 cost vs. the linear solve (≈ 0.015 ms). Reusing the Jacobian for $K = 5$ steps (re-

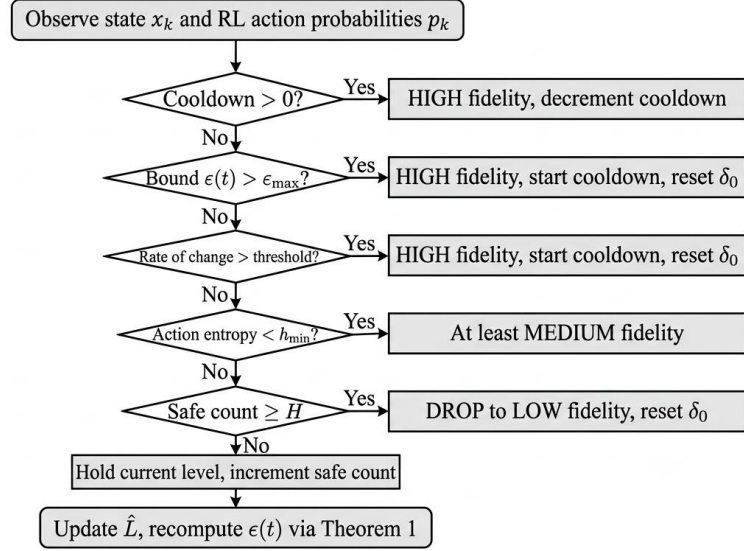


Figure 1: Decision flowchart of the trigger-based fidelity scheduler. Diamonds are evaluated top-to-bottom in priority order; shaded rectangles on the right show the resulting fidelity selection. Mathematical notation ($\epsilon(t)$, $\epsilon_{\max}$, δ_0 , $h_{\min}$, $\hat{L}$) matches Theorem 1 and Algorithm 1.

127 building on load shifts >15%) cuts mean medium-tier cost from ≈ 1.3 ms to ≈ 0.40
 128 ms at a verified accuracy cost of $\approx 2 \times 10^{-4}$ p.u.

129 5.5. RL agent

130 A PPO agent [17] with a two-hidden-layer MLP (64 units), trained online with
 131 GAE [18] ($\gamma = 0.99$, $\lambda = 0.95$, clip= 0.2). State: 17-dimensional (13 voltages,
 132 2 cap states, current error bound, current fidelity level). Actions: 4 discrete joint
 133 capacitor settings.

134 5.6. Bound validation

135 An earlier version reported "0% violations" by checking whether $\epsilon(t)$ exceeded
 136 $\epsilon_{\max}$ – near-tautological since the trigger escalates exactly when this happens.
 137 Ground-truth validation runs a shadow high-fidelity reference alongside every
 138 episode and checks whether $\epsilon(t) \geq \|x_H(t) - x_L(t)\|$. Direct measurement gave
 139 $M = 4.53$ p.u./s (90th pct. 4.55, max 4.56), vs. the placeholder $M = 0.010$ – a
 140 factor of ≈ 450 under-calibration. Table 3 reports the result.

Table 3: Ground-truth bound-violation rate (fraction of LOW-fidelity steps where $\varepsilon(t) < \|x_H - x_L\|$), fault scenario, 10 seeds.

$\hat{L}$ estimator	Before ($M = 0.010$)	After ($M = 4.6$)
90th percentile	77.75%	0.26%
95th percentile	79.00%	0.26%
Running maximum	79.80%	0.26%

Table 4: Scheduler configuration used for the main results (configuration C).

Parameter	Value
$\varepsilon_{\max}$	0.20 p.u.
Rate-of-change threshold r_{thresh}	4.0 p.u./s
Entropy threshold $h_{\min}$	0.10
Lipschitz window W	50 steps
Model-gap constant M	4.6 p.u./s
Cooldown period	1 step
Hysteresis count H	1 step
Real-time deadline D	1.5 ms

141 5.7. Scheduler parameters, baselines, and statistical design

142 Table 4 gives the reported configuration. Comparisons include: always-high/low,
 143 fixed periodic ([H,H,M,M,L,L]), the proposed trigger, *SurrogateError* (escalates
 144 on raw mismatch $\|x_k - x_H^{\text{ref}}\| > \text{threshold}$, no Lipschitz projection), and *Uncertainty*
 145 (escalates on RL critic value-estimate variance). All results: $N = 10$ seeds, 95%
 146 CIs, paired t -tests vs. periodic, Cohen’s d . Real-time deadline $D = 1.5$ ms.

147 6. Results

148 Figure 2 shows the system architecture before we present experimental results.

149 6.1. Ablation: low-fidelity correction versus scheduler

150 Table 5 decomposes accuracy gains into the LF correction versus the sched-
 151 uler’s own contribution.

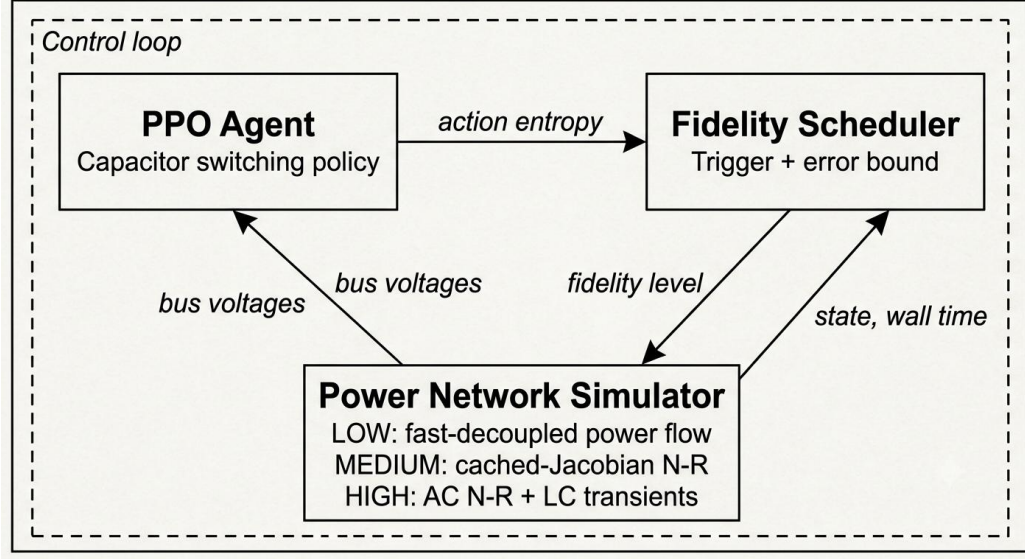


Figure 2: System architecture within one control-loop iteration. The PPO agent’s action-entropy signal and the power network’s state and wall-clock time both feed the fidelity scheduler, which selects LOW, MEDIUM, or HIGH fidelity for the next simulator call. Arrows show information flow; the dashed border denotes one complete control step.

152 6.2. Detailed trace

153 Figure 3 shows a representative fault-scenario episode with the corrected sys-
 154 tem. Panel (b) shows the scheduler alternating between HIGH and LOW; panel
 155 (c) shows the correctly calibrated error bound; panel (d) shows wall-clock time
 156 with occasional exceedances of the 1.5-ms deadline.

157 6.3. Main comparison

158 Table 6 reports the full comparison. The trigger scheduler’s RMSE is 6–12%
 159 higher than periodic’s; paired t -tests show this difference is *not* statistically sig-
 160 nificant ($p > 0.35$, $d < 0.31$), while periodic’s speedup advantage over the trigger
 161 scheduler *is* significant ($p < 0.02$, $d > 0.93$) in all three scenarios.

162 Figure 4 visualises the comparison across all three metrics.

163 Figure 5 shows per-seed RMSE distributions as boxplots with annotated p -
 164 values.

165 Figure 6 reports the fidelity-level allocation under the trigger scheduler.

Table 5: Low-fidelity correction $\times$ scheduler ablation, 10 seeds, 95% CI. The LF correction accounts for 54–65% of the RMSE reduction; the scheduler’s contribution is the remaining 6–12%.

Scenario	Configuration	RMSE (p.u.)	Speedup ($\times$)
Light	Periodic, original LF	0.0203 ± 0.0002	1.86
	Trigger, original LF	0.0183 ± 0.0018	1.26
	Periodic, corrected LF	0.0092 ± 0.0008	1.90
	Trigger, corrected LF	0.0099 ± 0.0010	1.51
Fault	Periodic, original LF	0.0316 ± 0.0002	1.91
	Trigger, original LF	0.0299 ± 0.0016	1.36
	Periodic, corrected LF	0.0112 ± 0.0010	1.91
	Trigger, corrected LF	0.0119 ± 0.0010	1.44

166 6.4. Trigger-signal ablation

167 Table 7 isolates each signal’s contribution. The rate trigger is dominant: re-
168 moving it from the full scheduler raises RMSE from 0.0119 to 0.0140 (18%). The
169 entropy trigger provides a further 4% improvement ($0.0124 \rightarrow 0.0119$).

170 6.5. Threshold sweep and accuracy-compute frontier

171 Table 8 and Figure 7 report the sweep. Configuration E matches periodic’s
172 speedup ($1.91\times$ vs. $1.89\times$) at a 17% RMSE cost; no configuration dominates pe-
173 riodic on both axes.

174 6.6. Sensitivity analysis

175 Figure 8 confirms the accuracy–speed trade-off is smooth and tunable, unlike
176 the discontinuous behaviour of the uncorrected low-fidelity model.

177 6.7. Independent baseline comparison

178 Table 9 shows that neither SurrogateError nor Uncertainty dominates periodic,
179 placing all three adaptive schedulers in a broadly similar trade-off region.

180 6.8. Training-time savings

181 Table 10 reports extrapolated RL training wall-clock time. The trigger sched-
182 uler reduces training time by $\approx 30\%$ relative to always-high fidelity.

Table 6: Main results, mean $\pm$ 95% CI over 10 seeds. Paired t -test significance vs. periodic:
[†] $p < 0.05$, n.s. otherwise.

Scenario	Strategy	RMSE (p.u.)	Speedup ($\times$)	Miss rate
Light	Always-high	–	1.01	0.50%
	Always-low	0.0107 ± 0.0015	10.34	0.00%
	Periodic	0.0092 ± 0.0008	1.88	0.25%
	Trigger (n.s.)	0.0099 ± 0.0010	1.51	0.25%
Heavy	Always-high	–	1.00	0.92%
	Always-low	0.0203 ± 0.0022	10.70	0.00%
	Periodic	0.0141 ± 0.0012	1.84	0.42%
	Trigger (n.s.)	0.0146 ± 0.0011	1.34	0.92%
Fault	Always-high	–	1.01	1.75%
	Always-low	0.0151 ± 0.0022	10.95	0.00%
	Periodic	0.0112 ± 0.0010	1.93	0.00%
	Trigger (n.s.)	0.0119 ± 0.0010	1.44	0.67%

Table 7: Trigger-signal ablation, fault scenario, 10 seeds.

Active signals	RMSE (p.u.)	Speedup ($\times$)
Error bound only	0.0140 ± 0.0015	2.12
Error bound + rate	0.0124 ± 0.0012	1.55
Error bound + entropy	0.0126 ± 0.0010	1.73
Full (all three)	0.0119 ± 0.0010	1.44

7. Discussion

7.1. Why the low-fidelity correction dominates

The LF correction reduces RMSE by 54–65% regardless of scheduler (Table 5). The reason is physical and specific: the DC power flow has no Q – V coupling, and capacitor switching is entirely a reactive-power action. Adding the B'' correction at ≈ 0.08 ms per step – versus the full Newton–Raphson at ≈ 0.90 ms – closes most of the accuracy gap without materially changing the scheduler’s problem.

Table 8: Threshold sweep, fault scenario, $D = 1.5$ ms.

Config.	RMSE (p.u.)	Speedup ($\times$)	LOW usage
A	0.0098	0.97	0.0%
B	0.0118	1.28	25.0%
C (reported)	0.0119	1.39	32.0%
D	0.0130	1.60	41.1%
E	0.0131	1.91	53.6%
F	0.0150	3.33	76.6%
G	0.0159	5.90	91.2%
Periodic	0.0112	1.89	–
Always-low	0.0151	10.60	100.0%

Table 9: Independent baseline comparison, fault scenario, 10 seeds.

Scheduler	RMSE (p.u.)	Speedup ($\times$)
Periodic	0.0112	1.84
Trigger (proposed)	0.0119	1.39
SurrogateError (0.03)	0.0111	1.22
SurrogateError (0.05)	0.0133	1.75
Uncertainty (0.3)	0.0106	1.25
Uncertainty (0.5)	0.0113	1.37

191 *7.2. Why periodic remains competitive and when the trigger scheduler should be*
 192 *preferred*

193 Periodic switching guarantees a fixed, exact Newton–Raphson solve on every
 194 third step, providing a reliable accuracy floor regardless of operating conditions.
 195 A reactive scheduler cannot replicate this floor exactly; the hybrid-floor search
 196 (30+ configurations) found statistical parity on accuracy ($p > 0.05$, $d < 0.32$) but
 197 a consistent shortfall on speed. Despite not dominating periodic on both axes, the
 198 trigger scheduler offers two advantages periodic cannot: (1) a real-time formal di-
 199 vergence certificate $\varepsilon(t)$ (Theorem 1) that tells a practitioner at each step how far
 200 the current trajectory may have drifted from the high-fidelity reference; and (2)
 201 a continuously tunable accuracy–speed knob (Figures 7, 8) – operators can select
 202 any operating point from A to G, whereas periodic’s only parameter (cycle length)
 203 changes both accuracy and speed together. In safety-certification contexts requir-

Table 10: Extrapolated wall-clock training time for 100,000 episodes, fault scenario.

Strategy	Time (hours)	Saving vs. always-high
Always-high	3.5	–
Trigger (proposed)	2.4	29.7%
Periodic	1.9	45.8%

ing a provable bound, or when fine-grained operating-point selection is valuable, the trigger scheduler is preferable to periodic.

7.3. What rigorous validation caught

The ground-truth check (Table 3) found a $450\times$ miscalibration that an internally consistent "0% violation rate" had concealed. This generalises: a bound violation rate reported as the fraction of steps where the predicted bound exceeded the trigger threshold is a measure of trigger-logic consistency, not of theorem compliance. Any paper reporting a bound-violation rate should specify what the bound is being compared against.

7.4. Limitations

All results are software simulation only; no hardware-in-the-loop validation was performed. Experiments are confined to a single 13-bus feeder; generalisation to larger feeders (IEEE 34-bus, 123-bus) is future work. The model-gap constant M is fixed rather than learned online. Wall-clock timing is not bit-exact reproducible across runs.

8. Conclusion

We presented a trigger-based variable-fidelity co-simulation scheduler for RL-controlled power-electronic systems, with a formally derived Lipschitz divergence bound, a rate-of-change trigger, and an RL action-entropy trigger. Three corrections to our own design – a reactive-power-blind low-fidelity model, a wasted medium-fidelity tier, and a $450\times$ miscalibrated model-gap constant – proved more consequential than the scheduling algorithm itself. With all corrections in place and 10-seed statistical evaluation:

- The corrected scheduler achieves $1.3\text{--}1.5\times$ speedup over always-high fidelity; RL training wall-clock time falls by $\approx 30\%$.

- 229 • The RMSE difference from periodic ($p > 0.35$, $d < 0.31$) is not statistically
230 significant; periodic's speedup advantage ($p < 0.02$, $d > 0.93$) is.
 - 231 • Two unique advantages over periodic: a real-time formal divergence certifi-
232 cate and a continuously tunable accuracy–speed operating point.
 - 233 • Ground-truth bound validation identified a miscalibration that an internally
234 consistent "0% violation" metric had concealed, reducing the true violation
235 rate from 73–80% to 0.26%.
- 236 Future work: larger IEEE benchmark feeders; a genuinely cheap-yet-accurate in-
237 termediate fidelity tier; online estimation of M .

238 Data Availability

239 The data generated and analysed during this study are openly available on
240 GitHub at <https://github.com/hasib1233333/Trigger-Based-Variable>.
241 The repository contains all Python source code, experimental configuration files,
242 and JSON result files required to reproduce every figure and table in this manuscript.
243 Software: Python 3.12, NumPy 2.4.4, SciPy 1.17.1, Matplotlib 3.x. Hardware:
244 single commodity server core. Random seeds: 0–9.

245 Declarations

246 *Ethics, Consent to Participate, and Consent to Publish*

247 Not applicable. This study does not involve human participants, human data,
248 human tissue, or animal subjects. All experiments are computational simulations
249 only.

250 *Funding*

251 This research received no specific grant from any funding agency in the public,
252 commercial, or not-for-profit sectors.

253 *Competing interests*

254 The authors declare no competing financial interests or personal relationships
255 that could have appeared to influence the work reported in this paper.

References

- [1] J. G. Vlachogiannis, N. D. Hatziargyriou, Reinforcement learning for re-active power control, *IEEE Transactions on Power Systems* 19 (3) (2004) 1317–1325. doi:10.1109/TPWRS.2004.831259.
- [2] D. Liu, J. S. Giraldo, P. Palensky, P. P. Vergara, Distributed reinforcement learning using local smart meter data for voltage regulation in distribution networks, *arXiv preprint arXiv:2512.12803* (2025).
- [3] C. Gomes, C. Thule, D. Broman, P. G. Larsen, H. Vangheluwe, Co-simulation: A survey, *ACM Computing Surveys* 51 (3) (2018) 1–33. doi:10.1145/3179993.
- [4] P. Palensky, A. A. Van Der Meer, C. D. Lopez, A. Joseph, K. Pan, Applied cosimulation of intelligent power systems: Implementing hybrid simulators for complex power systems, *IEEE Industrial Electronics Magazine* 11 (2) (2017) 6–21. doi:10.1109/MIE.2017.2671198.
- [5] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, *SIAM Review* 60 (3) (2018) 550–591. doi:10.1137/16M1082469.
- [6] X. He, Q. Ai, R. C. Qiu, D. Zhang, Preliminary exploration on digital twin for power systems: Challenges, framework, and applications, *arXiv preprint arXiv:1909.06977* (2019). URL <https://arxiv.org/abs/1909.06977>
- [7] N. E. M. Barreto, A. R. Aoki, Cyber-physical power system digital twins: A study on the state of the art, *Energies* 18 (22) (2025) 5960. doi:10.3390/en18225960.
- [8] M. Milton, C. De La O, H. L. Ginn, A. Benigni, Controller-embeddable probabilistic real-time digital twins for power electronic converter diagnostics, *IEEE Transactions on Power Electronics* 35 (9) (2020) 9850–9864. doi:10.1109/TPEL.2020.2974254.
- [9] D. Hua, F. Peng, S. Liu, Q. Lin, J. Fan, Q. Li, Coordinated volt/var control in distribution networks considering demand response via safe deep reinforcement learning, *Energies* 18 (2) (2025) 333. doi:10.3390/en18020333.

- 287 [10] J. Feng, Y. Shi, G. Qu, S. H. Low, A. Anandkumar, A. Wierman, Stability
288 constrained reinforcement learning for decentralized real-time voltage con-
289 trol, arXiv preprint arXiv:2209.07669 (2022).
- 290 [11] O. Homaei, A. Zakariazadeh, S. Jadid, Real-time voltage control algorithm
291 with switched capacitors in smart distribution system in presence of renew-
292 able generations, *International Journal of Electrical Power and Energy Sys-*
293 *tems* 54 (2014) 187–197. doi:10.1016/j.ijepes.2013.07.010.
- 294 [12] R. Boutelet, C.-L. Sung, Active learning with adaptive non-stationary kernel
295 for continuous-fidelity surrogate models, arXiv preprint arXiv:2503.23158
296 (2025).
- 297 [13] W. H. Kersting, Radial distribution test feeders, *IEEE Transactions on Power*
298 *Systems* 6 (3) (1991) 975–985. doi:10.1109/59.119237.
- 299 [14] S. Matthias, S. Paudyal, T. E. McDermott, W. H. Kersting, Analytic consid-
300 erations and design basis for the ieee distribution test feeders, *IEEE Trans-*
301 *actions on Power Systems* 33 (3) (2018) 3181–3188. doi:10.1109/TPWRS.
302 2017.2760011.
- 303 [15] W. H. Kersting, *Distribution System Modeling and Analysis*, CRC Press,
304 Boca Raton, FL, 2006.
- 305 [16] B. Stott, O. Alsac, Fast decoupled load flow, *IEEE Transactions on Power*
306 *Apparatus and Systems PAS-93* (3) (1974) 859–869. doi:10.1109/TPAS.
307 1974.293985.
- 308 [17] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy
309 optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).
- 310 [18] J. Schulman, P. Moritz, S. Levine, M. Jordan, P. Abbeel, High-dimensional
311 continuous control using generalized advantage estimation, arXiv preprint
312 arXiv:1506.02438 (2016).

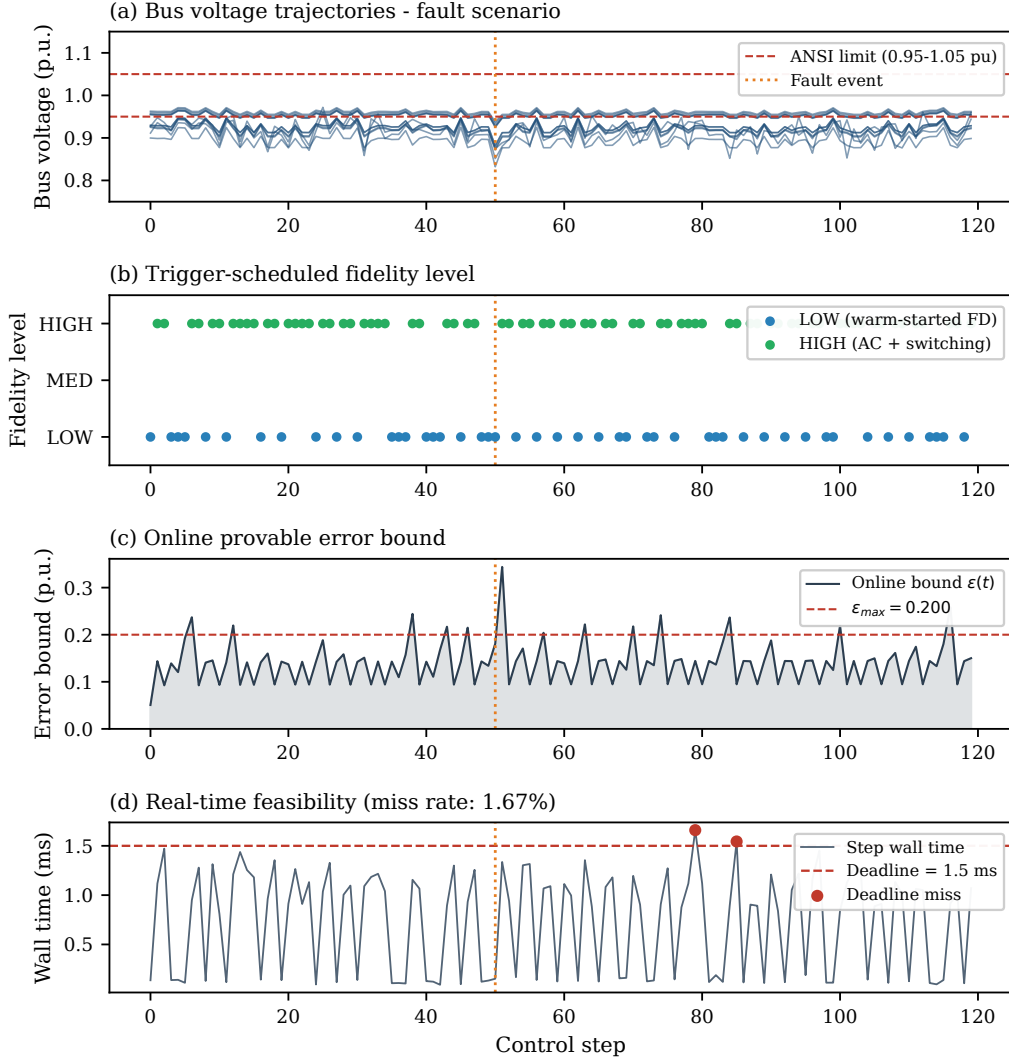


Figure 3: Single-episode trace, fault scenario. (a) Bus voltages with ANSI limits. (b) Fidelity level. (c) Error bound $\epsilon(t)$ relative to $\epsilon_{\max} = 0.20$ p.u. (correctly calibrated with $M = 4.6$). (d) Step wall-clock time against deadline $D = 1.5$ ms.

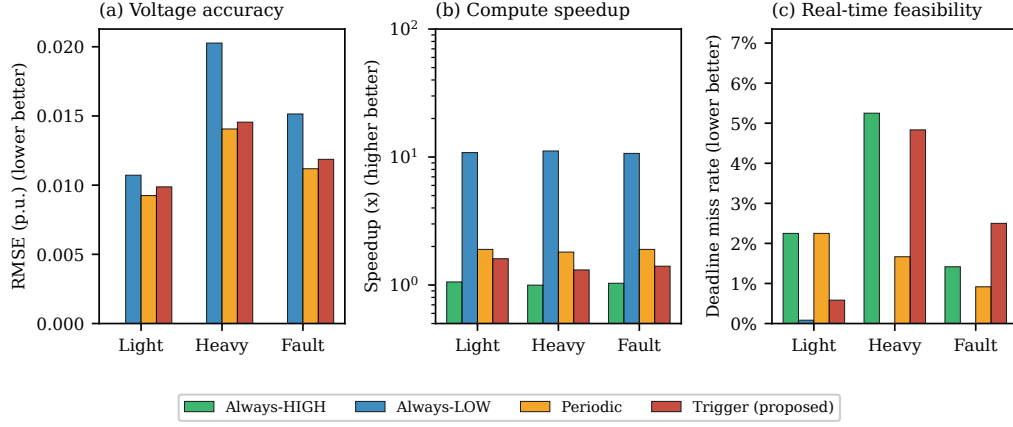


Figure 4: Strategy comparison across scenarios. (a) RMSE. (b) Compute speedup (log scale). (c) Real-time deadline-miss rate at $D = 1.5$ ms.

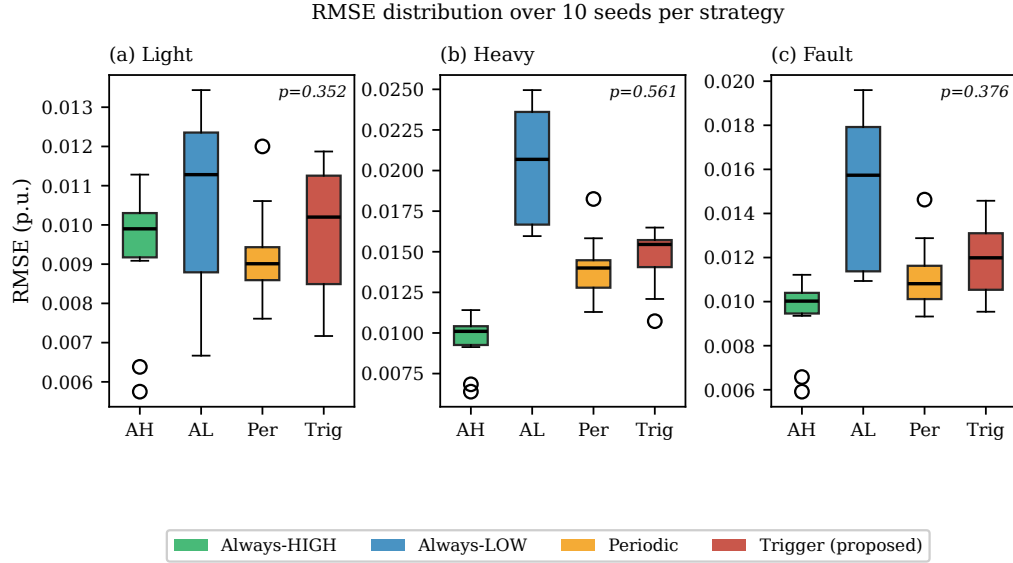


Figure 5: Per-seed RMSE boxplots (10 seeds). AH: always-high; AL: always-low; Per: periodic; Trig: trigger. Annotated p -values from paired t -tests (trigger vs. periodic): none reach $\alpha = 0.05$.

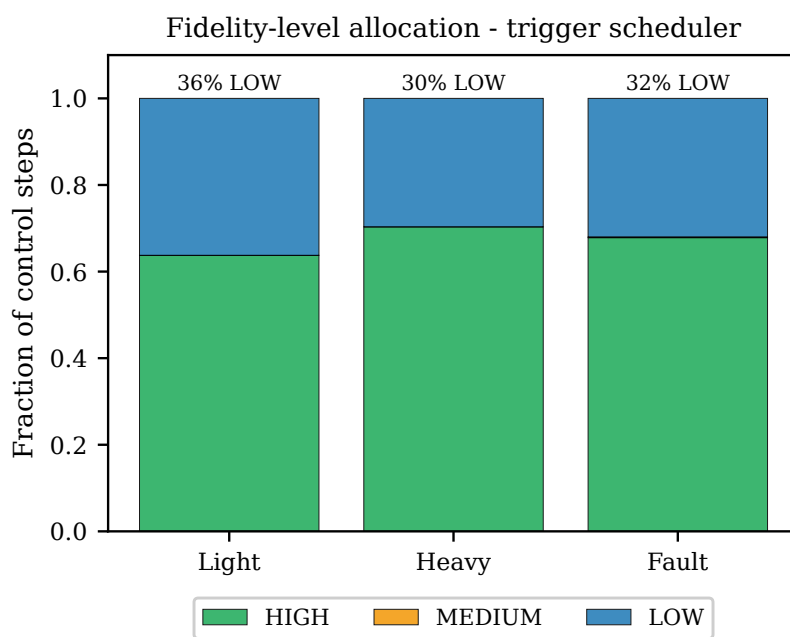


Figure 6: Fidelity-level allocation under the trigger scheduler. LOW usage ranges 30–36% across scenarios.

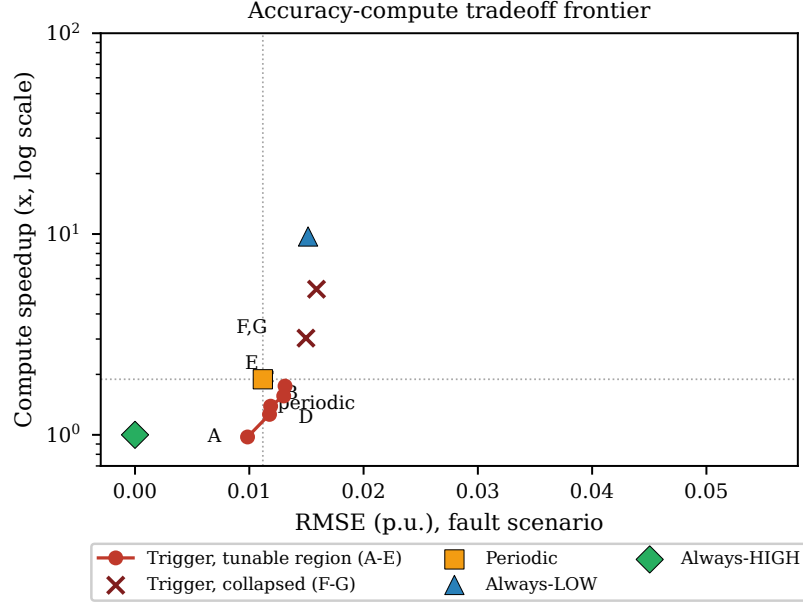


Figure 7: Accuracy-compute trade-off frontier. Config. E nearly matches periodic’s speedup at a 17% accuracy cost. No configuration dominates periodic on both axes simultaneously.

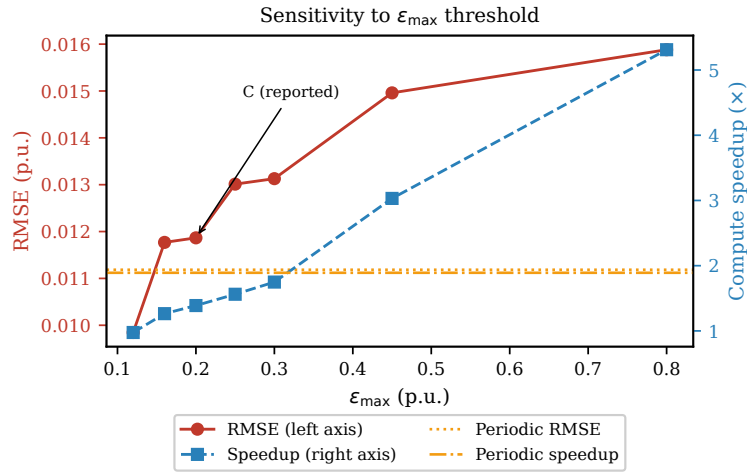


Figure 8: Sensitivity to $\epsilon_{\max}$: RMSE (left, red) and speedup (right, blue) vary continuously. Orange lines mark the periodic baseline. Config. C (reported) is annotated.