

Clockless analogue MVP for physical EP research

Prepared: 2026-07-11

Starting point: COMPONENT_HW_MAP.md

Scope: academic prototype, not a product demonstrator; analogue learning core; no processor, ADC, DAC, FPGA, sample-and-hold bank, phase clock, or digital optimizer in the learning loop.

1. Executive decision

The current \$5k–\$20k Demo-0 is not an MVP. It is an architecture demonstrator that tries to validate, at once, an SRAM-CIM module, bidirectional transpose reads, RoPE mixers, QK/RMS normalization, analogue softmax, SwiGLU multipliers, analogue state handling, ADC/DAC boundaries, a phase sequencer, write programming, and digital supervisory logic. A failure would be difficult to attribute to one mechanism.

The lower-risk research question is narrower and more publishable:

Can a reciprocal nonlinear analogue network continuously learn by a local equilibrium contrast, with the free and nudged states physically present at the same time, without a clock or processor in the learning loop?

Build that first. The recommended core is a **twin-equilibrium analogue learning tile**:

- two small physical replicas of the same reciprocal nonlinear resistor network;
- corresponding trainable edges share one analogue weight capacitor;
- both replicas run continuously, so there are no sequential free/nudged phases;
- an output OTA supplies a true EP-style current nudge; a jumper also provides a voltage-clamped Coupled Learning control condition;
- each edge updates its own capacitor from a local contrast;
- manual boundary-condition selection and a manual learn/freeze switch replace an FPGA;
- a scope or DMM may observe the circuit but is not part of the feedback or learning path.

Recommended first populated board: **8 sign-update edges plus one exact multiplier reference channel**, approximately **\$170–\$300** excluding instruments, tax, and shipping. This is about **17×–118× below** the current \$5k–\$20k board class. A one-edge calibration rig is approximately **\$70–\$130**.

2. What “completely analogue and clockless” can honestly mean

A defensible claim is:

The state evolution, nudge, local learning rule, and weight storage are continuous-time analogue processes. No periodic control signal, processor, converter, sampled state memory, or digitally computed parameter update participates in learning.

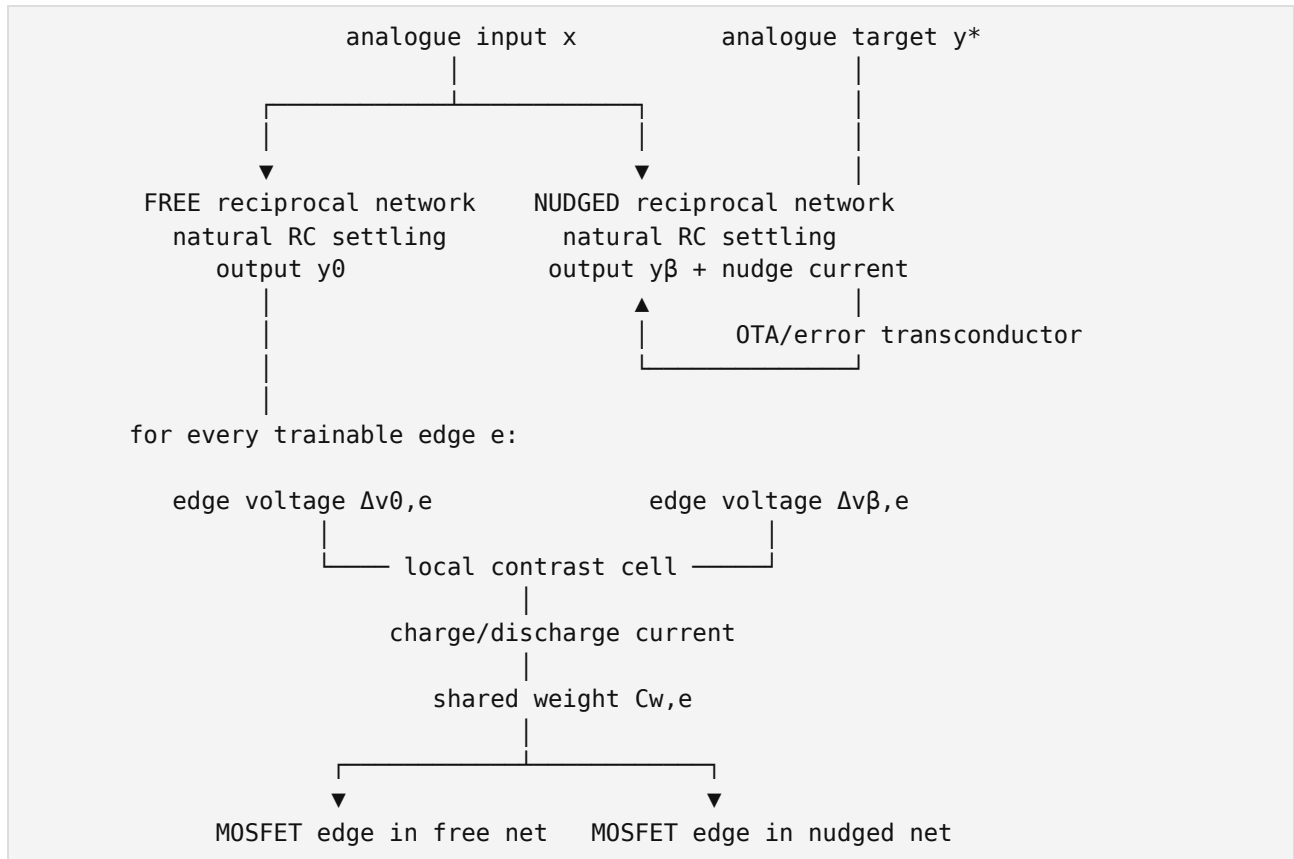
Allowed outside the claim:

- a bench supply;

- an oscilloscope or DMM used only for observation;
- a human changing input/target switches;
- a static mechanical learn/freeze switch;
- offline analysis after an experiment.

A full language-model demonstration cannot honestly be entirely analogue and clockless at the token interface: symbol lookup, presentation of a sequence, cross-entropy labels, and token sampling are discrete operations. The academic MVP should therefore validate the **physical learning primitive**, then add a very small reciprocal attention cell as a second experiment.

3. Core topology



The two replicas must be laid out symmetrically and use matched transistor pairs where practical. The shared capacitor ensures both edge copies always use the same learned weight.

3.1 State dynamics

The node voltages are the states. Resistors/MOSFET conductances and node capacitances produce the relaxation automatically. Add small capacitors only where needed to set a reproducible pole and suppress oscillation; do not build a capacitor-plus-OTA integrator for every abstract model state.

Design target, not a theorem:

$$\tau_{\text{weight}}/\tau_{\text{state}} \geq 10^2, \quad \text{preferably } 10^3.$$

The state should equilibrate much faster than the weight capacitors move. This time-scale separation replaces a settle detector and phase sequencer.

3.2 True EP nudge and the CL control

Populate both modes on the same board.

EP mode — current/force nudge. For voltage outputs, compute the output error and inject a proportional current at the nudged output. A practical small-nudge implementation is

$$I_{\beta} = -g_{\beta} (y_0 - y^*)$$

with an LM13700-class OTA or a discrete transconductor. The polarity is chosen to push the nudged output toward the target. This crosses voltage error with current nudge and is the mode to use for the formal EP/gradient-flow claim.

CL mode — voltage constraint. Buffer a weighted voltage between the free output and the target and impose it on the second replica. This is easier and reproduces the demonstrated clockless Coupled Learning architecture, but it should not be called exact EP.

A two-position switch should select EP-current or CL-voltage nudge. That comparison is itself a useful experiment.

3.3 Exact local contrast

For a conductance-like parameter, the local energy derivative is proportional to the squared voltage drop. Use

$$C_{\{w,e\}} \frac{dV}{dt}_{\{w,e\}} = s_e k \left[(\Delta v_{\{\beta,e\}})^2 - (\Delta v_{\{0,e\}})^2 \right],$$

where $s_e = \pm 1$ accounts for whether increasing capacitor voltage increases or decreases effective conductance.

Do not square twice. Use

$$a^2 - b^2 = (a - b)(a + b),$$

so one four-quadrant multiplier can implement an exact reference channel. The established laboratory circuit used an AD633 with op-amp conditioning and a local capacitor.

3.4 Low-cost sign contrast

The exact multiplier dominates cost. A much cheaper local rule is

$$C_{\{w,e\}} \frac{dV}{dt}_{\{w,e\}} = s_e I_0 \operatorname{sgn}(|\Delta v_{\{\beta,e\}}| - |\Delta v_{\{0,e\}}|)$$

outside a deadband δ , with zero current inside the deadband. Implement it with absolute-value/rectifier stages, a comparator with hysteresis, and two matched charge/discharge current sources.

This rule has precedent in simulated memristor EP hardware because it removes the analogue multiplier. The **continuous capacitor implementation proposed here is an engineering adaptation**, not an already demonstrated result. That is a legitimate research contribution, but it must be labeled correctly.

3.5 Freeze and retention

Use a mechanical toggle or relay to disconnect update current from every weight capacitor. Avoid a clocked switch matrix. Use film capacitors initially; characterize leakage and dielectric absorption. Volatile analogue weights are acceptable for an academic demonstrator, but not for a storage product.

4. Recommended hardware ladder

Rung A — one-edge metrology tile

Purpose: validate the physics before assembling a network.

Populate:

- one free/nudged MOSFET pair;
- one shared weight capacitor;
- one exact AD633 contrast path;
- one sign-only contrast path selectable by jumper;
- one EP-current nudge OTA;
- one CL-voltage nudge path;
- test points for all four edge terminals, multiplier inputs/output, capacitor current, and weight voltage.

Estimated cost: **\$70–\$130**.

Required measurements:

1. transfer surface $I_{\text{update}}(\Delta v_0, \Delta v\beta)$;
2. exact-channel gain and offset;
3. sign-channel boundary, hysteresis, and deadband;
4. zero-contrast drift at $\Delta v_0 = \Delta v\beta$;
5. weight retention with learning frozen;
6. temperature drift;
7. nudge linearity versus β ;
8. time-scale ratio between node settling and weight motion.

Do not build the multi-edge board until this tile gives a stable null at zero contrast.

Rung B — two modular four-edge cards

Build a four-edge card and populate two cards for eight trainable edges. Each card contains:

- four twin MOSFET edges;
- four sign-update cells;

- four weight capacitors;
- local trim/deadband points;
- a shared learn/freeze bus that is static, not clocked;
- edge terminals on headers so the network topology is patchable.

Recommended population: eight sign channels plus **one parallel exact AD633 channel on a selected edge**. This gives a continuously measured exact-versus-sign comparison without buying eight multipliers.

Estimated cost: **\$170–\$300**.

Initial task:

- one-input nonlinear regression with 4–8 manually selected static examples;
- then a two-input task after topology simulation and one-edge characterization.

Do not promise XOR at eight edges. The published nonlinear clockless network used 32 twin edges for XOR. Design the card so eight identical four-edge modules can be stacked later.

Rung C — 32-edge replication-class network

Use eight four-edge cards and the sign-update rule. This approaches the scale of the published nonlinear demonstration while avoiding 32 AD633 multipliers.

Estimated cost: **\$450–\$900**.

This is still below the low end of the original plan by approximately **5.5×–44×** and is large enough for a serious robustness and nonlinear-learning study.

Rung D — reciprocal attention microcell

Only after Rungs A–C work, add a transformer-adjacent cell:

- two tokens;
- one head;
- scalar or two-dimensional state;
- tied key/value or another explicitly energy-based reciprocal construction;
- hardwired causal connectivity;
- no embedding memory, RoPE, RMSNorm, QK norm, SwiGLU, ADC, LM head, or token sampler.

For two alternatives, softmax reduces to a logistic function of a score difference, so a differential pair can replace a general N-way entropic-resistor array. This is an **attention-shaped energy cell**, not an OLMo2 block.

Estimated total including the learning core: **\$300–\$700**.

5. Costed options

Budgetary single-quantity catalogue prices checked on 2026-07-11; prices exclude tax, shipping, instruments, assembly labor, and rework.

Item	Planning role	Unit price used	Planning quantity, 8-edge board	Extended
ALD1106PBL	matched N-MOS array; two twin edges/package	\$9.29	4	\$37.16
AD633ANZ	exact four-quadrant contrast multiplier	\$21.48	0, 1, or 8	\$0 / \$21.48 / \$171.84
TLV274IPWR	quad rail-to-rail op amp	\$1.43	6–8	\$8.58–\$11.44
LM13700	dual OTA; output nudge/current sources	\$1.70	1–2	\$1.70–\$3.40
LM339-class comparator	sign/deadband channels	about \$0.67	2–4	about \$1.34–\$2.68
CD4066-class switch	optional static freeze/routing	about \$0.73–\$0.97	2–4	about \$1.46–\$3.88
Film capacitors	weight storage	about \$0.63 at 1 μ F	8–12	about \$5–\$8
Diodes, resistors, trims	rectifiers, limits, biasing	—	lot	\$15–\$40
PCB/protoboard, headers, test points	physical implementation	—	lot	\$35–\$120
Power rails, protection, spare parts	laboratory overhead	—	lot	\$30–\$90

Resulting envelopes:

Variant	Estimated build cost	Recommendation
One-edge exact/sign calibration tile	\$70–\$130	Build first
8-edge sign-only	\$140–\$250	Cheapest useful network
8-edge sign + one exact reference channel	\$170–\$300	Recommended MVP
8-edge all-exact AD633	\$300–\$500	Only after reference channel works
32-edge sign-update network	\$450–\$900	Replication-class nonlinear demo
Reciprocal two-token attention add-on	+\$100–\$250	Phase 2 only

Aggressive transistor substitution

A CD4007UBE costs about \$0.89 and contains a CMOS dual complementary pair plus inverter. It can be explored on the one-edge tile, but it is not the main-board recommendation: matching, body connections, and device operating region become the dominant uncertainty. Saving roughly \$30 of matched-transistor cost on an eight-edge board is not worth sacrificing the experiment's interpretability.

6. Delete list for the original plan

For the academic MVP, delete these entirely:

- SRAM-CIM, Mythic, HERMES, or other accelerator evaluation board;

- FPGA and phase sequencer;
- DAC and ADC arrays;
- token embedding lookup;
- RoPE DDS and mixers;
- full causal softmax array;
- QK normalization;
- RMSNorm banks;
- SwiGLU multiplier array;
- digital cross-entropy head;
- Adam/Muon or any digital optimizer;
- crossbar write-programming machinery;
- sampled free-phase memories;
- digital settle/retry logic;
- GPU gradient telemetry.

Keep or replace as follows:

Original function	MVP replacement
MVM/crossbar	patchable reciprocal nonlinear resistor network
state integrators	natural RC/KCL state dynamics
sequential free/nudged phases	two continuously operating physical replicas
DAC nudge	OTA current injection
digital phase memory	simultaneous physical state comparison
outer-product update/programming	local capacitor charge/discharge
FPGA control	manual boundary selection and learn/freeze
ADC telemetry	buffered scope/DMM observation outside loop

7. Experimental matrix

The cleanest paper is not merely “it learned.” It should isolate the choices that make the clockless implementation possible.

Factor 1 — nudge type

- EP: voltage error, current nudge;
- CL: voltage error, voltage constraint.

Factor 2 — local update

- exact difference of squares;
- sign-only magnitude comparison;
- sign-only with deadband/hysteresis.

Factor 3 — device quality

- matched ALD1106 edges;
- one-edge CD4007 substitution;
- deliberate mismatch or trim-offset injection.

Factor 4 — nudge magnitude

- small nudge;
- moderate/overclamped nudge;
- measured bias-versus-SNR curve.

Factor 5 — operating conditions

- temperature;
- supply variation;
- capacitor leakage/hold time;
- damaged or disconnected edge;
- topology size.

Proposed acceptance criteria, to be declared before network training:

1. no periodic control signal in the learning loop;
2. no processor, ADC, DAC, or stored free-phase sample used for an update;
3. exact reference channel has a stable zero-contrast null;
4. sign channel agrees with the exact channel on update direction in at least 95% of the specified operating grid;
5. frozen weights remain within the chosen tolerance over the measurement interval;
6. loss decreases consistently across multiple resets, with all failures reported;
7. EP-current and CL-voltage modes are labeled separately.

8. Six-week execution plan

Week 1 — SPICE and one-edge schematic

- simulate the MOSFET edge over the intended node/gate voltage range;
- choose C_w , update current, and state capacitance for at least 100× time-scale separation;
- simulate exact and sign contrast transfer surfaces;
- simulate OTA current nudge and output compliance;
- add hard rails on weight voltage.

Week 2 — one-edge breadboard

- build only the exact path first;
- measure offsets and null drift;
- add sign path and deadband;
- add manual freeze;
- decide whether breadboard leakage is tolerable or proceed directly to PCB.

Week 3 — one-edge characterization

- automate only measurement if desired; do not put automation in the learning loop;
- produce measured update-vector fields;
- identify the safe voltage region;
- lock component values for the modular card.

Week 4 — first four-edge card

- symmetrical placement;
- dense test points;
- patchable node topology;
- one exact reference edge routed in parallel with its sign cell.

Week 5 — eight-edge network

- build a topology selected by SPICE;
- train one static example, then 4–8 manually selected regression points;
- compare exact-reference and sign decisions continuously.

Week 6 — robustness and paper figures

- CL versus EP;
- exact versus sign;
- mismatch and temperature;
- retention and drift;
- energy/power by function;
- scaling estimate to 32 edges and the reciprocal attention cell.

9. Claims this MVP can and cannot support

It can support

- a continuous-time analogue physical-learning core;
- clockless simultaneous free/nudged operation;
- local analogue weight storage and updates;
- processor-free learning after boundary conditions are applied;
- measured exact-versus-sign update behavior;
- measured EP-current-versus-CL-voltage behavior;
- nonlinear supervised learning in a small reciprocal network;
- a credible scaling BOM.

It cannot support

- a full analogue OLMo2 block;
- ordinary independent-Q/K/V attention on a passive reciprocal network;
- end-to-end clockless language modeling;

- nonvolatile weights;
- a claim that the sign-only capacitor rule is exact EP;
- a claim that voltage-clamped Coupled Learning is identical to EP;
- production energy/area numbers derived from a breadboard.

A passive reciprocal network avoids the adjoint/transposed-Jacobian problem by construction, but it also excludes normal non-reciprocal attention. The transformer connection must therefore use tied, energy-based attention or be deferred to an active-adjoint phase.

10. Why this is the better academic MVP

It tests one scientific claim at a time. Every expensive item in the original plan exists to preserve a nearly complete transformer block. That is appropriate after the physical learning primitive is established, not before it.

The proposed sequence produces publishable intermediate results even if the final nonlinear task fails:

- measured clockless local-gradient circuit;
- exact-versus-sign learning rule;
- EP-versus-CL physical comparison;
- mismatch-induced bias and its suppression;
- analogue weight retention and drift;
- modular scaling law;
- reciprocal attention microcell.

The main recommendation is therefore:

Build the \$70–\$130 one-edge tile, then the \$170–\$300 eight-edge hybrid board. Do not purchase a CIM evaluation module, FPGA, DAC/ADC bank, or softmax hardware for the MVP.

11. Primary references used

1. Sam Dillavou et al., **“Machine Learning Without a Processor: Emergent Learning in a Nonlinear Electronic Metamaterial,”** arXiv:2311.00537v2 / PNAS (2024). Demonstrated a fully analogue, clockless nonlinear twin-resistor network; 32 twin edges; local capacitive learning; XOR and nonlinear regression. Its implemented rule is Coupled Learning.
2. Sam Dillavou et al., **“Understanding and Embracing Imperfection in Physical Learning Networks,”** arXiv:2505.22887v2 (2026 revision). Characterizes bias, limit cycles, and system-agnostic suppression of analogue learning imperfections.
3. Joshua A. McGinnis, Xinbo Li, and Yoichiro Mori, **“Coercivity and Local Convergence of Physical Learning in Linear Circuits,”** arXiv:2606.15443 (2026). Distinguishes EP’s force/current nudge from CL’s voltage constraint and analyzes their gradient-flow structure.
4. Seung-Hyun Oh et al., **“Memristor Crossbar Circuits Implementing Equilibrium Propagation for On-Device Learning,”** *Micromachines* 14(7):1367 (2023). Proposes simultaneous free/nudged networks and a sign-only fixed-step update to remove analogue multipliers; results are simulation-oriented for memristor programming.

5. I. M. Elfadel and J. L. Wyatt, **“The ‘Softmax’ Nonlinearity: Derivation Using Statistical Mechanics and Useful Properties as a Multiterminal Analog Circuit Element,”** NeurIPS 1993. Establishes reciprocal/passive circuit properties of an entropic softmax element.
6. **“How to Train Your Resistive Network: Generalized Equilibrium Propagation and Analytical Learning,”** arXiv:2602.03546 (2026). Gives a single-network projector-based route for linear resistive networks; useful as a future replica-reduction direction, not the first nonlinear build.