PhaseWeave

Graph reasoning and cyclic qudits for historical market replay

Research methods and experimental results
20 November 2026

Abstract

We forecast two directed FX triangle margins one second ahead. A paired width-200 experiment adds currency, pair and overnight funding references to quotes, trades and document context on the same historical windows. October mean absolute error falls from 0.156397 to 0.155883 log basis points. A paired UTC-day interval favors the reference model in October and includes zero on September stress and unseen CHF triangles. A six-arm width–data study finds a split-dependent effect from doubling the August window count. We report the evidence clocks and verify the oriented boundary of a genus-one embedding.

1 Graph and observation model

At decision time t, let $G_t = (V, E_t)$ be a directed attributed graph. A vertex is a currency or spot asset. An edge records its source, side, timestamp, rate, and when available the displayed quantity and unit. At most the last observed quote with timestamp at or before t is used. For pair A/B with bid b and ask a,

$$w_{AB} = \log b, \qquad w_{BA} = -\log a.$$

A path $P = (v_0, \dots, v_k)$ with per-leg log fees $f_i \geq 0$ has margin $10^4 \sum_{i=1}^{k} (w_{v_{i-1}v_i} - f_i)$ in log basis points. The verifier treats a path with a stale or absent edge as unavailable.

Proposition 1 (Change of units)**.** *For a closed path, replacing every directed log rate by* $w'_{uv} = w_{uv} + h(v) - h(u)$ *preserves its margin, with unchanged log fees.*

Proof. The additional terms telescope to $h(v_k) - h(v_0) = 0$. □

Bid and inverse ask are separate directed rates. Their sum is $\log(b/a)$, while a mid-price connection can be antisymmetrized for curvature analysis. The observable log margin retains its sign and magnitude even when a compact phase representation wraps. Fenn et al. analyze the distinction between positive quote products and execution of all legs. [2]

2 Forecasts and historical replay

A training sample contains a graph history $[t - L, t]$ and outcomes at later timestamps. For the pilot's one-second future target, all three pairs must refresh strictly after t. The input samples use their own preceding as-of records. Latest-known state is defined by exchange or archive timestamp, and transport latency requires a separate delay model. The broad model will predict both directional margins, a confidence measure, and a continuous field evaluated at held-out quote query points.

The predeclared chronological partitions have disjoint support intervals. An embargo excludes any candidate whose history or target enters a neighboring partition. Transformations, vocabularies, reference graphs, and calibration parameters are fit on training only. Duplicate snippets, overlapping

events, and near-identical trajectories must be grouped before the final held-out scoring. Separate a new currency-family or venue test from a later-time test.

Marginal forecasts are distinct from route decisions. A replay of a multi-leg route advances simulated execution time at each leg, tracks the quantity remaining, and accounts for depth, fees and refusals. For a directed edge with output rate r_i and input capacity c_i, the capacity expressed at the first leg is $c_i / \prod_{j<i} r_j$ under the assumed simultaneous book and filled-prefix model. The minimum of these capacities is an upper bound for the initial size under that model. Displayed orders, observed trades, and hypothesized fills must be recorded separately.

3 Surface incidence and cyclic features

Follow the clean-dessin construction in the GRH research atlas. [1] Let α reverse darts and let σ rotate darts around their tails. Require a fixed-point-free involution, a valid permutation, and transitivity. The permutation $\phi = \alpha\sigma^{-1}$ traverses the inverse oriented face boundaries in this convention. The genus of the corresponding connected oriented surface is determined by

$$2 - 2g = c(\sigma) - \frac{|H|}{2} + c(\phi),$$

where H is the dart set and c counts permutation cycles. Market graph connectivity does not choose a rotation system. The rotation rule must be explicit, and orientation and genus controls test for dependence on that choice.

A six-dart torus illustrates why the oriented attaching word matters. After subdividing its three parallel edges, the counts become $(V, E, F) = (5, 6, 1)$. With all six new edges oriented from black to white, set $p = (1, -4)$, $q = (2, -5)$ and $r = (3, -6)$. Each pair travels between the same two black vertices through a different white vertex. The face $pr^{-1}qp^{-1}rq^{-1}$ closes in twelve steps, uses each edge once each way, and leaves pq^{-1}, pr^{-1} as an integral basis of H_1. The six-letter word printed in the historical subdivision has nonzero boundary $3(v_{\text{black},2} - v_{\text{black},1})$. [10]

Let B_1 be the oriented vertex–edge boundary and B_2 the edge–face boundary. Each face telescopes, giving $B_1 B_2 = 0$. For a real antisymmetric link field x, the face curls are $B_2^\top x$. Functionals on independent cycles in $\ker B_1 / \operatorname{im} B_2$ measure global periods after representatives have been chosen. Under rescaling $x \mapsto x + B_1^\top h$, both face curls and closed-cycle functionals are invariant. A global period can depend on the representative when the link field has nonzero face curl.

The locally gentle algebra uses a vertex for each paired dart, arrows $a_h : [h] \to [\sigma^{-1}h]$, and the prohibited composite $a_h a_{\phi h}$ in travel order. A graph feature map records selected finite observables of the cellulation. The completed order, representations of its finite observables and the market field have different mathematical types. With matrix-valued transport, the product on a closed path transforms by conjugation, so traces are natural invariant readouts. This finite construction motivates lattice-style comparisons.

The ToricGT decrement convention for a q-level state is $Ce_j = \omega^j e_j$, $Se_j = e_{j-1 \bmod q}$, with $\omega = e^{2\pi i/q}$. Direct action on e_j gives $SC = \omega CS$ and $C^q = S^q = I$. The historical surface-order companion shift uses the increment direction and a formal uniformizer at the wrap. The unitary circuit uses a unit-modulus complex wrap. Use interference or Hermitian components of a Weyl expectation to observe relative phases.

4 Graph models and ablations

Compare GCN, GAT, typed MPNN and a Node/Edge TokenGT backbone with matched data and compute. [3] The shared graph schema also represents the ConvexTok byte DAG, the prediction trace

and typed evidence. A conditioned discrete-flow head proposes valid graph edits or ordered cycles. A basis operator head reconstructs the time-varying margin field. Tropical attention and cyclic-qudit features are varied independently. Compare the circuit against trigonometric, frozen, unentangled and phase-scrambled features.

The reasoning model receives the graph and verified correspondence with retrieved analogies. Candidate reasoning branches use typed edges, an evidence reference, and a terminating action. A specialized track receives the same graph, scoring function, and candidate budget. GraphCG-style concept directions are tested by interventions and equal-norm random alternatives. [4] The GFlowNet terminal sampler has a positive reward $R(x)$ tied to legal completed actions and abstention. Its trajectory-balance residual is

$$\log Z + \sum_i \log P_F(s_{i+1} \mid s_i) - \sum_i \log P_B(s_i \mid s_{i+1}) - \log R(x).$$

Evaluate this residual and the empirical terminal distribution against finite enumeration. [5] A pre-trained graph-conditioned language model, its QLoRA variant and a language-free graph model share the same verifier. [6]

The completed pilot included 24 width-200 proxies, two wider transfers and 12 matched circuit and attention controls. Validation selected a separate checkpoint for each width in subsequent full-epoch training. The quote-only width-200 model reached October MAE 0.155829 under its own schedule. We use separately matched arms below to estimate the effect of adding reference inputs. [9, 11]

5 Measured feasibility

On the verified RTX 4090 host, the HistData pilot validated 22,819,287 quote rows from nine files. The sources were EUR/USD, USD/JPY and EUR/JPY in August through October 2026. Tied timestamps use the minimum bid and maximum ask. The pilot used a 250 ms maximum age, a one-second forecast horizon, and 0.1 log-bps extra fee per leg.

Month	Future samples	Positive margins	Persistence MAE (bps)
August	148,646	2	0.1992
September	247,696	18,878	0.2192
October	241,314	4	0.1999

September 17 and 18 account for 17,511 positive labels. Every source volume was zero. These facts motivate session-level uncertainty and checking the positive-margin regime with an independent feed. Three Gate spot depth logs provide a follow-on reconstruction dataset with 1,940,033 rows. These measurements are feasibility diagnostics.

Explicit statevector simulations used batches of four 16-qubit states, eight eight-qutrit states and eight six-ququint states at four layers each. The maximum state norm error on the GPU was 6.20×10^{-6}. The CPU–CUDA differences were at most 1.06×10^{-6} for the mean observable and 2.90×10^{-6} for the tested derivative. The largest difference between automatic differentiation and a centered numerical derivative was 1.23×10^{-5}. These checks establish reference backend agreement for the tested patches.

6 Unique data at matched exposure

Chinchilla compares text-training compute allocations, while recent data-constrained studies explicitly vary unique data and reuse [7, 8]. A PhaseWeave graph window allocates 144 spatial positions and yields 16 derived temporal positions. These counts track graph compute and provide no word-token

equivalence. In a calibration on actual August windows, each arm trained the same 211,329-parameter width-64 model for 64 updates and 4,096 window presentations, with the same implementation and initial weights per seed. The nested window counts 256, 1,024 and 4,096 correspond to sixteen, four and one passes.

Fit seed	256 unique	1,024 unique	4,096 unique
20261125	0.245708	0.241932	0.241897
20261126	0.238083	0.237282	0.239582
20261127	0.240125	0.242018	0.241801

September validation subset MAE, log bps. Persistence is 0.242495 on the same 1,024 validation windows.

The lowest validation MAE occurs at a different unique training count for each of the three seeds. This calibration addresses reuse and does not identify a general compute-optimal scaling frontier. We transferred proxy-selected settings to three widths with 65,536 or 131,072 distinct August windows per width [9]. The expansion is nested and preserves all evaluation splits byte for byte. Within each width, both arms used 64 windows per update and shared a 1,024-step warmup. September validation selected from 131,072, 262,144, and 393,216 presentations. October results use those selected models.

Parameters	Original MAE	Expanded MAE	Change [95% interval]
2,020,265	0.157922	0.158077	+0.000155 [+0.000010, +0.000308]
10,080,513	0.157773	0.157866	+0.000093 [−0.000068, +0.000251]
34,695,297	0.156831	0.158445	+0.001614 [+0.001388, +0.001851]

October MAE in log bps. Exploratory paired UTC-day bootstrap.

Expansion decreased stress error at all widths, with intervals excluding zero at the two smaller widths. The 34.70M expanded arm increased unseen CHF error by 0.005277 log bps. Its October and CHF checkpoints were selected at 393,216 presentations, versus 262,144 for the original arm. The contrast includes this training-duration difference. Each arm has one initialization, leaving variation across seeds unmeasured. Scaling data, exposure curves and run configuration.

7 Paired reference experiment

The frozen corpus has 294,912 FX windows. Each width-200 arm forecasts the one-second margins for both triangle orientations and receives identical quote histories, seven Gate trade slots and two document slots. The added reference inputs are three currency records, three BIS pair records and five overnight funding series. Both arms use the same 20-slot architecture and 2,150,866 trainable parameters. The control zeros the eleven reference vectors and masks them. Each arm normalizes its observed August inputs. Matched initialization, epoch permutations and learning-rate changes give both arms 16,384 updates across 16 full passes of 65,536 windows each. Validation selected epoch 15 with MAE 0.175562 for the control and epoch 16 with MAE 0.174222 for references. [12]

Holdout	Control	References	Paired change [95% interval]
September stress	0.159801	0.160031	+0.000230 [−0.000170, +0.000730]
October	0.156397	0.155883	−0.000515 [−0.000862, −0.000185]
Unseen CHF	0.177248	0.176691	−0.000557 [−0.001459, +0.000540]

One-second margin MAE in log bps. Change equals references minus control, computed before rounding.

The 2,500-draw paired bootstrap resamples UTC days. September stress and October contain 65,536 windows each over 12 and 26 days, respectively. Unseen CHF contains 32,768 windows over 26 days. October favors references under this shared initialization. Both other intervals span zero. The separately trained quote-only width-200 model has a slightly lower October MAE than either paired model. [12, 11]

Training contrast and fixed-weight masks. The two arms shared their August order and initialized trainable tensor bytes. The mean validation MAE across arms governed the common learning-rate and plateau schedule. Holdout labels were opened after validation selection. An additional intervention kept the reference checkpoint fixed. Masking its BIS pair slots changed October MAE by $+0.007558$ $[+0.006708, +0.008588]$ log bps. Masking its five overnight funding slots changed it by -0.000683 $[-0.000945, -0.000420]$. The contrasts estimate sensitivity to withholding observed inputs from this fitted model. Static references may encode triangle identity and funding records may encode the calendar. The single shared initialization and overlapping currencies limit inference across seeds and markets. [12]

Continuous controls. Eight-neighbor retrieval used 5,751 August windows from the frozen forecaster's representation. Validation selected zero mixture weight for all four retrieval methods. In October, standalone latent retrieval scored 0.173232 against 0.155829 for the frozen model. A separate supervised continuous search used 48 verifier evaluations for either branching or independent paths. October MAEs were 0.166094 for branching, 0.177207 for independent paths and 0.164349 for direct use of the fitted quantile median. In a separate two-seed GFlowNet test against a specified conditional Gaussian reward, held-out October terminal KL was 0.0235 and 0.0197 nats. That test measures sampler agreement with a known target density. [14]

Gate derivative and depth coverage. A separate UTC-aligned companion contains 1,104 applied funding settlements for four USDT perpetuals. All 294,912 FX anchors have all four rates when their settlement labels plus 60 seconds precede the anchor, subject to a 24-hour carry limit. Historical subscriber receipt times are unavailable. These derivative rates are distinct from the five overnight funding inputs in the paired training experiment. The only spot depth hour is 30 October, 12:00–13:00 UTC. It intersects 457 holdout anchors and zero training anchors. At the anchor, all three books meet a 250 ms age limit in 113 cases. Across all 16 frames the same limit admits zero windows. These companion data establish coverage and timing assumptions. [13] At the one-second age limit, 286 anchors have all three books fresh and 24 windows pass at every frame (14 October test and 10 CHF). All 1,104 archived rates match the historical endpoint. Their retrieval receipts postdate the FX windows, leaving zero anchors with proof of pre-anchor retrieval. The rate's unit is a signed fraction of contract position value per settlement. The spot book replay adds 100 ms to merged update labels. Trade, funding, depth and FX instruments retain their venue-specific identities. [13]

References

[1] A. Schreiber, *GRH research atlas*, pinned source e941b50.

[2] D. Fenn et al., *The mirage of triangular arbitrage in the spot foreign exchange market*, arXiv:0812.0913.

[3] J. Kim et al., *Pure transformers are powerful graph learners*, arXiv:2207.02505.

[4] S. Liu et al., *Unsupervised discovery of steerable factors when graph deep generative models are entangled*, arXiv:2401.17123.

[5] N. Malkin et al., *Trajectory balance*, arXiv:2201.13259.

[6] T. Dettmers et al., *QLoRA*, arXiv:2305.14314.

[7] J. Hoffmann et al., *Training Compute-Optimal Large Language Models*, arXiv:2203.15556.

[8] N. Muennighoff et al., *Scaling Data-Constrained Language Models*, arXiv:2305.16264.

[9] G. Yang et al., *Tensor Programs V*, arXiv:2203.03466.

[10] A. Schreiber, *Historical torus subdivision*, oriented edge list and face word.

[11] PhaseWeave, *Width and circuit comparisons*, experiment results.

[12] PhaseWeave, *Paired reference forecast*, epoch history, matched arms and masks.

[13] PhaseWeave, *Gate derivative and spot-depth coverage*, frozen replay specification.

[14] PhaseWeave, *Continuous forecast and sampling experiments*, methods and controls.