CIRCUIT DEPTH AND GRAPH ARCHITECTURE

Six-qutrit clock circuits for one-second FX forecasts

We compared the effect of circuit depth across three width-200 graph backbones. The 144 fitted models share the same market windows, optimization schedule and parameter count. The outcome is absolute forecast error for two directed currency-triangle margins, measured in log basis points. A negative depth contrast means that four applications of the circuit produced lower error than one.

Ten prespecified comparisons on November afternoons

For each forecast origin, we average absolute error over the two directions. Within each backbone and attention rule we then subtract the depth-one loss from the depth-four loss and average over two fixed training seeds. The upper six rows report that effect for the coherent entangling arm. The lower four subtract the corresponding TokenGT depth effect at the same attention setting. An interaction can favor a backbone even if its absolute forecast error is larger.

Backbone	Attention	Mean change	Pointwise 95%	Family 95%
GCN	softmax	−0.00797	[−0.00955, −0.00639]	[−0.00977, −0.00618]
GCN	tropical	−0.00086	[−0.00189, +0.00006]	[−0.00265, +0.00094]
GCN + TokenGT	softmax	−0.00718	[−0.00857, −0.00568]	[−0.00898, −0.00539]
GCN + TokenGT	tropical	−0.00111	[−0.00186, −0.00034]	[−0.00291, +0.00068]
TokenGT	softmax	−0.00191	[−0.00330, −0.00061]	[−0.00371, −0.00012]
TokenGT	tropical	+0.00111	[+0.00033, +0.00191]	[−0.00068, +0.00291]
GCN	softmax	−0.00606	[−0.00750, −0.00453]	[−0.00785, −0.00426]
GCN	tropical	−0.00197	[−0.00332, −0.00070]	[−0.00377, −0.00018]
GCN + TokenGT	softmax	−0.00527	[−0.00629, −0.00416]	[−0.00706, −0.00347]
GCN + TokenGT	tropical	−0.00223	[−0.00283, −0.00172]	[−0.00402, −0.00043]

Negative values indicate error reduction. In the lower four rows, negative values indicate a greater reduction (or smaller increase) relative to TokenGT. Every interval uses the same eligible origins and direction convention.

How to read the bands

All 2,500 bootstrap draws select ten represented UTC dates with replacement, using seed 20261224. A selected date carries all its eligible windows and contrasts. Dividing the sum of selected daily loss differences by the selected row count reproduces the row-weighted mean. The pointwise interval uses the 2.5th and 97.5th percentiles. For each draw we also take the largest absolute departure from the observed mean across the ten contrasts. The 95th percentile of these maxima supplies a common, unstudentized radius for the family bands.

Observed pattern. All three softmax backbones reduced November error with greater depth, with family bands entirely below zero. Each of the three tropical depth bands included zero. Both GCN variants had a more favorable depth response than TokenGT under each attention rule. The largest reduction was 0.00797 log bps for GCN with softmax. For both GCN/softmax and TokenGT/softmax, the depth effect changed direction on CHF (page 2). These paired comparisons concern two fixed seeds and previously scored market dates. They measure forecast error, and provide no estimate of trading profit or performance on new sessions.

MODEL, COHORT AND REPRODUCTION

What changed across the 144 fits?

The circuit and backbone

Each 16-frame window contains three currency nodes and six directed edges per frame. A GCN uses two directed degree-normalized convolution layers. The hybrid runs one GCN layer and one token-attention layer. TokenGT uses two token-attention layers. All three apply a temporal attention block. The softmax/tropical switch therefore changes only temporal attention on GCN, and both spatial and temporal attention on the other backbones. All attention blocks have eight heads.

The six edge rates generate six input angles. Each entangling layer applies the same local and adjacent-pair phases to the $3^6 = 729$ complex amplitudes, then local unitary mixers. The product control removes joint bond phases while keeping coherent local operations. In all arms, neighbor angles and the five coupling parameters still enter the local mixers. The dephased control propagates probabilities by squared mixer magnitudes. The max-plus control keeps the best log path and normalizes its final scores. Each arm returns the six one-wire expectations of $\mathrm{diag}(1, -1/2, -1/2)$. Their projection enters the pooled graph state; its final temporal readout adds two residuals to persistence. These are classical simulations.

There are three depths (1, 2, 4), four propagation arms and two seeds (20261225, 20261226) for each backbone/attention pair. Repeated layers share weights. Each model has **1,538,073 parameters**, including 1,447 in the patch. Each fit sees 8,192 August windows over three epochs, or **384 updates** at batch size 64. Training uses μAdamW with learning rate 0.0003, smooth L_1 loss ($\beta = 0.25$) and gradient-norm clipping at 1. The minimum-MAE checkpoint on 2,048 September validation rows wins, with earlier epochs breaking ties. Initialization matches within backbone and seed, and shuffles match across all fits.

How the same depth contrast varies by panel

Backbone	Attention	Validation	Sept. stress	October	New CHF
GCN	softmax	−0.00460	−0.00416	−0.00488	+0.00587
GCN	tropical	−0.00110	−0.00034	−0.00043	−0.00285
GCN + TokenGT	softmax	−0.00564	−0.00584	−0.00487	−0.00173
GCN + TokenGT	tropical	−0.00130	−0.00187	−0.00067	−0.00072
TokenGT	softmax	−0.00139	−0.00154	−0.00036	+0.00893
TokenGT	tropical	+0.00068	+0.00226	+0.00201	+0.01052

Entries are coherent entangling depth-four minus depth-one mean absolute errors in log basis points, with 2,048 rows per column. Validation participated in epoch selection. The CHF triangles share some market edges with the development triangles.

The November panel forecast 2,018 origins between 14:00 and 15:00 UTC on ten previously examined sessions. At 1,329 origins, all target quotes were strictly later than the anchor and within 250 ms of the one-second target. Before scoring, the forecast seal bound the selected checkpoints, row clocks and all 145 forecast arrays on each panel, including persistence. Independent replay checked the 432 epoch checkpoint hashes and the ten family bands. All **24** selected TokenGT/softmax checkpoint digests matched the earlier repeated-depth study. The full run took 601.4 seconds with peak allocated GPU memory ranging from 365.9 to 709.9 MiB per fit. The source package includes the aggregate inputs and the table generator.