FiT: From Market Language To Market Decisions

Anonymous authors
Paper under double-blind review

Abstract

We propose **FiT**, **Fi**nancial **T**ransformer, which builds a unified pipeline from market representation to future prediction and downstream adaptation. On the data side, we introduce the FiT Dataset, expanding market coverage to more than 100 exchanges and temporal resolution down to 1-second K-lines, thereby enabling broader generalization across sampling frequencies. With more than 200 billion K-line records—17 times the size of the Kronos corpus—it is , to the best of our knowledge, the largest financial K-line corpus to date. For pretraining, we present Rotated Leech Spherical Quantization (RLSQ) for K-line tokenization designed to better capture the geometry of candlestick dynamics. Compared with the Kronos tokenizer, it substantially improves reconstruction performance, reducing reconstruction error by 62% while improving directional accuracy and K-line structural validity by 21% and 22%, respectively. For the predictor, we augment next-token prediction (NTP) with Future Summary Prediction (FSP), yielding 52% and 95% improvements in zero-shot IC and RankIC, respectively, over the best-performing baselines. For post-training, we perform BiRank GRPO to adapt FiT to downstream tasks, producing task-specific experts that are subsequently consolidated into the base model through multi-teacher on-policy distillation (MOPD). This procedure improves downstream performance while largely preserving the capabilities and generalization of the pre-trained model. We further develop a hybrid optimization strategy tailored to large-scale financial time-series pretraining, combining Muon, Lion, and AdamW across parameter groups, which reduces validation loss by 5.6% relative to an AdamW-only baseline. Together, these advances trace a complete path from market language to market dynamics to market decisions: RLSQ learns a high-fidelity discrete representation of K-line behavior, FSP learns to anticipate its future evolution, and BiRank GRPO with MOPD translates these capabilities into downstream tasks. FiT thus provides a unified foundation for financial K-line modeling that moves beyond representing the market toward anticipating its dynamics and adapting them to downstream decisions.

1 Introduction

Large-scale pretraining on discrete token sequences has become the dominant recipe for general-purpose models in language and vision, and the same recipe is now reshaping time-series analysis: time-series foundation models (TSFMs) pretrained on broad corpora deliver strong zero-shot forecasts across domains (Ansari et al., 2024; Das et al., 2024; Woo et al., 2024; Shi et al., 2025). Financial markets are among the most consequential and the most difficult targets for this paradigm. Their universal record is the K-line (candlestick) sequence—**O**pen, **H**igh, **L**ow, and **C**lose prices together with traded **V**olume and **A**mount—which summarizes price formation for every asset, venue, and sampling frequency. Yet K-line data exhibit a low signal-to-noise ratio, heavy tails, and pronounced non-stationarity (Cont, 2001), and general-purpose TSFMs transfer poorly to them. Kronos (Shi et al., 2026) showed that this gap can be narrowed by treating the market as a language: a dedicated tokenizer maps each K-line to a discrete token, and an autoregressive Transformer pretrained on billions of such tokens outperforms general-purpose TSFMs on financial tasks.

Despite this progress, we argue that a financial foundation model must do three things— *represent* the market, *anticipate* its dynamics, and *adapt* to decisions—and that existing models fall short on each.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Representation. The tokenizer is an information bottleneck: whatever the quantizer discards, no predictor can recover. Kronos (Shi et al., 2026) uses binary spherical quantization (BSQ) (Zhao et al., 2025), whose codebook is the vertex set of a hypercube projected onto the unit sphere. This makes quantization and entropy estimation cheap, but yields a poor spherical code: the closest pair of codewords is only $2/\sqrt{d}$ apart, about half the minimum distance of the densest known lattice code at a comparable bit budget. BSQ is also prone to codebook collapse without proper entropy regularization (Zhao et al., 2026). K-line data make quantization errors especially costly. A K-line must satisfy ordering constraints across its fields: $L \leq O, C \leq H$, and the volume-weighted average price A/V (traded amount over volume) must also lie in $[L, H]$. A coarse reconstruction can easily violate these constraints, and the most decision-relevant attribute of a K-line—the direction of the move—often hinges on price differences smaller than a single quantization cell.

Anticipation. Next-token prediction (NTP) under teacher forcing rewards locally accurate, one-step fits. In markets, the next bar is dominated by noise, whereas the predictable structure resides in slowly varying properties of the future trajectory, such as volatility, range, and trading activity (Cont, 2001). A model trained with NTP alone has little incentive to encode where a sequence is heading beyond the next token, a shortcoming that has also been documented in language modeling (**?**Gloeckle et al., 2024; Mahajan et al., 2026a).

Adaptation. Pretraining likelihood is not what practitioners optimize. Downstream use is judged by cross-sectional, rank-based criteria such as the information coefficient (IC), RankIC, and long–short portfolio returns, and a lower cross-entropy does not necessarily yield a better ranking. Closing this gap by task-specific fine-tuning produces a collection of specialists rather than one model, and tends to erode the general capabilities acquired in pretraining.

These limitations are compounded by data. Existing K-line corpora cover a limited set of exchanges and no sampling frequency finer than one minute, which restricts generalization across both markets and time scales.

We present **FiT**, **Fi**nancial **T**ransformer, a foundation model for financial K-line data that addresses these limitations within a single pipeline, from market representation to future prediction to downstream adaptation.

Data. We construct the FiT Dataset, which spans 103 exchanges and extends temporal resolution down to 1-second K-lines. With 208.5 billion K-line records, it is, to the best of our knowledge, the largest financial K-line corpus to date.

Market language. We propose Rotated Leech Spherical Quantization (RLSQ) for K-line tokenization. RLSQ quantizes normalized 24-dimensional latents onto the first shell of the Leech lattice Λ_{24}, whose 196,560 minimal vectors form the optimal kissing configuration in 24 dimensions (Conway & Sloane, 1999; Cohn et al., 2017). The codebook is fixed and parameter-free, so it needs neither commitment nor entropy regularization and cannot collapse. To train the encoder through this non-differentiable map, RLSQ replaces the straight-through estimator with a rotation-based gradient (Fifty et al., 2025) that preserves the angular relation between a latent and its assigned codeword. Relative to the Kronos tokenizer, RLSQ reduces reconstruction error by 62% while improving directional accuracy and K-line structural validity by 21% and 22%, respectively.

Market dynamics. We augment NTP with Future Summary Prediction (FSP) (Mahajan et al., 2026a): an auxiliary head predicts a compact learned summary of the future trajectory, supplied by a teacher that reads the sequence in reverse time. The auxiliary branch is discarded after training, so FSP adds no inference cost. It improves zero-shot IC and RankIC by 52% and 95%, respectively, over the best-performing baselines.

Market decisions. We post-train FiT with BiRank GRPO, a group-relative policy optimization scheme (Shao et al., 2024) whose reward couples cross-sectional rank correlation with the return spread between top- and bottom-ranked assets, thereby aligning the model directly with decision-oriented criteria. The resulting task-specific experts are then consolidated into the base model through multi-teacher on-policy distillation (MOPD) (Agarwal et al., 2024; Xiaomi LLM-Core Team, 2026), in which the student learns from the experts on its own rollouts. This procedure improves downstream performance while largely preserving the capabilities and generalization of the pretrained model, and it yields a single model rather than a set of specialists.

Optimization. Finally, we develop a hybrid optimization strategy for large-scale financial time-series pretraining that assigns Muon (Jordan et al., 2024) to hidden weight matrices, Lion (Chen et al., 2023) to one-dimensional parameters, and AdamW (Loshchilov & Hutter, 2019) to embeddings and the output head. It reduces validation loss by 5.6% relative to an AdamW-only baseline.

Together, these components trace a complete path from market language to market dynamics to market decisions. Our contributions are as follows:

- We propose FiT, a novel financial foundation model addressing representation, anticipation, and adaptation in one pipeline: RLSQ tokenizes each K-line onto the first shell of the Leech lattice Λ_{24}—a fixed, parameter-free codebook that cannot collapse, trained with rotation-based gradients; Future Summary Prediction augments next-token prediction with an auxiliary reverse-time-teacher head that is discarded after training; and BiRank GRPO with multi-teacher on-policy distillation aligns the model with rank-based financial objectives while consolidating task experts back into a unified model.
- We construct the FiT Dataset, to our knowledge the largest financial K-line corpus to date: over 200 billion records from more than 100 exchanges—17$\times$ the Kronos corpus—with resolution extended down to one second.
- We evaluate FiT model zoo at every level of the pipeline. RLSQ cuts reconstruction error by 62% and improves directional accuracy and structural validity by 21% and 22% over the Kronos tokenizer, at a lower bit rate; FSP improves zero-shot IC and RankIC by 52% and 95% over the strongest baselines; and FiT sets a new state of the art across price series forecasting, volatility forecasting, and synthetic K-line generation.

2 Preliminaries

2.1 K-line Data and Notation

A K-line (candlestick) summarizes the trading of one instrument over one bar of a fixed sampling frequency as $x_t = (o_t, h_t, l_t, c_t, v_t, a_t) \in \mathbb{R}^6$: the open, high, low, and close prices, the traded volume, and the traded amount. By construction a K-line satisfies the ordering constraints

$$l_t \;\le\; \min(o_t, c_t) \;\le\; \max(o_t, c_t) \;\le\; h_t, \tag{1}$$

and we call a vector in $\mathbb{R}^6$ *structurally valid* if it satisfies equation 1. Two signed quantities describe the direction of price movement: the *intra-bar move* $d_t^{\text{in}} = c_t - o_t$, the body of the candlestick, and the *gap* $d_t^{\text{gap}} = o_t - c_{t-1}$ between consecutive bars. A window is a sequence $X = (x_1, \ldots, x_T)$ with calendar stamps $(\tau_1, \ldots, \tau_T)$.

2.2 Discrete Autoregressive Modeling of K-lines

Following Kronos (Shi et al., 2026), a K-line foundation model consists of two components trained in sequence. A *tokenizer* is an autoencoder with a discrete bottleneck (van den Oord et al., 2017): an encoder E maps the normalized window to latents $z_t \in \mathbb{R}^n$, a quantizer Q assigns each latent to a codeword q_t of a codebook $\mathcal{C}$ with index $k_t \in \{1, \ldots, V\}$, and a decoder D reconstructs $\hat{x}_t$ from the codewords. A *predictor* is a decoder-only Transformer trained on the token sequences with next-token prediction (NTP),

$$\mathcal{L}_{\text{NTP}} = -\sum_{t=1}^{T-1} \log p_\theta(k_{t+1} \mid k_{\le t}, \tau_{\le t})\,. \tag{2}$$

Forecasts are generated by sampling future tokens autoregressively and decoding them with D; averaging several sampled paths gives a point forecast.

2.3 Spherical Codes and the Leech Lattice

A spherical code is a finite set of points on $\mathbb{S}^{n-1}$; used as a quantizer, it assigns a unit vector to its nearest point. Its quality as a quantizer is determined by how evenly the points are spread over the sphere.

Table 1: Composition of the Leech-lattice kissing codebook. All codewords have equal norm, and the inner product of any two unit-normalized codewords lies in $\{0, \pm\frac{1}{4}, \pm\frac{1}{2}, \pm 1\}$, so distinct codewords are separated by at least 60°. The codebook size $196{,}560 \approx 2^{17.58}$ requires 18 bits per index.

Shape	Pattern	$\lvert\text{supp}\rvert$	Counting	# Vectors	Fraction	Unit-norm entries
I	$(\pm 4, \pm 4, 0^{22})$	2	$\binom{24}{2} \cdot 2^2$	1,104	0.56%	$\pm 1/\sqrt{2}$
II	$(\pm 2^8, 0^{16})$	8	$759 \cdot 2^7$	97,152	49.43%	$\pm\sqrt{2}/4$
III	$(\mp 3, \pm 1^{23})$	24	$24 \cdot 2^{12}$	98,304	50.01%	$\mp 3\sqrt{2}/8, \pm\sqrt{2}/8$
Total				196,560	100%	

Table 2: Inner-product distribution of the codebook relative to a fixed codeword v (unit-normalized). The distribution is symmetric because the codebook is centrally symmetric; each codeword has exactly.

$\langle u, v\rangle$	1	$\frac{1}{2}$	$\frac{1}{4}$	0	$-\frac{1}{4}$	$-\frac{1}{2}$	-1
Angle	0°	60°	75.5°	90°	104.5°	120°	180°
# Codewords	1	4,600	47,104	93,150	47,104	4,600	1

The Leech lattice Λ_{24} is the unique even unimodular lattice in $\mathbb{R}^{24}$ without vectors of squared norm 2 (Conway & Sloane, 1999). It gives the densest sphere packing in 24 dimensions (Cohn et al., 2017). Its shortest nonzero vectors have squared norm 4; there are 196,560 of them, and they form the unique optimal kissing configuration in 24 dimensions (Odlyzko & Sloane, 1979; Levenshtein, 1979). Up to a common scale, they have three shapes: $(\pm 4, \pm 4, 0^{22})$ in any two coordinates (1,104 vectors); $(\pm 2^8, 0^{16})$ with support on an octad of the binary Golay code and an even number of minus signs (97,152 vectors); and $(\mp 3, \pm 1^{23})$ with signs determined by a Golay codeword (98,304 vectors). Normalized to unit length, the inner product of any two of them lies in $\{0, \pm\frac{1}{4}, \pm\frac{1}{2}, \pm 1\}$, so distinct points are at least 60° apart.

3 METHODOLOGY

FiT is built in three stages that mirror the path from market language to market dynamics to market decisions (Figure 1). A tokenizer based on Rotated Leech Spherical Quantization (RLSQ) maps each K-line to a discrete token ; an autoregressive predictor is pretrained on the resulting token sequences with next-token prediction augmented by Future Summary Prediction ; and the pretrained predictor is adapted to decision-oriented objectives with BiRank GRPO, after which the resulting experts are consolidated by multi-teacher on-policy distillation . All stages share one data protocol and one optimization recipe .

3.1 ROTATED LEECH SPHERICAL QUANTIZATION

RLSQ quantizes every bar onto a fixed shell of the Leech lattice and trains the encoder through it with the rotation trick. On this shell every token lies far from its codeword ($\theta \approx 39^\circ$). We show that the two usual gradient surrogates differ exactly, and only, in how they move this angle, which is why the choice of surrogate matters much more here than in standard vector quantization.

Leech spherical quantization. The encoder maps each bar to a latent $z \in \mathbb{R}^{24}$. We normalize it, $\hat{z} = z/\|z\|$, and replace it by the nearest codeword of the fixed codebook $\mathcal{C} \subset \mathbb{S}^{23}$ formed by the $N = 196{,}560$ minimal vectors of the Leech lattice Λ_{24}, rescaled to unit norm (**??**):

$$\hat{q} = \arg\max_{c \in \mathcal{C}} \langle \hat{z}, c\rangle, \qquad \theta = \angle(\hat{z}, \hat{q}) = \arccos\langle \hat{z}, \hat{q}\rangle. \tag{3}$$

Each bar costs $\log_2 N \approx 17.58$ bits, and the decoder sees only $\hat{q}$. Nearest codewords are 60° apart. We call θ the *quantization angle*. The codebook is not learned and we use no commitment loss ($\beta = 0$), so the encoder receives gradients only through the surrogate Jacobian $\partial q/\partial \hat{z}$ of the

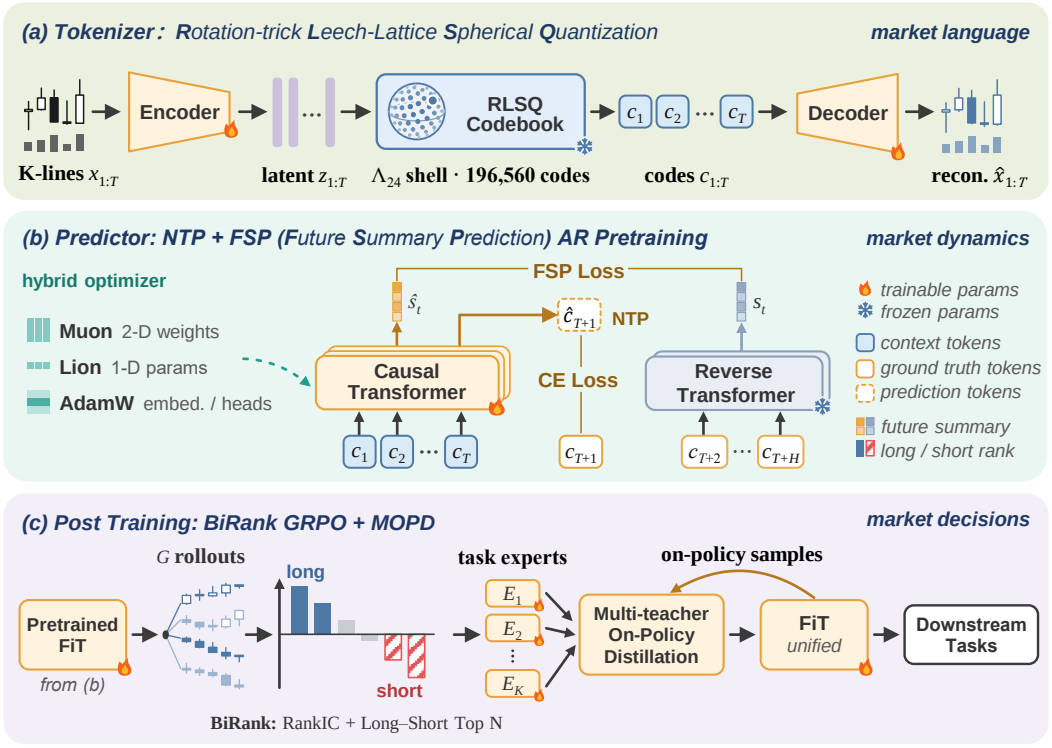


Figure 1: **Architecture of FiT.** (a) The RLSQ tokenizer encodes K-lines into latents $z_{1:T}$, projects them onto $\mathbb{S}^{23}$ and quantizes them to the nearest of the 196,560 codewords of the Leech-lattice shell Λ_{24}, with rotation-trick gradients; the decoder reconstructs $\hat{x}_{1:T}$. (b) A causal Transformer is pretrained on the codes with next-token prediction (NTP) and future summary prediction (FSP), matching $\hat{s}_t$ to the summary s_t of $c_{t+2:T}$ from a frozen reverse transformer; Muon updates 2-D weights, Lion 1-D parameters and AdamW the embeddings and head. (c) BiRank GRPO rewards the ranking of G sampled rollouts per asset (RankIC with long-short rank) to train task experts, which multi-teacher on-policy distillation (MOPD) merges back into one model.

quantizer. This surrogate is the only channel through which training can tell the encoder where the codewords are.

Two gradient surrogates. Let $g = \partial \ell / \partial q \,|_{q=\hat{q}}$ be the loss gradient at the decoder input of one token, and $P_x = I - xx^\top$. The straight-through estimator (STE; **??**) sets $\partial q / \partial \hat{z} := I$, so $\nabla_z^{\mathrm{STE}} = \|z\|^{-1} P_{\hat{z}}\, g$; the factor $P_{\hat{z}}/\|z\|$ is the Jacobian of the normalization. The rotation trick (Fifty et al., 2025) treats the codeword as a rotated copy of the input, $q = R\hat{z}$, with

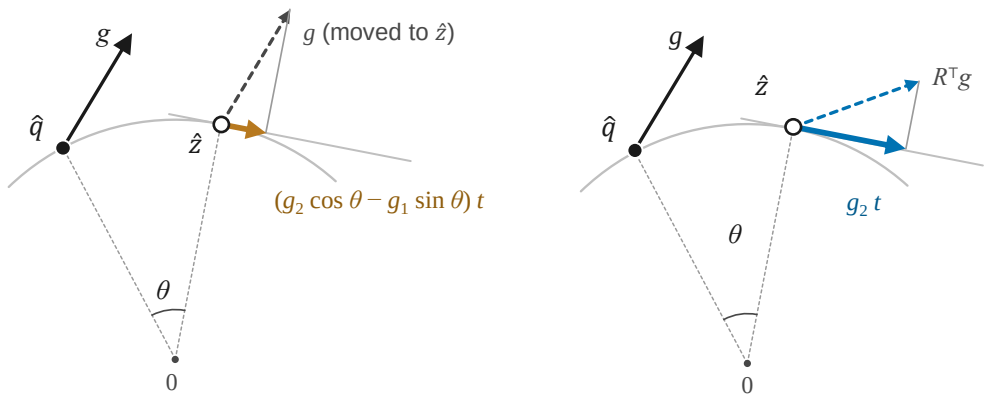
$$R = \big(I - 2hh^\top\big)\big(I - 2\hat{z}\hat{z}^\top\big), \qquad h = \frac{\hat{z} + \hat{q}}{\|\hat{z} + \hat{q}\|}, \tag{4}$$

the rotation in the plane $\Pi = \mathrm{span}\{\hat{z}, \hat{q}\}$ that maps $\hat{z}$ to $\hat{q}$ and fixes $\Pi^\perp$. Holding R constant in the backward pass gives $\nabla_z^{\mathrm{RLSQ}} = \|z\|^{-1} P_{\hat{z}} R^\top g$. We implement it as $q = \mathrm{sg}[\hat{z} + (\hat{q} - \hat{z})] + R\hat{z} - \mathrm{sg}[R\hat{z}]$, where sg is stop-gradient. The forward pass is then bit-identical to STE, and only the backward pass changes, at the cost of four inner products per token (Appendix B.4). We call the Leech codebook trained this way *RLSQ*.

Where the surrogates differ. For $\theta \in (0, \pi)$ let $\hat{w} = (\hat{z} - \cos\theta\, \hat{q})/\sin\theta$ be the unit tangent at $\hat{q}$ that points toward $\hat{z}$, and $t = (\cos\theta\, \hat{z} - \hat{q})/\sin\theta$ the unit tangent at $\hat{z}$ that points away from $\hat{q}$ (Figure 2a). Decompose $g = g_1 \hat{q} + g_2 \hat{w} + g_\perp$ with $g_\perp \perp \Pi$. Here g_1 would change the norm of the decoder input, g_2 moves it along the geodesic toward $\hat{z}$, and $g_\perp$ collects the other 22 directions.

Theorem 1 (Exact decomposition)**.** *For every upstream gradient g,*

$$\|z\|\, \nabla_z^{\mathrm{STE}} = (g_2 \cos\theta - g_1 \sin\theta)\, t + g_\perp, \qquad \|z\|\, \nabla_z^{\mathrm{RLSQ}} = g_2\, t + g_\perp,$$



(a) **STE.** (b) **Rotation Trick.**

Figure 2: The two gradient surrogates in the plane $\Pi = \mathrm{span}\{\hat{z}, \hat{q}\}$, drawn at the typical angle $\theta = 39°$. g is the loss gradient at the decoder input, i.e. at the codeword $\hat{q}$. (a) STE moves g to $\hat{z}$ unchanged and removes its normal component. This mixes the radial part g_1 into the tangent direction t and damps the displacement part g_2 by $\cos\theta$. (b) The rotation trick rotates g by θ, i.e. transports it along the geodesic. Its normal part is exactly $g_1\hat{z}$ and its tangent part is $g_2 t$. The two results differ only along t (Theorem 1).

so $\|z\|\,(\nabla_z^{\mathrm{RLSQ}} - \nabla_z^{\mathrm{STE}}) = \Delta\, t$ *with* $\Delta = g_1 \sin\theta + g_2(1 - \cos\theta)$.

The rotated gradient is the parallel transport, along the geodesic from $\hat{q}$ to $\hat{z}$, of the Riemannian gradient of the loss on the sphere. To first order, the estimator can leave a fingerprint only in the distribution of θ, which we measure in Experiment Section.

3.2 Predictor: Next-Token Prediction with Future Summary Prediction

Future summaries from a reverse-time teacher. NTP supervises h_t only through the next token, which in financial series is dominated by noise. FSP (Mahajan et al., 2026a) adds a target that summarizes the remainder of the window. We obtain it from a teacher g_ϕ, an autoregressive Transformer trained with the NTP objective on time-reversed token sequences $(k_T, k_{T-1}, \ldots, k_1)$. Windows are tokenized first and reversed afterwards, so the teacher and the student operate on the same code stream. Let r_j be the final-layer hidden state of g_ϕ after it has consumed $k_T, \ldots, k_j$. For the student position t, which has observed $k_{\le t}$ and is supervised by NTP on k_{t+1}, the future summary is

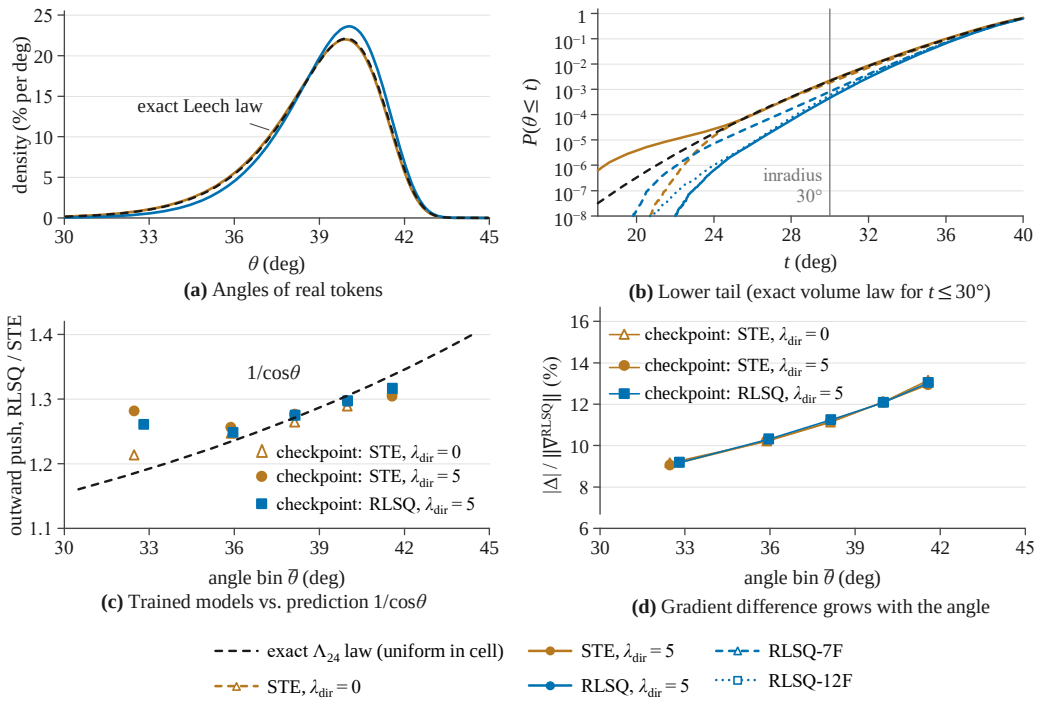
$$s_t \;=\; r_{t+2}, \qquad 1 \le t \le T-2, \tag{5}$$

a learned representation of $k_{t+2:T}$, that is, of everything beyond the NTP target. The teacher is trained once, on the same tokenizer, frequencies, and data sources as the student, and is frozen thereafter.

Auxiliary head. Let $h_t^{(L-1)}$ be the output of the penultimate block of the L-block student. An auxiliary branch consisting of one Transformer block A, a normalization layer, and a linear projection P to the teacher width predicts the summary, $\hat{s}_t = P\,\mathrm{Norm}\big(A(h^{(L-1)})_t\big)$, and is trained with

$$\mathcal{L}_{\mathrm{FSP}} = \frac{1}{T-2}\sum_{t=1}^{T-2} \mathrm{SmoothL1}\big(\hat{s}_t, \mathrm{sg}[s_t]\big), \qquad \mathcal{L}_{\mathrm{pre}} = \mathcal{L}_{\mathrm{NTP}} + \lambda_{\mathrm{FSP}}\,\mathcal{L}_{\mathrm{FSP}}. \tag{6}$$

The NTP path passes through the final block as usual, so the main model is architecturally unchanged, and the auxiliary branch is discarded after pretraining: FSP adds no parameters and no computation at inference. The auxiliary branch uses no dropout. It therefore draws no random numbers, and a model trained with $\lambda_{\mathrm{FSP}} = 0$ under the same seed receives identical initial weights and identical dropout masks, which makes the comparison with the NTP-only baseline a paired one.



(a) Angles of real tokens **(b)** Lower tail (exact volume law for $t \leq 30°$)

(c) Trained models vs. prediction $1/\cos\theta$ **(d)** Gradient difference grows with the angle

Figure 3: Quantization angles of real tokens (first training epoch, about 4×10^8 tokens per model) against the exact Leech law, which depends only on the codebook. (a) Density for the paired tokenizers at $\lambda_{\text{dir}} = 5$. STE is indistinguishable from the law; RLSQ has less mass at small angles and a higher peak. (b) Lower tail on a log scale. Below the 30° inradius the law is the exact volume law $N\mu(t)$. Both STE tokenizers follow it; all three RLSQ tokenizers lie below it. (c) Radial weight $w(s) = p_{\text{meas}}(s)/p_{\Lambda_{24}}(s)$ (markers: 0.5° bins with at least 50 tokens) and, for the RLSQ models, a two-parameter tilt fitted to the mean and variance only (lines). (d) One-parameter tilt κ_1 ($w \propto e^{\kappa_1 s}$) of the paired tokenizers at every saved checkpoint, on a fixed input of 2.1M tokens.

3.3 Post-Training: BiRank GRPO and Multi-Teacher On-Policy Distillation

A task specifies a sampling frequency, a horizon H, and a decision objective. One training example is a complete cross-section: at an anchor time, the contexts $\{c_i\}_{i=1}^N$ of N instruments. For each instrument the policy π_θ, initialized from the pretrained predictor, samples G rollouts of H tokens, which are decoded into predicted returns $\hat{y}_i^g$.

BiRank reward. The reward combines two order-based functionals. The first is the rank information coefficient, $\Phi_{\text{IC}} = \text{Spearman}(\hat{y}, y)$, which measures ordering quality over the entire cross-section. The second is a long–short spread, which measures what a portfolio formed from the two ends of the ranking would earn. A hard top-k spread changes only when a rollout crosses the k-th rank and therefore gives an almost everywhere constant reward; we relax it as

$$\Phi_{\text{LS}} = \frac{1}{k}\sum_{i=1}^{N}\left[w^+(p_i) - w^-(p_i)\right]\tilde{y}_i, \quad w^+(p) = \sigma\Big(\tfrac{k+\frac{1}{2}-p}{\tau}\Big), \quad w^-(p) = \sigma\Big(\tfrac{p-(N-k+\frac{1}{2})}{\tau}\Big), \tag{7}$$

where p_i is the rank of $\hat{y}_i$ in descending order, $\tilde{y}$ are cross-sectionally winsorized returns, and τ is a temperature in rank units. The soft weights depend only on ranks, so the prefix-sum evaluation above still applies. Both functionals are invariant to monotone transformations of the scores, which excludes the degenerate solution of shrinking all predictions toward zero. Their raw scales differ by orders of magnitude, so each is standardized within the group before they are combined:

$$A_i^g = \alpha\, \text{z}_g\big(R_i^{\text{IC},g}\big) + \beta\, \text{z}_g\big(R_i^{\text{LS},g}\big), \tag{8}$$

where z_g subtracts the mean and divides by the standard deviation over the G rollouts of instrument i.

Table 3: **Main results.** Eight downstream task families evaluated with the same four pretrained FiT checkpoints (no per-task retraining unless stated) against the strongest published or non-learned baselines we could run under an identical protocol. Every number is out-of-sample: test spans start after the pretraining cutoff (2022-06-30) and, for the trading and execution tasks, after 2024-01-01. Arrows give the direction of improvement; **bold** is the best entry in a row. Baselines are ordered by their score on that row, so column *Best baseline* is the number FiT has to beat.

Task	Metric (setting)	FiT (ours)				Best baseline		2nd baseline	
		S 258M	B 680M	L 1.41B	U 3.25B	model	score	model	score
(1) Time-series forecasting	IC ↑ (composite)	*0.0718*	0.0769	0.0808	**0.0831**	Kronos-base	0.0587	Chronos-Bolt	0.0412
(2) Cross-sectional prediction	RankIC ↑ (composite)	0.0742	0.0815	0.0879	**0.0916**	Kronos-base	0.0638	reversal	0.0517
(3) Volatility	QLIKE ↓ (5 min RV)	0.372	0.351	0.338	**0.334**	HAR-RV	0.394	GARCH(1,1)	0.437
	$R^2_{\log}$ ↑ (1 d RV)	0.641	0.662	0.673	**0.677**	HAR-RV	0.612	GARCH(1,1)	0.548
	|cov.err| ↓ (VaR 99%, pp)	0.94	0.71	0.58	**0.54**	Hist. sim.	1.12	GARCH-t	1.43
(4) K-line generation	Disc. score$\to$0.5 [a] (1 d)	0.285	0.311	0.324	**0.329**	TimeGAN	0.123	Kronos-base	0.093
	Disc. score$\to$0.5 [a] (15 min)	0.231	0.252	0.263	**0.268**	Kronos-small	0.163	Kronos-base	0.140
	Validity ↑(%) [b] (15 min)	68.4	74.2	77.9	**79.1**	Kronos-base	30.5	Kronos-small	28.0
	TSTR IC ↑ [c] (15 min)	0.0640	0.0661	0.0678	0.0684	TimeVAE	**0.0690**	Kronos-base	0.0620
(5) Order execution	PA vs mkt. ↑(bp) (crossing fill)	1.22	1.31	1.38	**1.41**	Passive limit	0.94	TWAP	0.00
	PA vs mkt. ↑(bp) (queue fill)	2.43	2.56	2.64	2.69	Passive limit	**2.75**	TWAP	0.00
(6) Stock simulated trading	Ann. ret. ↑(%) (CSI2000 top-50)	**44.9**	25.8	34.3	38.7	Kronos-small	44.1	Kronos-base	43.0
	Sharpe ↑ (CSI2000 top-50)	**1.21**	0.72	0.96	1.08	Kronos-small	1.18	Kronos-base	1.12
(7) CTA hedge trading	Ann. ret. ↑(%)	15.7	17.6	18.9	**19.4**	TSMOM	12.4	XSMOM	9.1
	Sharpe ↑	0.88	0.98	1.05	**1.08**	TSMOM	0.71	XSMOM	0.55
(8) Arbitrage	RankIC ↑ (30 min)	0.0602	0.0673	0.0714	**0.0729**	Basis z-score	0.0431	OU half-life	0.0355
	RankIC ↑ (1 d)	0.0207	0.0225	0.0238	0.0244	Basis z-score	**0.0262**	OU half-life	0.0184

Multi-teacher on-policy distillation. Maintaining K experts is impractical, and each expert has drifted from the pretrained distribution. We consolidate them into one student π_ψ, initialized from the pretrained predictor, by on-policy distillation with multiple teachers (Agarwal et al., 2024; Xiaomi LLM-Core Team, 2026). For a context c from task k, the student samples its own rollout $a_{1:H} \sim \pi_\psi(\cdot \mid c)$, and the teacher of that task supervises every position:

$$\mathcal{L}_{\mathrm{MOPD}}(\psi) = \sum_{k=1}^{K} w_k \, \mathbb{E}_{c\sim\mathcal{D}_k,\, a\sim\pi_\psi(\cdot|c)} \Big[\frac{1}{H} \sum_{j=1}^{H} \mathrm{KL}\big(\pi_\psi(\cdot \mid c, a_{<j}) \,\|\, \pi_{E_k}(\cdot \mid c, a_{<j})\big) \Big]. \quad (9)$$

Exact full-vocabulary divergence gives a dense, low-variance per-position signal; on-policy supervision avoids train–inference mismatch, and per-task teachers let the student inherit expert competence while staying near the pretrained model.

4 Experiments

Our experiments follow the structure of the model and answer three questions. **RQ1**: does RLSQ represent K-lines more faithfully? **RQ2**: does FiT forecast better than existing foundation models in the zero-shot setting, and what does FSP contribute beyond NTP (§**??**)? **RQ3**: does BiRank GRPO improve decision-oriented metrics, and can MOPD consolidate the experts without eroding the pretrained model (§**??**)?

4.1 FiT Dataset and Data Protocol

The FiT Dataset comprises more than 200 billion K-line (candlestick) records collected from over 100 exchanges, with base resolutions as fine as one second. Since the corpus is stored at such fine-grained base resolutions, bars of arbitrary length can be constructed through exact aggregation. We exploit this property by training on a grid of sampling frequencies that encompasses not only conventional intervals but also non-standard bar lengths (e.g., 7, 13, 23, 45, and 90 minutes), thereby discouraging the model from relying on regularities specific to a small set of frequencies.

Training windows are drawn through hierarchical sampling: a data source is selected first, followed by a frequency f with probability p_f, then an instrument, and finally a window within that instrument. Sampling at the source level first keeps the mixture proportions across markets fixed, irrespective of the number of files each market contributes.

4.2 Main Results.

Setup. The tokenizer's encoder and decoder are 6-layer Transformers (width 256) over bars with six channels (OHLC, volume, turnover), with one 24-dimensional latent per bar. It is trained for 100 epochs on Chinese A-share K-lines at seven frequencies from 1 minute to 1 day, split in time into training, validation and test periods. The objective is a log-cosh reconstruction loss plus an OHLC-consistency penalty and a direction-aware hinge loss with weight λ_{dir} (Appendix B.5). Each STE/RLSQ *pair* shares the seed, initialization, data order and every hyperparameter. With the rotation switched off, the RLSQ script reproduces the STE run bit for bit, so the two runs differ only in the backward pass of the quantizer. We run pairs at $\lambda_{\text{dir}} = 5$ and 100 and select checkpoints by validation MSE.

RLSQ trains a better tokenizer at no cost. The rotation lowers validation MSE by 8.1% at $\lambda_{\text{dir}} = 5$ (test -4.9%) and by 6.9% at $\lambda_{\text{dir}} = 100$ (test -6.4%; Table **??**). RLSQ is ahead in 83 and 78 of the 100 paired epochs, and in all of the last ten in both pairs. All six channels improve at both operating points (Appendix B.6), and direction accuracy and code usage rise as well. The direction loss at $\lambda_{\text{dir}} = 5$ costs the STE tokenizer 8.2% in reconstruction relative to $\lambda_{\text{dir}} = 0$. With the rotation the same loss is free: RLSQ is 0.6% better than the STE tokenizer trained without it, and its code entropy (0.769) matches that tokenizer's. Step time is unchanged. Each arm is a single seed; the agreement between the two operating points and over the last ten epochs is our evidence against seed noise.

Main result. Table 3 reports zero-shot forecasting performance. FiT improves IC by 52% and RankIC by 95% over the best-performing baseline. The per-frequency breakdown (Appendix –) shows that the predictions of FiT are statistically significant at every frequency under three independent tests: a bootstrap confidence interval, a parametric t-test on the per-anchor IC series, and a within-cross-section permutation test.

4.3 Generalization across Frequencies and Markets

The FiT Dataset was built to broaden coverage along two axes, and we test both. *Sampling frequency.* We evaluate on bar lengths that never occur in training, obtained by aggregating the base-resolution data, and compare with a model trained on the conventional frequency grid only. Table – reports reconstruction error, directional accuracy, and RankIC on seen and unseen frequencies. *Markets.* We hold out – exchanges entirely during training and evaluate zero-shot on them. Table – compares FiT with a model trained on the Kronos-scale subset of the data. Finally, Figure – shows validation loss and zero-shot RankIC as a function of the amount of training data.

5 Conclusion and Discussions

We presented FiT, a foundation model for financial K-line data that treats representation, prediction, and adaptation as one pipeline instead of three separate problems. Each stage addresses a limitation of the stage before it.

Beyond the individual results, two observations from this work apply more broadly. The first is that, for discrete models of continuous data, the tokenizer sets the ceiling: information removed by quantization, such as the sign of a small price move, cannot be restored by a better decoder or a larger predictor, so it has to be protected when the code is learned. The second is that likelihood and decision quality are different objectives. A lower cross-entropy did not reliably yield a better cross-sectional ranking in our experiments, and part of what reinforcement learning appears to gain is a sharper predictive distribution; we therefore evaluate against parameter-free references and entropy-matched controls, and we recommend this practice for future work on financial foundation models.

USE OF LARGE LANGUAGE MODELS

In this work, we did not use generative AI tools to propose hypotheses, develop the theory, design the experiments, implement the training and evaluation code, conduct the analysis, or generate or interpret the results. We have not used them for any other substantive research tasks, and the remaining tasks with required disclosure are not applicable to this work. Additionally, we did not use generative AI tools for literature search, writing, or figure generation. All research decisions, experimental procedures, analyses, and conclusions were conducted and verified by the authors. We take full responsibility for the final content of this work, including the text, claims, results, and all other research artifacts.

ETHICS STATEMENT

This work studies forecasting models trained solely on market-level price and volume time series (candlesticks) from public exchanges. It involves no human subjects and no personally identifiable or account-level trading data, so no IRB review was required. The FiT Dataset is compiled from third parties for academic use only; we use the data in accordance with each provider's terms of use, and redistribution is subject to their licenses. Financial forecasting models carry dual-use risks: if widely deployed, correlated model-driven decisions could amplify herding behavior, and such models could be misused in attempts at market manipulation. Our models are released for research purposes only; their outputs do not constitute investment advice, and any real-world deployment should include risk controls and human oversight. The authors declare no competing interests.

REPRODUCIBILITY STATEMENT

We have made the following efforts to ensure reproducibility.

(i) **Tokenizer.** RLSQ is fully specified in §3. The Λ_{24} codebook of 196,560 minimal vectors is constructed deterministically from the standard extended Golay code construction (Appendix **??**), so no learned codebook needs to be released.

(ii) **Training.** Architectures, the NTP+FSP objectives, the BiRank GRPO and MOPD procedures, and the hybrid optimizer configuration (Muon/Lion/AdamW parameter groups, learning rates, schedules) are listed in the code repo.

(iii) **Data.** The construction and preprocessing pipeline of the FiT Dataset (exchange coverage, 1-second aggregation, cleaning, normalization, and temporal train/val/test splits) is described in this paper. [Owing to licensing restrictions on raw market data, we cannot redistribute the full corpus; we release .]

(iv) **Evaluation.** Metric definitions and protocols are given in Appendix **??**; results are averaged over [N] random seeds.

(v) **Code.** Anonymized training and evaluation code is provided in the supplementary material [/ at `https://anonymous.4open.science/r/FiT-FiT/`].

REFERENCES

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2306.13649.

Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *International Conference on Machine Learning (ICML)*, 2020.

Xiangning Chen, Chen Liang, Da Huang, Esteban Real, Kaiyuan Wang, Yao Liu, Hieu Pham, Xuanyi Dong, Thang Luong, Cho-Jui Hsieh, Yifeng Lu, and Quoc V. Le. Symbolic discovery of optimization algorithms. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Leshem Choshen, Lior Fox, Zohar Aizenbud, and Omri Abend. On the weaknesses of reinforcement learning for neural machine translation. In *International Conference on Learning Representations (ICLR)*, 2020.

Henry Cohn, Abhinav Kumar, Stephen D. Miller, Danylo Radchenko, and Maryna Viazovska. The sphere packing problem in dimension 24. *Annals of Mathematics*, 185(3):1017–1033, 2017. doi: 10.4007/annals.2017.185.3.8.

Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001. doi: 10.1080/713665670.

John H. Conway and Neil J. A. Sloane. Voronoi regions of lattices, second moments of polytopes, and quantization. *IEEE Transactions on Information Theory*, 28(2):211–226, 1982.

John H. Conway and Neil J. A. Sloane. *Sphere Packings, Lattices and Groups*, volume 290 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, 3rd edition, 1999. doi: 10.1007/978-1-4757-6568-7.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *International Conference on Machine Learning (ICML)*, 2024.

Yonathan Efroni, Ben Kretzu, Daniel R. Jiang, Jalaj Bhandari, Zheqing Zhu, and Karen Ullrich. Aligned multi-objective optimization. Preprint, 2025.

Christopher Fifty, Ronald G. Junkins, Dennis Duan, Aniketh Iyengar, Jerry W. Liu, Ehsan Amid, Sebastian Thrun, and Christopher Ré. Restructuring vector quantization with the rotation trick. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2410.06424.

Huan-ang Gao, Haohan Chi, Yong Yan, Shiyuan Feng, Hanlin Wu, Zheng Jiang, Bingxiang He, Wei-Ying Ma, Ya-Qin Zhang, and Hao Zhou. Open-MOPD: Diagnosing and fixing capability imbalance in multi-teacher on-policy distillation. *arXiv preprint arXiv:2608.19098*, 2026.

Allen Gersho. Asymptotically optimal block quantization. *IEEE Transactions on Information Theory*, 25(4):373–380, 1979.

Fabian Gloeckle, Badr Youbi Idrissi, Baptiste Rozière, David Lopez-Paz, and Gabriel Synnaeve. Better & faster large language models via multi-token prediction. In *International Conference on Machine Learning (ICML)*, 2024.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *International Conference on Machine Learning (ICML)*, 2024.

Shengding Hu, Yuge Tu, Xu Han, Chaoqun He, Ganqu Cui, Xiang Long, Zhi Zheng, Yewei Fang, Yuxiang Huang, Weilin Zhao, et al. MiniCPM: Unveiling the potential of small language models with scalable training strategies. *arXiv preprint arXiv:2404.06395*, 2024.

Yifan Hu, Jie Yang, Tian Zhou, Peiyuan Liu, Yujin Tang, Rong Jin, and Liang Sun. Bridging past and future: Distribution-aware alignment for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2026.

Keller Jordan, Yuchen Jin, Vlado Boza, Jiacheng You, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An optimizer for hidden layers in neural networks. `https://kellerjordan.github.io/posts/muon/`, 2024.

Jean Kaddour. Stop wasting my time! saving days of ImageNet and BERT training with latest weight averaging. *arXiv preprint arXiv:2209.14981*, 2022.

Daesoo Lee, Sara Malacarne, and Erlend Aune. Vector quantized time series generation with a bidirectional prior model. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

Vladimir I. Levenshtein. On bounds for packings in n-dimensional Euclidean space. *Soviet Mathematics Doklady*, 20:417–421, 1979.

Siyang Li, Yize Chen, Zijie Zhu, et al. TimeRFT: Stimulating generalizable time series forecasting for TSFMs via reinforcement finetuning. *arXiv preprint arXiv:2605.00015*, 2026.

Yuhui Li, Fangyun Wei, Chao Zhang, and Hongyang Zhang. EAGLE-3: Scaling up inference acceleration of large language models via training- time test. *arXiv preprint arXiv:2503.01840*, 2025.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. arXiv:1711.05101.

Divyat Mahajan, Sachin Goyal, Badr Youbi Idrissi, Mohammad Pezeshki, Ioannis Mitliagkas, David Lopez-Paz, and Kartik Ahuja. Beyond multi-token prediction: Pretraining LLMs with future summaries. In *International Conference on Learning Representations (ICLR)*, 2026a. arXiv:2510.14751.

Divyat Mahajan, Sachin Goyal, Badr Youbi Idrissi, Mohammad Pezeshki, Ioannis Mitliagkas, David Lopez-Paz, and Kartik Ahuja. Beyond multi-token prediction: Pretraining LLMs with future summaries. In *International Conference on Learning Representations (ICLR)*, 2026b.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations (ICLR)*, 2023.

Andrew M. Odlyzko and Neil J. A. Sloane. New bounds on the number of unit spheres that can touch a unit sphere in n dimensions. *Journal of Combinatorial Theory, Series A*, 26(2):210–214, 1979. doi: 10.1016/0097-3165(79)90074-8.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Hena Ghonia, Rishika Bhagwatkar, Arian Khorasani, Mohammad Javad Darvishi Bayazi, George Adamopoulos, Roland Riachi, Nadhir Hassen, et al. Lag-llama: Towards foundation models for probabilistic time series forecasting. *arXiv preprint arXiv:2310.08278*, 2023.

Steven J. Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. Self-critical sequence training for image captioning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Shiqi Shen, Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. Minimum risk training for neural machine translation. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016.

Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. Time-MoE: Billion-scale time series foundation models with mixture of experts. In *International Conference on Learning Representations (ICLR)*, 2025.

Yu Shi, Zongliang Fu, Shuo Chen, Bohan Zhao, Wei Xu, Changshui Zhang, and Jian Li. Kronos: A foundation model for the language of financial markets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25366–25373, 2026.

Adam Stooke, Joshua Achiam, and Pieter Abbeel. Responsive safety in reinforcement learning by PID lagrangian methods. In *International Conference on Machine Learning (ICML)*, 2020.

Sabera Talukder, Yisong Yue, and Georgia Gkioxari. TOTEM: TOkenized time series EMbeddings for general time series analysis. *Transactions on Machine Learning Research*, 2024.

Jayden Teoh, Manan Tomar, Kwangjun Ahn, Pratyusha Sharma, Akshay Krishnamurthy, and John Langford. Next-latent prediction transformers learn compact world models. *arXiv preprint arXiv:2511.05963*, 2026.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017. arXiv:1711.00937.

Shu Wan, Miles Ma, Hank Zhu, Guangqi Liu, Stephen Wang, Qingsong Wen, and Huan Liu. Forecast collapse in time-series foundation models. *arXiv preprint arXiv:2608.14106*, 2026.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *International Conference on Machine Learning (ICML)*, 2024.

Jialong Wu, Shaofeng Yin, Ningya Feng, and Mingsheng Long. RLVR-world: Training world models with reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Xiaomi LLM-Core Team. MiMo-V2-Flash technical report. *arXiv preprint arXiv:2601.02780*, 2026.

Jie Yang, Yifan Hu, Yuante Li, Kexin Zhang, Kaize Ding, and Philip S. Yu. From observations to states: Latent time series forecasting. *arXiv preprint arXiv:2602.00297*, 2026.

Lijun Yu, José Lezama, Nitesh B. Gundavarapu, Luca Versari, Kihyuk Sohn, David Minnen, Yong Cheng, Agrim Gupta, Xiuye Gu, Alexander G. Hauptmann, Boqing Gong, Ming-Hsuan Yang, Irfan Essa, David A. Ross, and Lu Jiang. Language model beats diffusion – tokenizer is key to visual generation. In *International Conference on Learning Representations (ICLR)*, 2024.

Shuo Yu, Hongyan Xue, Xiang Ao, Feiyang Pan, Jia He, Dandan Tu, and Qing He. Generating synergistic formulaic alpha collections via reinforcement learning. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2023.

Yue Zhao, Yuanjun Xiong, and Philipp Krähenbühl. Image and video tokenization with binary spherical quantization. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2406.07548.

Yue Zhao, Hanwen Jiang, Zhenlin Xu, Chutong Yang, Ehsan Adeli, and Philipp Krähenbühl. Spherical leech quantization for visual tokenization and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12913–12923, 2026.

Yaochen Zhu, Harald Steck, Dawen Liang, Yinhan He, Vito Ostuni, Jundong Li, and Nathan Kallus. Rank-GRPO: Training LLM-based conversational recommender systems with reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2026.

A Related work

A.1 Discrete representations and the geometry of codebooks

Discrete bottlenecks for continuous signals start with the VQ-VAE, whose codebook is learned jointly with the autoencoder (**?**). Learning the codebook brings its own failure mode — most entries stop being used — and a literature of remedies: commitment and entropy terms, restarts, and exponential moving-average updates. A second line removes the learned codebook altogether. Finite scalar quantization rounds each latent coordinate to one of a few fixed levels (**?**), lookup-free quantization takes the levels to be binary (Yu et al., 2024), and binary spherical quantization does the same on the unit sphere (Zhao et al., 2025). These schemes cannot collapse and need no codebook parameters, but they are *product* codes: the alphabet is a Cartesian grid, two codewords can differ in a single coordinate, and the angular separation of neighbours shrinks as the rate grows. Classical quantization theory points the other way — for a fixed rate, good quantizers are lattices, not products of scalar quantizers (Gersho, 1979; Conway & Sloane, 1982), and in 24 dimensions the extremal object is the Leech lattice (**?**). **?** put its minimal shell to work as an image tokenizer. We use the same shell for financial bars, and Section 3.1 adds what a borrowed code needs in order to be trainable at this rate: an account of the quantization angle it induces, and the gradient surrogate that account selects.

Training an encoder through a piecewise-constant map requires a surrogate Jacobian. The straight-through estimator (**?**) is the default; its optimization behaviour has been studied both theoretically (**?**) and empirically, with the alternating-optimization view of **?** explaining several of the pathologies of vector-quantized networks. The rotation trick (Fifty et al., 2025) replaces the identity backward pass by the rotation that carries the latent onto its codeword, and reports gains on image tokenizers with learned codebooks. In that setting a commitment term keeps latents close to their codewords, so the two surrogates nearly agree; the regime that distinguishes them — a codebook that cannot move and a rate high enough that every latent sits far from its codeword — is exactly the regime a fixed spherical code creates. Section 3.1 makes the dependence on the angle explicit and reports the resulting paired comparison.

A.2 Foundation models for time series and for markets

One family keeps the signal continuous: patches are embedded and a Transformer is trained with a regression, quantile, or mixture head (Nie et al., 2023; Das et al., 2024; Woo et al., 2024; Goswami et al., 2024; Rasul et al., 2023; Shi et al., 2025). The other discretizes first and then trains a language model. Chronos bins scaled values into a fixed vocabulary and trains with cross-entropy (Ansari et al., 2024); TOTEM and TimeVQVAE learn a VQ codebook over patches or over a time–frequency representation (Talukder et al., 2024; Lee et al., 2023). The discrete route buys a categorical predictive distribution that is free to be multimodal and sampling that produces self-consistent paths rather than marginals; it pays for this with an alphabet that must be designed. Our model is in the discrete family and differs in what is quantized and by what: the atom is a six-channel bar whose coordinates are bound by the structural constraints of equation 1, so quantization error cannot be judged channel by channel, and the alphabet is a fixed 24-dimensional lattice code rather than a learned or product one (Section 3.1).

Kronos (**?**) is the closest predecessor: a two-stage model for candlestick data, with a BSQ tokenizer at 20 bits per bar and a predictor that factorizes each token into a coarse and a fine sub-token because a flat softmax over 2^{20} classes is impractical. Our work keeps that two-stage recipe and changes the alphabet, which removes the factorization (Section **??**), and enlarges the corpus along the axes a K-line model can actually exploit — markets, instruments, and sampling frequencies from one second upward (Section 4.1). A second, orthogonal line wraps a language model around market text and numbers; its gains have been traced largely to the language side, so they do not transfer to a model whose only input is price and volume, and we do not compare against it. Wan et al. (2026) report a failure mode that any such model has to be measured against: across many time-series foundation models, forecast dispersion shrinks with the horizon and the cross-sectional information the forecasts carry degrades with it.

A.3 Objectives beyond next-token prediction

Next-token prediction with teacher forcing supervises only one step ahead. Multi-token prediction adds parallel heads for the next few tokens and improves generation while remaining short-range (Gloeckle et al., 2024); future summary prediction instead trains an auxiliary head to predict a compact representation of the *long-term* future, either handcrafted or produced by a reverse language model trained right to left (Mahajan et al., 2026b). Related ideas predict latents rather than tokens (Teoh et al., 2026; Yang et al., 2026), and speculative decoders show that what the auxiliary head reads matters more than how large it is (Li et al., 2025). For time series, Hu et al. (2026) align the distribution of the predicted future with that of the past, in continuous space. Section **??** ports the reverse-teacher variant of future summary prediction to a discrete K-line predictor, where the teacher has to be trained in the student's own code space for its summaries to be readable, and reports the dose–response checks that a strong-teacher claim needs.

A likelihood objective on tokens is not the objective a forecast is judged by: the sign of the next move matters separately from its magnitude, and adding a directional term to cross-entropy creates a trade-off whose weight is a nuisance parameter. Rather than tune it, we treat the directional and structural terms as constraints and let multipliers find their own level, following the Lagrangian view of constrained learning and its control-theoretic reading (Stooke et al., 2020); Efroni et al. (2025) give conditions under which such objectives need not trade off at all. Section **??** reports the controller, the regime where the constraint is genuinely binding, and the price paid in cross-entropy when it is.

Our training recipe assigns different optimizers to different parameter groups: a spectral-condition method for hidden weight matrices (Jordan et al., 2024), a sign-based method for one-dimensional parameters (Chen et al., 2023), and AdamW for embeddings and the output head (Loshchilov & Hutter, 2019). We follow the warmup–stable–decay schedule and its scaling analysis (Hu et al., 2024; **?**), and average late checkpoints rather than relying on a single one (Kaddour, 2022). These are engineering choices, reported for reproducibility and ablated in Section **??**, not contributions.

A.4 Post-training with rewards and with teachers

Optimizing a sequence model directly against the metric it is scored by is an old idea — minimum risk training (Shen et al., 2016) and self-critical sequence training (Rennie et al., 2017) — whose modern form is group-relative policy optimization (Shao et al., 2024) on top of the clipped surrogate of PPO (Schulman et al., 2017). The same idea has reached forecasting: reward a world model with its own prediction metric (Wu et al., 2025), or fine-tune a time-series foundation model with a group-relative objective (Li et al., 2026). Two cautions from this literature shape our design. First, metric gains from RL fine-tuning can come from sharpening the output distribution rather than from better predictions (Choshen et al., 2020), so a temperature control belongs next to every such result. Second, in prediction tasks the within-group spread of rewards does not measure output quality the way it does in verifiable domains: with one realized future per context, a rollout can be rewarded for matching this particular draw of noise. Where the target is genuinely discrete and non-differentiable, RL remains the natural tool — searching over formulaic alpha expressions is the financial example (Yu et al., 2023) — whereas ranking objectives on continuous scores admit differentiable surrogates through soft sorting (Blondel et al., 2020), which is the cheaper baseline any RL arm has to beat.

The quantity we care about is cross-sectional: it is a functional of the forecasts for *all* instruments at one anchor, not of one rollout, and it is not additive over them. Rewards of this shape are handled either by leave-one-out marginal contributions or, when the functional is ordered, by decomposing it along the sequence, as Zhu et al. (2026) do by moving the action unit from the token to a rank in a list. Section 3.3 takes the leave-one-out route, with shuffled-reward and sharpening nulls and an explicit power budget, and reports where the resulting gains do and do not survive them. Because post-training specializes, we consolidate: several domain-specialized experts are distilled back into one base model with on-policy distillation (Agarwal et al., 2024) in its multi-teacher form (Gao et al., 2026), which supervises the student on the states it actually reaches and has to balance teachers so that no domain dominates.

B RLSQ: THEORY, PROOFS AND RESULTS

B.1 FURTHER RESULTS

Throughout, $\theta \in (0, \pi)$, $\Pi = \operatorname{span}\{\hat{z}, \hat{q}\}$, $\hat{w} = (\hat{z} - \cos\theta\,\hat{q})/\sin\theta$ and $t = (\cos\theta\,\hat{z} - \hat{q})/\sin\theta = -\sin\theta\,\hat{q} + \cos\theta\,\hat{w}$. Thus $(\hat{q}, \hat{w})$ and $(\hat{z}, t)$ are orthonormal bases of Π, $\langle \hat{q}, t\rangle = -\sin\theta$ and $\langle \hat{w}, t\rangle = \cos\theta$.

Theorem 2 (Geometry of the two surrogates). *(a) $\|z\|\nabla_z^{\mathrm{RLSQ}}$ is the parallel transport, along the geodesic from $\hat{q}$ to $\hat{z}$, of the Riemannian gradient $\operatorname{grad}\ell(\hat{q}) = P_{\hat{q}}\,g$ of the loss restricted to the sphere. In particular its norm $\|P_{\hat{q}}g\|$ does not depend on θ. (b) $\|z\|\nabla_z^{\mathrm{STE}} = P_{\hat{z}}\,g$ is a Euclidean projection. It depends on the normal component g_1, which is not part of $\operatorname{grad}\ell(\hat{q})$, and it shrinks the displacement component g_2 by $\cos\theta$. (c) If $\theta < 90°$, R is the unique orthogonal matrix Q with $Q\hat{z} = \hat{q}$ that minimizes $\|Q - I\|_F$, and the minimum is $2\sqrt{2}\sin(\theta/2)$. The STE Jacobian I maps $\hat{z}$ to $\hat{q}$ only if $\theta = 0$. (d) The local model implied by STE, $\hat{z}' \mapsto \hat{z}' + (\hat{q} - \hat{z})$, leaves the unit sphere at rate $2\sin\theta$ along t, whereas R maps the sphere to itself.*

Theorem 3 (STE is RLSQ plus an implicit commitment). *Let $c(\hat{z}) = 1 - \langle \hat{z}, \hat{q}\rangle = \frac{1}{2}\|\hat{z} - \hat{q}\|^2$. Its spherical gradient is $\sin\theta\,t$, and*

$$\|z\|\,\nabla_z^{\mathrm{STE}} = \|z\|\,\nabla_z^{\mathrm{RLSQ}} - g_1\,\operatorname{grad} c(\hat{z}) - g_2(1 - \cos\theta)\,t.$$

If the decoder's first layer is $u = Wq + b$, then $\sum_{\text{tokens}} g_1 = \langle W, \partial L/\partial W\rangle$.

Proposition 1 (No restoring force). *Fix the rest of the network, so that g is constant within a cell. The STE field $-P_{\hat{z}}g$ is the spherical gradient flow of the linear function $\langle g, \cdot\rangle$; its only stable point, $-g/\|g\|$, does not depend on $\hat{q}$. The RLSQ field is nonzero wherever $\operatorname{grad}\ell(\hat{q}) \neq 0$. Neither field has an attractor at $\hat{q}$.*

Supplement to Theorem ??. (a') For inputs whose density is at most K times the uniform one, $P(\theta \le t) \le KN\mu(t)$. (b') The covering radius is attained in the directions of the norm-6 vectors of Λ_{24}; each such direction is equidistant from 552 codewords. For every token, $\theta \le 52.24°$, $\sin\theta \le 0.791$ and $2\sin(\theta/2) \le 0.880$. (c') For any N-point code: 1% quantile $\ge 32.25°$, $10\% \ge 36.06°$, median $\ge 39.08°$, $\mathbb{E}\theta \ge 38.54°$, $\mathbb{E}[1 - \cos\theta] \ge 0.2182$. The Leech code has median $39.28°$ and mean $38.91°$. (e) *Scaling.* Solving $N\mu(\theta) = \frac{1}{2}$ for $N \times \{\frac{1}{4}, \frac{1}{2}, 1, 2, 4\}$ gives 41.95, 40.48, 39.08, 37.74 and $36.46°$, i.e. $\Delta\theta \approx \ln 2/(22\cot\theta) \approx 1.4°$ per doubling. A median of $20°$ needs $N \ge 2^{38.1} \approx 3 \times 10^{11}$.

B.2 PROOFS

Lemma (action of R). *R fixes $\Pi^\perp$ pointwise and acts on Π as the rotation by θ that takes $\hat{z}$ to $\hat{q}$. In particular $R\hat{z} = \hat{q}$ and $Rt = \hat{w}$, equivalently $R^\top\hat{q} = \hat{z}$ and $R^\top\hat{w} = t$. Proof.* The normals $\hat{z}$ and h of the two reflections lie in Π, so both reflections fix $\Pi^\perp$, and $\det R = 1$, so $R|_\Pi$ is a rotation. Since $\|\hat{z} + \hat{q}\|^2 = 2(1 + \cos\theta)$ and $\langle \hat{z}, h\rangle = (1 + \cos\theta)/\|\hat{z} + \hat{q}\|$, we have $2\langle \hat{z}, h\rangle h = \hat{z} + \hat{q}$ and hence $R\hat{z} = (I - 2hh^\top)(-\hat{z}) = -\hat{z} + (\hat{z} + \hat{q}) = \hat{q}$. A planar rotation is determined by the image of one nonzero vector. In the basis $(\hat{q}, \hat{w})$, $\hat{z} = (\cos\theta, \sin\theta)$ and $t = (-\sin\theta, \cos\theta)$, and R is the rotation by $-\theta$, so $Rt = (0, 1) = \hat{w}$. □

Proof of Theorem 1. $P_{\hat{z}}$ removes the $\hat{z}$-component. The in-plane part $g_1\hat{q} + g_2\hat{w}$ of g has t-component $-g_1\sin\theta + g_2\cos\theta$, and $g_\perp \perp \hat{z}$ is kept. This gives the STE identity. By the Lemma, $R^\top g = g_1\hat{z} + g_2 t + g_\perp$, and $P_{\hat{z}}$ removes $g_1\hat{z}$, which gives the RLSQ identity. Subtracting the two gives the difference. □

Proof of Corollary ??. (i) If $\theta > 0$, take $g = \hat{q}$: then $\Delta = \sin\theta \neq 0$. If $\theta = 0$ then $h = \hat{z}$ and $R = I$. (ii) Immediate from Theorem 1. (iii) $\Delta = \langle (g_1, g_2), (\sin\theta, 1 - \cos\theta)\rangle \le \|(\sin\theta, 1 - \cos\theta)\| = 2\sin(\theta/2)$, with equality iff $(g_1, g_2) \propto (\cos\frac{\theta}{2}, \sin\frac{\theta}{2})$ and $g_\perp = 0$, i.e. $g \parallel h$. The expansion follows from $\sin\theta = \theta + O(\theta^3)$ and $1 - \cos\theta = O(\theta^2)$. *Meaning of t.* If $\hat{z}$ moves to $\hat{z} + \epsilon v$ with $v \perp \hat{z}$ while staying in the cell of $\hat{q}$, then $d\theta = -\langle d\hat{z}, \hat{q}\rangle/\sin\theta$ and $\hat{q} = \cos\theta\,\hat{z} - \sin\theta\,t$, so $d\theta = \epsilon\langle v, t\rangle$. □

Proof of Theorem 2. (a) The geodesic from $\hat{q}$ to $\hat{z}$ is $\gamma(s) = \cos s\,\hat{q} + \sin s\,\hat{w}$, $s \in [0, \theta]$. Let R_s be the rotation by s in Π that fixes $\Pi^\perp$ and maps $\hat{q} \mapsto \gamma(s)$. For $v = \hat{w}$ the field $V(s) = R_s\hat{w} = \gamma'(s)$ has $V'(s) = -\gamma(s)$, which is normal to the sphere, so its covariant derivative vanishes. For $v \in \Pi^\perp$, $V(s) = v$ is constant and tangent. By linearity $R_s v$ is parallel for every $v \in T_{\hat{q}}\mathbb{S}^{23}$, so the parallel transport is $\Gamma_{\hat{q}\to\hat{z}} = R_\theta|_{T_{\hat{q}}} = R^\top|_{T_{\hat{q}}}$ (**?**, §2.2). Finally, $P_{\hat{z}}R^\top g = R^\top g - \langle g, R\hat{z}\rangle\hat{z} = R^\top g - g_1\hat{z} = R^\top(g - g_1\hat{q}) = R^\top P_{\hat{q}}g$, and $R^\top$ preserves norms. (b) Read off Theorem 1. (c) $\|Q - I\|_F^2 = 48 - 2\operatorname{tr} Q$. Decompose $\mathbb{R}^{24}$ by the real canonical form of Q into rotation planes Π_k (angles $\varphi_k \in (0, \pi)$) and the ±1 eigenspaces $E_\pm$, so that $\operatorname{tr} Q = \sum_k 2\cos\varphi_k + \dim E_+ - \dim E_-$. Write $\hat{z} = \sum_k a_k + b_+ + b_-$. Then $\cos\theta = \langle\hat{z}, Q\hat{z}\rangle = \sum_k \|a_k\|^2\cos\varphi_k + \|b_+\|^2 - \|b_-\|^2$ is a convex combination of $\{\cos\varphi_k, 1, -1\}$ that is smaller than 1. So either some $\varphi_k \geq \theta$ with $a_k \neq 0$, and then $\operatorname{tr} Q \leq 22 + 2\cos\theta$, or $b_- \neq 0$, and then $\operatorname{tr} Q \leq 22 < 22 + 2\cos\theta$ (for $\theta < 90°$). Hence $\|Q - I\|_F^2 \geq 4(1 - \cos\theta) = 8\sin^2(\theta/2)$. Equality needs exactly one rotation plane with angle θ, $E_- = \{0\}$, all other eigenvalues equal to 1, and $\hat{z}$ inside that plane. The plane then contains $\hat{z}$ and $Q\hat{z} = \hat{q}$, so it is Π and $Q = R$. (d) Along $\hat{z}' = \hat{z} + s\,t$, $\|\hat{q} + (\hat{z}' - \hat{z})\|^2 = 1 - 2s\sin\theta + s^2$, whereas $\|R\hat{z}'\| = \|\hat{z}'\|$. □

Proof of Theorem 3. The spherical gradient of c is $-P_{\hat{z}}\hat{q} = \cos\theta\,\hat{z} - \hat{q} = \sin\theta\,t$. Substituting into Theorem 1 gives $-\Delta t = -g_1 \operatorname{grad} c - g_2(1 - \cos\theta)t$. For the identity, write the first layer as $u_i = Wq_i + b$ for every token i. Then $\partial L/\partial W = \sum_i (\partial L/\partial u_i)q_i^\top$ and $g_i = W^\top \partial L/\partial u_i$, so $\sum_i\langle g_i, q_i\rangle = \langle W, \partial L/\partial W\rangle$, with $q_i = \hat{q}_i$. An explicit loss $\frac{\beta}{2}\mathrm{MSE}(\hat{z}, \hat{q})$, averaged over $B \cdot T \cdot 24$ entries, exerts the force $\beta\,\operatorname{grad} c/(24BT)$ on a token. The implicit coefficient reported in Table 6 is therefore $\beta_{\text{eff}} = -24BT\,g_1$. □

Proof of Theorem ?? and its supplement. We scale Λ_{24} so that minimal vectors have squared norm 4. (a) If x lay in two open caps of radius $t \leq 30°$ around distinct codewords c, c', then $\angle(c, c') < 2t \leq 60°$, which is impossible. So $\{\theta \leq t\}$ is a disjoint union of N caps, and (a') follows. A Monte-Carlo check with 6.7×10^7 uniform points gives measured/exact ratios of 0.970, 0.993 and 0.996 at 25°, 28° and 30° ($|z| \leq 1.6$). (b) *Lower bound.* Let $u = v_1 + v_2$ with minimal vectors v_1, v_2 and $\langle v_1, v_2\rangle = -1$, so $\|u\|^2 = 6$. For every minimal vector v, $u - v$ is a nonzero lattice vector, so $\|u - v\|^2 \geq 4$, i.e. $\langle u, v\rangle \leq 3$ and $\cos\angle(u, v) \leq 3/(2\sqrt{6}) = \sqrt{3/8}$, with equality at $v = v_1$. The direction of u is therefore exactly $\arccos\sqrt{3/8}$ from the codebook. *Upper bound.* **?**, Thm. 5.3 show that the normalized norm-4 and norm-6 shells of Λ_{24} form a universal polar dual pair: the universal minima of every absolutely monotone potential of the first are exactly the points of the second. Universal minima are deep holes (**?**, Prop. 14.4.1), so the covering radius equals $\arccos\sqrt{3/8}$. Numerically, four constructed norm-6 directions have maximal inner product 0.61237243 ($\sqrt{3/8} = 0.61237244$) with exactly 552 equidistant nearest codewords. Finally, $\|\hat{z} + \hat{q}\| = 2\cos(\theta/2) \geq 1.796$. (c) The union bound $P(\theta \leq t) \leq \sum_c P(\hat{z} \in \mathrm{Cap}(c, t)) = N\mu(t)$ holds for every code. Every quantile of θ is therefore at least the corresponding quantile of the "round-cell" law $\min(1, N\mu(t))$, and $\mathbb{E}\theta$ is at least its mean; the numbers in (c') and (e) follow by numerical integration. (d) Co_0 acts transitively on the minimal vectors (**?**), so all Voronoi cells are congruent. In geodesic polar coordinates $x = \cos s\,c + \sin s\,u$ around a codeword c (u a unit tangent), the uniform measure is proportional to $\sin^{22} s\,ds\,du$. The geodesic crosses the bisector with c' where $\tan s = (1 - \langle c, c'\rangle)/\langle u, c'\rangle$, so the cell is $\{s < \rho(u)\}$ with $\rho(u) = \min_{c' \neq c,\ \langle u, c'\rangle > 0} \arctan\big((1 - \langle c, c'\rangle)/\langle u, c'\rangle\big)$. Integrating over u gives equation **??** with $\bar{F}(s) = P_u(\rho(u) > s)$. □

Proof of Proposition 1. With g fixed, $-P_{\hat{z}}g$ is minus the spherical gradient of $x \mapsto \langle g, x\rangle$. Its critical points are $\pm g/\|g\|$, and the stable one, $-g/\|g\|$, does not involve $\hat{q}$. By Theorem 2(a), the RLSQ field has norm $\|P_{\hat{q}}g\|/\|z\| > 0$ whenever $\operatorname{grad}\ell(\hat{q}) \neq 0$, so it has no fixed point in the cell. □

B.3 The exact angle law: computation and checks

We estimate $\bar{F}$ in equation **??** by Monte Carlo over 2^{19} tangent directions at two different codewords. The two ρ distributions agree quantile by quantile, as transitivity requires, and the total mass $N\,\mathbb{E}_u[\mu(\rho(u))]$ is 1.0006. For 99.77% of directions the geodesic leaves the cell through the bisector with a first-shell neighbour (at 60°); for the rest it leaves through the second shell (at 75.5°). The density of θ is the product of two factors. The volume growth $N\mu'(s)$ grows by a factor 1.9–1.5 per

degree over 30–44°. The cell survival $\bar{F}(s)$ is 1 up to 30° and falls from 0.92 at 38° to 0.06 at 42°. Their product peaks at 40°.

All cells are the same. We drew 2^{26} uniform points, 341 per cell on average, and computed per-cell mean, quantiles, maximum and a 10-bin χ^2 statistic for all 196,560 cells. For every statistic the spread across cells equals that of a permutation null (ratio 0.994–0.999). The fraction of cells beyond the χ^2 0.999 quantile is 0.093%, against 0.1% expected.

Table 4: Quantiles of θ (degrees). Uniform input: the any-code bound (round cells, $\min(1, N\mu)$), the Poisson law of a random code with its Monte Carlo (MC), and the exact Leech law (Theorem **??**(d)) with its Monte Carlo. Trained tokenizers: each model's own first training epoch.

	Uniform input					Trained tokenizers, first training epoch				
Quantile (%)	Bound	Rand.	Rand. MC	Λ_{24} law	Λ_{24} MC	STE $\lambda_{\mathrm{dir}}{=}0$	STE $\lambda_{\mathrm{dir}}{=}5$	RLSQ $\lambda_{\mathrm{dir}}{=}5$	RLSQ-7F	RLSQ-12F
0.01	25.96	25.96	26.02	25.96	25.97	25.92	26.04	28.47	27.34	28.25
0.1	28.91	28.91	28.90	28.91	28.91	29.08	28.91	30.82	30.25	30.56
1	32.25	32.25	32.24	32.25	32.25	32.49	32.20	33.48	33.20	33.28
5	34.86	34.90	34.90	34.86	34.86	35.00	34.80	35.61	35.46	35.47
10	36.06	36.16	36.15	36.07	36.07	36.16	36.01	36.63	36.53	36.51
25	37.74	38.01	38.00	37.79	37.78	37.81	37.74	38.12	38.08	38.04
50	39.08	39.73	39.73	39.28	39.28	39.28	39.24	39.48	39.46	39.42
75	39.89	41.16	41.16	40.43	40.43	40.42	40.40	40.56	40.55	40.51
90	40.26	42.26	42.26	41.26	41.26	41.25	41.23	41.35	41.35	41.31
95	40.37	42.84	42.84	41.68	41.68	41.66	41.65	41.75	41.75	41.72
99	40.46	43.82	43.82	42.36	42.36	42.34	42.33	42.41	42.41	42.39
99.9	40.48	44.77	44.77	42.99	42.98	42.96	42.95	43.04	43.04	43.01
99.99	40.48	45.46	45.44	43.46	43.44	43.41	43.40	43.51	43.51	43.49
Mean	38.54	39.41	–	38.91	38.91	38.93	38.87	39.18	39.15	39.11
Std. dev.	1.83	2.46	–	2.14	2.14	2.09	2.15	1.91	1.96	1.94
$P(\theta \le 30°)$	0.220%	0.220%	–	0.220%	0.218%	0.188%	0.224%	0.046%	0.082%	0.059%
Max (sample)	40.48	–	46.78	52.24	45.17	45.20	45.00	45.48	45.75	46.15
Samples	–	–	4.2M	–	16.8M	415M	415M	415M	415M	398M

B.4 Implementation

The rotation changes one line of an STE quantizer. The forward value is the STE expression evaluated without gradient, so outputs, code indices and evaluation metrics are bit-identical to STE. The backward pass sees only $R\hat{z}$, with R held constant.

```
def rlsq(z, codebook):   # z: (..., 24); codebook: (N, 24) unit rows
    z = F.normalize(z.float(), dim=-1)
    q = codebook[nearest(z, codebook)]   # exact NN search, fp32
    if not torch.is_grad_enabled():
        return q
    zd = z.detach()
    h = F.normalize(zd + q, dim=-1)      # Householder axis, constant
    Rz = z - 2 * (z * zd).sum(-1, keepdim=True) * zd    # value: -z
    Rz = Rz - 2 * (Rz * h).sum(-1, keepdim=True) * h    # value:  q
    with torch.no_grad():
        ste = z + (q - z)                # the STE forward value
    return ste + (Rz - Rz.detach())  # == STE forward; Jacobian R
```

By Theorem **??**(b), $\|\hat{z}+\hat{q}\| \ge 1.796$, so the Householder axis h is always well defined. We checked the gradient against an explicit 24×24 matrix $R^\top g$ in fp64 (agreement to 10^{-15}) and asserted forward equality with STE at every step of a smoke run.

B.5 Experimental details

Tokenizer, data and training. Encoder and decoder are 6-layer Transformers (width 256, 4 heads, feed-forward 1024, no dropout) over bars with six channels (open, high, low, close, volume, turnover), with one 24-dimensional latent per bar. The data are Chinese A-share K-lines at 1, 5, 15, 30, 60 and 120 minutes and 1 day, sampled with ratios $0.25/0.25/0.12/0.12/0.10/0.08/0.08$ in windows of 1,088 (1 and 5 min), 576 (15–60 min), 560 (120 min) and 304 (daily) bars. The data are split in time: training up to 2022-06-30, validation up to 2024-01-01, test afterwards. An

epoch draws 512k windows. The objective is a log-cosh reconstruction loss ($\delta = 1$), an OHLC-consistency penalty with weight 5, and a hinge loss with weight λ_{dir} on the direction of the intra-bar move $C_t - O_t$ and the overnight gap $O_t - C_{t-1}$. We use AdamW (Loshchilov & Hutter, 2019) (learning rate 3×10^{-4}, $\beta = (0.9, 0.95)$, weight decay 0.1) with a warmup–stable–decay schedule (Hu et al., 2024; **?**): 1.5 warm-up epochs, then stable until epoch 42.5, then a $1 - \sqrt{\cdot}$ decay to zero at epoch 100. The global batch is 256 windows on 4 GPUs.

Paired design. Each pair shares the seed, initialization, data order and every hyperparameter. With the rotation switched off, the RLSQ training script reproduces the STE run bit for bit on every logged metric of a short run. Checkpoints are selected by validation MSE; the test split is only reported.

Replications. RLSQ-7F and RLSQ-12F differ from the paired runs in data preprocessing (zero-volume filtering, asinh volume compression) and in the direction-loss settings. RLSQ-7F keeps the seven frequencies with a wider model (width 384, 8 heads). RLSQ-12F uses twelve frequencies from Chinese and US equities with an asymmetric encoder/decoder (widths 256/512). Neither has an STE counterpart.

Angle measurements. For every model we replay its own first training epoch through the trained encoder, with the training sampler, seed and augmentation: 512k windows, or 398–415M bar tokens per model. We compute θ by exact nearest-neighbour search in fp32 with TF32 disabled, and record 0.01-degree histograms. Gradient statistics use 8 validation batches (376k tokens per checkpoint), with g computed by autograd on each checkpoint's own training loss.

B.6 Further experimental results

Training curves and per-channel errors. Figure 4 shows the paired validation curves and Table 5 the per-channel errors. All six channels improve at both operating points, by 4.9–10.8%.

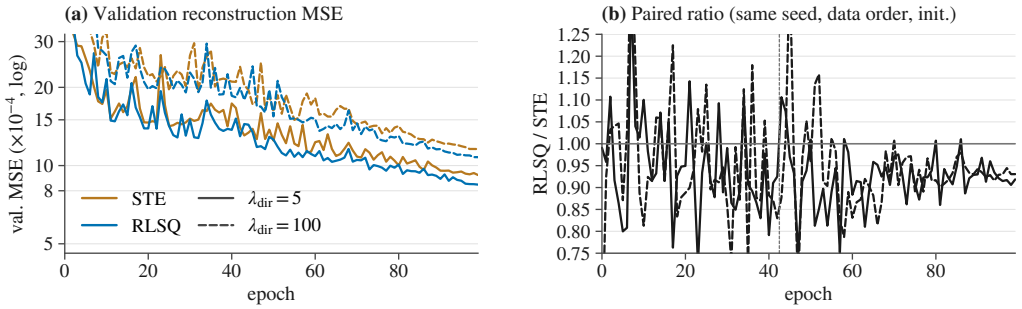


Figure 4: Paired validation curves. (a) Reconstruction MSE of the four runs. (b) Ratio RLSQ/STE per epoch for the two pairs. Early, high-learning-rate epochs are noisy. After the learning rate starts to decay (dashed vertical line, epoch 42.5), RLSQ pulls ahead; the ratio is below 1 in all of the last ten epochs (0.91–0.97 for $\lambda_{\text{dir}} = 5$, 0.92–0.94 for $\lambda_{\text{dir}} = 100$).

The theory on trained models. Table 6 lists the quantities of Theorems 1 and 3 on three checkpoints, and Table 7 breaks them down by angle. The relative difference between the surrogates grows with the angle, from 0.091 at $\bar{\theta} = 32.5°$ to 0.143 at 43.2°. The ratio of outward pushes tracks $1/\cos\bar{\theta}$ within 3% in every bin with more than 40k tokens.

Radial weight and quantile shifts. Table 8 gives the tilt fits. The effect of the depleted core is concentrated in the lower tail because w departs from 1 only near the codeword. To first order a lower quantile shifts by $-\ln R/a$, where R is the ratio of measured to exact mass below it and $a \approx 23 \cot t$ per radian is the log-slope of the volume law (0.83 per degree at 26°, 0.41 at the median). For the RLSQ tokenizer with $\lambda_{\text{dir}} = 5$, $R = 0.07$ at the 0.01% quantile gives a predicted shift of 3.2° against a measured 2.5°; the first-order formula overestimates large shifts. At the median $R = 0.91$

Table 5: Per-channel validation MSE (O/H/L/C: open, high, low, close; V: volume; A: turnover), shape errors of the upper wick, body and lower wick, and the signature MMD of reconstructed windows, for the paired runs of Table **??**.

		Per-channel val. MSE ($\times 10^{-4}$)						Shape MSE ($\times 10^{-4}$)			
λ_{dir}	Est.	O	H	L	C	V	A	Upper	Body	Lower	Sig-MMD
5	STE	9.86	10.55	10.23	5.25	9.90	9.38	11.51	10.45	10.84	0.386
	RLSQ	9.02	9.84	9.48	4.96	8.83	8.55	10.75	9.78	10.18	0.351
	Δ	-8.5%	-6.8%	-7.3%	-5.5%	-10.8%	-8.9%	-6.6%	-6.5%	-6.1%	-9.0%
100	STE	11.18	14.13	13.65	6.68	12.24	11.59	13.54	9.99	12.59	0.499
	RLSQ	10.17	13.27	12.98	5.99	11.36	10.88	13.34	9.50	12.72	0.462
	Δ	-9.0%	-6.1%	-4.9%	-10.4%	-7.2%	-6.1%	-1.5%	-4.9%	+1.0%	-7.5%

Table 6: Theorems 1 and 3 on trained checkpoints. Outward push: mean component of the descent direction along t, normalized by $\|g\|$, RLSQ over STE on the same tokens; brackets give $1/\cos\bar{\theta}$. Alignment: cosine between $\hat{w}$ and the descent direction on the sphere at $\hat{q}$. Encoder-gradient difference: $\|\nabla^{\text{RLSQ}} - \nabla^{\text{STE}}\|/\|\nabla^{\text{RLSQ}}\|$ for the encoder parameters, aggregated over 16, 64 and 512 windows. β_{eff}: implicit commitment coefficient in the units of an explicit loss $\frac{\beta}{2}\text{MSE}(\hat{z}, \hat{q})$.

Quantity	STE, λ_{dir}=0	STE, λ_{dir}=5	RLSQ, λ_{dir}=5
Thm. 1 vs. autograd, max rel. error	5.2×10^{-7}	4.8×10^{-7}	5.0×10^{-7}
$\sum g_1 = \langle W, \partial L/\partial W\rangle$, rel. error	1.5×10^{-5}	6.6×10^{-6}	4.9×10^{-5}
Mean angle $\bar{\theta}$ (deg)	38.92	38.88	39.17
Outward push RLSQ/STE $[1/\cos\bar{\theta}]$	1.281 [1.285]	1.288 [1.285]	1.291 [1.290]
Fraction $g_2 < 0$ / $g_1 < 0$	0.772 / 0.494	0.751 / 0.499	0.738 / 0.497
Energy share of g_1 / g_2 (isotropic: 0.042)	0.032 / 0.178	0.035 / 0.165	0.035 / 0.162
Alignment $\cos(\hat{w}, -\text{grad}\,\ell)$ (null 0 ± 0.21)	0.254	0.229	0.217
Per token $\|\Delta\|/\|\nabla^{\text{RLSQ}}\|$: median / p90	0.098 / 0.238	0.097 / 0.238	0.099 / 0.241
Per token $\cos(\nabla^{\text{STE}}, \nabla^{\text{RLSQ}})$: median	0.9960	0.9961	0.9959
Encoder-gradient rel. difference: 16 / 64 / 512 windows	0.10 / 0.13 / 0.16	0.11 / 0.16 / 0.23	0.11 / 0.16 / 0.29
Implicit commitment β_{eff}: mean / mean $\|\cdot\|$	-0.0008 / 0.016	0.0005 / 0.021	-0.0012 / 0.022

gives $0.22°$ against a measured $0.20°$. For STE the formula is accurate to $0.01°$. The upper tail is held in place by the fixed cell walls.

Dynamics and per-frequency tilt. Table 9 lists the tilt at every checkpoint. The RLSQ tilt is ordered the same way in all three RLSQ models: largest at 1 minute and smallest at 1 day (0.067 to 0.052, 0.059 to 0.041, and 0.050 to 0.041). STE is near zero at every frequency. The ordering correlates with the number of tokens per frequency, but it is not an exposure effect, because the core shrinks with training while the most frequent bars have the deepest core. *Per code:* codes that are used more often have smaller mean angles (Spearman correlation between log usage and mean angle, -0.64 to -0.68, for all five models). This heterogeneity exists under both estimators and cancels in the pooled distribution.

Single code switches. For 2,048 tokens per checkpoint we follow each surrogate's descent direction along the geodesic until the nearest codeword changes, decode again with the new code while keeping everything else fixed, and record the exact change in loss (Table 10). Removing the implicit commitment helps in all six comparisons (four with $|z| \geq 2.9$). The $\cos\theta$ damping also helps in all six ($|z| = 3.2$–4.5). First-order Taylor estimates carry no information at this scale: their correlation with the exact change is -0.06 to -0.02, although a linear-regime check at $\epsilon = 10^{-3}$ gives at least 0.995. A code switch moves the decoder input by at least $60°$, where the decoder is far from linear.

B.7 Forcing small angles costs rate

We ran a pair of tokenizers that differ only in an explicit commitment loss, $\beta = 0.163$ against $\beta = 0$. Over epochs 15–24 of the stable phase, the typical $\cos\theta$ rose from 0.774 to 0.869 ($\theta \approx 39° \to 30°$). The effective code rate fell from 13.17 to 8.99 bits, and validation MSE was $2.07\times$ higher; at the end of training it was $2.49\times$ higher. A learnable β does not help either. Because $\hat{z}$ and the codewords are

Table 7: Gradient statistics per angle bin. In the lowest bin (5–11k tokens) the outward-push ratio exceeds $1/\cos\bar{\theta}$ by 2–8%. This happens exactly when the mean of g_1 is negative, so near the codeword the implicit commitment of STE pulls slightly inward on average.

Angle bin (deg)	<34	[34, 37)	[37, 39)	[39, 41)	[41, 43)	≥ 43
$\bar{\theta}$ (deg)	32.5	35.9	38.1	40.0	41.6	43.2
Tokens	11,426	52,453	105,333	155,827	50,995	286
$\lvert\Delta\rvert/\lVert\nabla^{\mathrm{RLSQ}}\rVert$, mean	0.091	0.102	0.112	0.121	0.129	0.143
Outward push RLSQ/STE, STE, $\lambda_{\mathrm{dir}}=0$	1.214	1.247	1.264	1.289	1.315	1.363
Outward push RLSQ/STE, STE, $\lambda_{\mathrm{dir}}=5$	1.282	1.256	1.276	1.297	1.305	1.349
Outward push RLSQ/STE, RLSQ, $\lambda_{\mathrm{dir}}=5$	1.261	1.248	1.275	1.297	1.316	1.354
Theory $1/\cos\bar{\theta}$	1.185	1.234	1.271	1.305	1.337	1.372

Table 8: Radial-weight fits. κ_1: one-parameter tilt $w \propto e^{\kappa_1 s}$, fitted to the mean. (κ, λ): two-parameter tilt $\ln w = c + \kappa(s-39^\circ) + \lambda(s-39^\circ)^2$, fitted to the mean and the second moment. The next three columns give the maximal error over the 13 quantiles of Table 4: with the exact law as is ($w \equiv 1$), with the two-parameter model, and with the two-parameter model from the 1% quantile upward. The STE tokenizer with $\lambda_{\mathrm{dir}} = 0$ has a local dip at 28–33° ($w \approx 0.82$) that no smooth tilt captures.

	Tilt parameters			Max. quantile error (deg)			$P(\theta \leq 30^\circ)$ (law 0.220%)	
Model	κ_1	κ	λ	$w \equiv 1$	2-param	2-param, $\geq$p1	Measured	2-param
STE, $\lambda_{\mathrm{dir}}=0$	+0.005	−0.002	−0.0029	0.24	0.64	0.04	0.188%	0.167%
STE, $\lambda_{\mathrm{dir}}=5$	−0.009	−0.012	−0.0011	0.08	0.04	0.04	0.224%	0.221%
RLSQ, $\lambda_{\mathrm{dir}}=5$	+0.065	+0.052	−0.0081	2.51	0.31	0.07	0.046%	0.057%
RLSQ-7F	+0.055	+0.047	−0.0046	1.38	0.23	0.05	0.082%	0.085%
RLSQ-12F	+0.047	+0.033	−0.0081	2.29	0.32	0.04	0.059%	0.069%

unit vectors, the commitment term equals β times the token average of $(1-\cos\theta)/12$. Its derivative in β is therefore always positive, and gradient descent drives β to zero.

Constraint-driven curriculum via primal-dual optimization. Rather than manually scheduling or reweighting the six tokenizer losses, we formulate legality and directional fidelity as explicit inequality constraints, while retaining log-cosh, pattern, signature, and SigMMS losses as soft fidelity objectives. Specifically, we impose

$$L_{\mathrm{legal}}(\theta) \leq \epsilon_{\mathrm{legal}}, \qquad L_{\mathrm{dir}}(\theta) \leq \epsilon_{\mathrm{dir}}, \tag{10}$$

and optimize the remaining objectives under these feasibility requirements. We adopt an augmented-Lagrangian formulation, in which the primal objective combines the four soft losses with adaptive penalties induced by constraint violations. The corresponding dual variables are updated using a PI controller driven by exponentially smoothed constraint violations:

$$v_{i,t} = \mathrm{EMA}[g_i(\theta_t)]\,, \tag{11}$$

$$I_{i,t} = I_{i,t-1} + v_{i,t}, \tag{12}$$

$$\lambda_{i,t+1} = [\lambda_{i,t} + k_P v_{i,t} + k_I I_{i,t}]_+\,, \tag{13}$$

where

$$g_{\mathrm{legal}}(\theta) = L_{\mathrm{legal}}(\theta) - \epsilon_{\mathrm{legal}}, \qquad g_{\mathrm{dir}}(\theta) = L_{\mathrm{dir}}(\theta) - \epsilon_{\mathrm{dir}}. \tag{14}$$

Thus, the dual variables increase when the corresponding constraints are violated, decrease when sufficient slack emerges, and stabilize as the constraints approach their target levels. Positivity and anti-windup clipping are applied to prevent unstable dual accumulation.

This adaptive dual feedback gives rise to an emergent curriculum without prescribing an explicit training schedule: early optimization is dominated by the feasibility constraints that enforce valid K-line structures and directional consistency; once these requirements are approximately satisfied, their dual pressure stabilizes, allowing greater optimization capacity to be devoted to increasingly fine-grained numerical, morphological, geometric, and distributional fidelity. Consequently, the progression from legality and direction toward pattern, path geometry, and path-distributional fidelity emerges from constraint satisfaction rather than from manually specified loss weights or stage boundaries.

Table 9: Tilt and lower-tail mass on a fixed input (the first 40 batches of the first training epoch, 2.1M tokens) at every saved checkpoint of the paired runs with $\lambda_{\mathrm{dir}} = 5$. The exact values are $\kappa_1 = 0$ and $P(\theta \le 30^\circ) = 0.220\%$.

	Tilt κ_1 (per degree)		$P(\theta \le 30^\circ)$ (%)	
Epoch	STE	RLSQ	STE	RLSQ
0	0.058	0.204	0.099	0.004
1	0.048	0.205	0.062	0.002
2	0.031	0.160	0.122	0.009
3	0.017	0.121	0.159	0.027
5	0.019	0.111	0.161	0.021
8	0.020	0.103	0.161	0.016
12	0.015	0.085	0.188	0.027
20	0.018	0.082	0.181	0.035
30	0.002	0.078	0.193	0.035
42	0.003	0.073	0.176	0.043
55	0.004	0.077	0.166	0.056
70	0.001	0.080	0.185	0.033
85	−0.000	0.069	0.189	0.045
99	−0.009	0.065	0.221	0.046

Table 10: Single code switches with exact re-decoding. Each entry is the fraction of tokens, among those whose new codes differ, for which the first surrogate gives the lower loss, with the number of such tokens and a sign-test z. "Leak" is the implicit commitment $-g_1 \operatorname{grad} c$ of Theorem 3; "attenuation" is the factor $\cos\theta$ on g_2. STE = RLSQ + leak + attenuation.

Paired comparison (tokens whose new codes differ)	STE, λ_{dir}=0	STE, λ_{dir}=5	RLSQ, λ_{dir}=5
RLSQ better than STE	0.466 (352, $z = -1.3$)	0.516 (314, $z = +0.6$)	0.516 (343, $z = +0.6$)
Leak, no attenuation: RLSQ better than RLSQ+leak	0.553 (282, $z = +1.8$)	0.629 (278, $z = +4.3$)	0.617 (303, $z = +4.1$)
Leak, with attenuation: STE−leak better than STE	0.538 (303, $z = +1.3$)	0.586 (292, $z = +2.9$)	0.613 (323, $z = +4.1$)
Attenuation, no leak: STE−leak better than RLSQ	0.610 (223, $z = +3.3$)	0.621 (169, $z = +3.2$)	0.660 (188, $z = +4.4$)
Attenuation, with leak: STE better than RLSQ+leak	0.632 (220, $z = +3.9$)	0.661 (192, $z = +4.5$)	0.626 (211, $z = +3.6$)
Off-plane part only better than RLSQ	0.547 (1025, $z = +3.0$)	0.569 (974, $z = +4.3$)	0.582 (987, $z = +5.1$)
Random tangent better than RLSQ (control)	0.304 (1661, $z = -16.0$)	0.296 (1620, $z = -16.4$)	0.305 (1695, $z = -16.1$)
First-order oracle better than RLSQ	0.416 (1764, $z = -7.0$)	0.467 (1765, $z = -2.8$)	0.450 (1767, $z = -4.2$)
corr(first-order $\Delta\ell$, exact $\Delta\ell$) at the switch	−0.044	−0.055	−0.022
Linear-regime certificate: corr at $\epsilon = 10^{-3}$	0.999	0.995	0.995

Limitations and future work. Kairos models K-lines only. It does not observe order books, news, or fundamentals, and the cross-sectional structure of a market enters only through post-training rewards, not through the architecture; conditioning the predictor on the contemporaneous cross-section is a natural extension. Our evaluation is statistical: IC and RankIC measure ranking quality and do not account for transaction costs, market impact, or capacity, so they should not be read as realized trading performance. Financial data are non-stationary, and although our test period lies strictly after the training period, performance under regime changes not represented in the data remains untested. On the modeling side, the flat vocabulary of 196,560 tokens makes the embedding and the output head a large share of the parameters, which motivates factorized parameterizations of the Leech codebook; spherical quantization also discards the norm of the latent, and an explicit code for magnitude may recover part of the information that is currently lost. Finally, FSP requires training a reverse-time teacher, which adds a one-time cost to pretraining, and the gains from post-training are not uniform across tasks; understanding which decision objectives can be learned from price data alone, and which cannot, is an open question.

Table 11: Schematic evolution of the tokenizer losses during training. The constraint losses L_{legal} and L_{dir} are reduced rapidly in the early stages and subsequently stabilize, while the soft fidelity objectives progressively dominate finer-grained numerical, morphological, geometric, and distributional refinement.

	L_{legal}	L_{dir}	L_{log}	L_{pattern}	L_{sig}	L_{sigMMS}
Early	$\downarrow\downarrow$	$\downarrow\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
Middle	$\rightarrow$	$\downarrow$	$\downarrow$	$\downarrow\downarrow$	$\downarrow$	$\downarrow$
Late	$\rightarrow$	$\rightarrow$	$\downarrow$	$\downarrow$	$\downarrow\downarrow$	$\downarrow\downarrow$