Four Engines, Four Commodity Microstructures

An Order-Flow-Aware, Roll-Integrity-Guarded Architecture for
Commodity Micro-Futures Calibration — Q7 AEGIS AI S4

Haneef R. Haqq, MBA[1]

[1]Quant7 Alpha, LLC

Technical Whitepaper · v1.0 · Strategy S4 (Commodity Micro Futures)

June 30, 2026

Abstract

We describe S4, the commodity member of the Q7 AEGIS AI family, which trades four CME micro-sized commodity futures—micro WTI crude oil (MCL), micro gold (MGC), micro copper (MHG), and micro natural gas (MNG)—across four intraday horizons (1, 3, 5, 15 minutes). The design premise is that commodity micros realize the four canonical Q7 regimes—trend persistence, volatility-expansion breakout, mean-reverting range drift, and dealer liquidity sweep—with substantially more *cross-asset heterogeneity* than digital assets or index micros: crude trends on inventory shocks but reverts in mid-week chop, gold breaks out on macro events but ranges in calm sessions, copper tracks overnight Asian demand flow, and natural gas is sweep-prone with sharp single-direction moves. S4 runs four concurrent engines that each target one regime, arbitrated by a seven-component confluence score and a five-state volatility-regime router. Position leverage is a function of order-flow conviction under the Pine `SIZE_MARGIN` model rather than a fixed multiplier, grounded in the empirical result that order-flow imbalance, not trade imbalance, is the dominant near-linear predictor of short-horizon drift. A path-geometry operator (velocity / jerk / curvature) supplies a confluence boost, and a per-configuration, train-equals-serve hybrid-AI gate filters entries. Distinctively, S4 carries an explicit roll-integrity control: continuous-contract back-adjustment introduces synthetic price discontinuities at each roll boundary that must be detected and neutralized rather than traded through. We detail the engine semantics, the confluence and gating logic, the order-flow leverage map, the adaptive four-leg exit, the contract-roll data-integrity control, the risk construction, and the calibration acceptance criteria, and we state the architecture's limitations explicitly. No realized-performance claims are made.

1. Edge Thesis

S4 trades directional commodity-micro prices, where the four canonical Q7 engines each correspond to a real microstructural regime, but the regime *mix* differs sharply across the four contracts. **Trend (TRND)** fires when the moving-average stack and trend-strength index align—the edge condition for supply-shock-driven persistence in crude (MCL) and in gold (MGC) during macro events. **Breakout (BRK)** fires on volatility expansion out of a Keltner–Bollinger squeeze, which on commodity micros captures inventory-day expansions and dollar-flip breakouts that dominate gold's opening sessions. **Mean-Reversion (MR)** fires on intraday extremes inside ranges, where dealers fade momentum bursts—the mid-session chop characteristic of copper (MHG) and range-bounded mid-week crude. **Sweep (SWP)** targets prior-high/low liquidity sweeps with rejection wicks, the single-direction stop-runs that dominate natural-gas (MNG) intraday. The four engines partition the regime space; the confluence and volatility-regime layers arbitrate which fires. The structural argument for the cascade is precisely the cross-asset heterogeneity: no single-engine specialist can cover crude's trend/revert duality, gold's breakout/range duality, copper's volume dependence, and gas's sweep dominance simultaneously.

2. Architecture

S4 is organized in three independently testable layers.

Signal. Per (asset, timeframe), the four engines compute entry candidates each bar and write a candidate direction plus an engine tag. Arbitration selects the first eligible engine whose direction-aligned confluence clears its per-engine floor, subject to a per-engine cooldown. Engines can be disabled per configuration; on the highest-volatility micros (MNG, MHG) at short horizons, calibration frequently selects mean-reversion-dominant profiles when trend and breakout structurally underperform that substrate.

Intelligence. Five subsystems run in parallel: the confluence aggregator (Section 3); the volatility-regime router (Section 6); the path-geometry operator (Section 7); the four-leg adaptive exit planner (Section 8); and the hybrid-AI overlay (Section 9).

Execution. Engine arbitration → confluence gate → cooldown → hybrid-AI gate → order-flow-scaled `SIZE_MARGIN` leverage sizing → entry. Exits run the four-leg cascade with a break-even arm, a three-tier MFE ratchet, and a runner. A NinjaScript C# port emits orders directly to NinjaTrader and a Pine v6 implementation emits TradingView alerts; the three implementations are held in specification parity.

Asset	Underlying	Horizons	Configs	Lev. ceiling
MGC	Micro gold	1·3·5·15m	4	100×
MHG	Micro copper	1·3·5·15m	4	75×
MCL	Micro WTI crude	1·3·5·15m	4	50×
MNG	Micro nat. gas	1·3·5·15m	4	25×

The 16 configurations (4 assets × 4 horizons) are each calibrated independently. The leverage ceiling is a per-instrument, margin-implied bound that *decreases with contract volatility*—gold the highest, natural gas the lowest—and it is a ceiling, never an operating point.

3. Confluence Scoring and Gating

Per bar, seven 0–100 component scores are computed—trend agreement, trend-strength (ADX) band, direction-specific oscillator (RSI) band, volume ratio, multi-timeframe agreement, Bollinger position, and candle-pattern strength—and combined into direction-conditional confluence:

$$\mathrm{conf}_{dir} = \Sigma_k\, w_k \cdot \mathrm{score}_{k,dir}, \quad \Sigma_k\, w_k = 1.$$

Each engine carries a calibrated confluence floor; the effective gate applies a per-direction asymmetry adjustment under a fixed feasibility cap (the Pine v6 hard ceiling). Engines can be disabled per configuration when structurally adverse for that (asset, horizon), which is how mean-reversion-dominant profiles emerge for the high-volatility micros—an outcome of calibration, not a hand-set rule. The volume-ratio component carries particular weight on copper, whose intraday regime is driven by overnight Asian demand flow.

4. The Four Engines and Regime Coverage

The engines are deliberately non-overlapping in their activation conditions, so that confluence arbitration is selecting among regimes rather than among redundant signals. TRND and BRK target directional persistence and expansion; MR and SWP target reversion and liquidity events. The coverage argument is empirical, not aesthetic: each engine's activation condition maps to a documented intraday behavior of a specific commodity book, and the union of conditions spans the observed regime distribution. The breakout engine is gated on a Keltner–Bollinger squeeze—the Bollinger band contracting inside the Keltner channel marks volatility compression, and the engine fires only on the expansion that resolves it, which suppresses the false breakouts that plague naive band-cross logic on choppy commodity sessions.

5. Order-Flow-Driven Leverage and SIZE_MARGIN

S4 does not size with a fixed multiplier; per-trade leverage is a function of confluence and *order-flow* conviction under the Pine `SIZE_MARGIN` model, capped per instrument. Position notional is set by a margin fraction times effective leverage, where effective leverage is attenuated—never the protective stop—to give risk room (Section 8). This choice is grounded in the limit-order-book literature. Order-flow imbalance (OFI)—the net change in size at the best bid and ask, inclusive of cancellations—explains short-horizon mid-price drift far better than trade imbalance (reported $R^2 \approx 65\%$ vs. 32%) [1, 3]. Crucially, the price-impact slope is *inversely proportional to market depth* [1], which dictates depth-normalized order-flow: leverage is attenuated when our own flow is large relative to local depth—a correction that matters most on the thin natural-gas and copper books and least on liquid gold. The Markovian queueing treatment further yields an order-flow-implied volatility, $\sigma_{OF} \propto \sqrt{(\lambda/D)}$, forecastable from flow statistics without observing price [2]—used to open the breakout gate and scale trend leverage when σ_{OF}/σ_{ATR} expands, and to favor mean-reversion when it contracts. The same theory supplies a closed-form queue-imbalance direction probability with empirical first-lag *negative* autocorrelation [2], the structural fact the MR engine exploits.

6. Volatility-Regime Router (Λ-Vol)

A deterministic five-state classifier—trend-calm, trend-active, mean-revert, stress-high, transition—maps the short/long volatility ratio and drift (in standard-deviation units) to regime-conditional stop and target multipliers. It is a rule-based state machine, not a learned model, chosen for interpretability and stability; it routes exit geometry and flags engine downsizing under stress. Commodity micros make the regime layer indispensable: natural gas resides in the stress-high state far more often than the other three contracts due to gas-day volatility blowouts, and a single static exit geometry serves the resulting multi-modal return distribution poorly [8]. Per-instrument volatility thresholds—particularly the stress-high boundary for MNG—are a calibration lever rather than a global constant.

7. Path-Geometry Operators (ζ-Field)

The ζ-Field supplies a curvature signal to the direction-aligned confluence channel via three operators on the log-price path: velocity (first derivative), jerk (third derivative, via a Savitzky–Golay coefficient), and curvature (discrete circulation). The three combine into a single signed score that boosts the aligned side additively, leaving the gate structure unchanged; a confluence-deadband guard prevents the boost weight from calibrating to a level that would suppress the underlying confluence below its floor. The rationale is that the *order* of moves and cross-channel geometry carry information that vanishes under bag-of-features representations—the same premise that motivates the signature method in machine learning [6]; the ζ-Field is a lightweight, deterministic instance of that geometric coverage, and it is most decisive for the breakout and sweep engines, where path *shape* rather than endpoint statistics is the signal.

8. Adaptive Four-Leg Exit with Runner

Each position scales out across four ATR-spaced target legs plus a runner, with engine-specific leg allocations and target multipliers (calibrated per configuration). A break-even stop arms after the first leg or once favorable excursion exceeds a threshold; a three-tier MFE ratchet advances the trail anchor in stages as maximum favorable excursion accumulates past the second leg; a minimum-bars-in-position guard protects short-horizon mean-reversion signals on the high-volatility micros from premature time-stops. Stop-loss governance is leverage-aligned: a protective noise-floor stop is preserved on every entry, per-engine SL multipliers scale the stop to engine character, and headroom is bought by *reducing leverage* rather than by widening or cutting the stop—so exit coherence is never traded for sizing. Any hard-risk event (circuit breaker, rolling halt, liquidation guard, macro gate, broker disconnect) closes all remaining legs immediately; the leg machinery operates only on the normal exit path.

9. The Hybrid-AI Gate (Train = Serve)

S4 filters rule-based entries through a hybrid-AI overlay that fuses three models—a supervised directional learner, a financial-language sentiment model, and a market-regime classifier—in the spirit of the multi-modal architecture of [9], but adapted to per-configuration, intraday, train-equals-serve operation. The motivation is the same one that paper documents empirically: trend and mean-reversion logic fail through regime transitions and are blind to news shocks, and a fusion of structured ML with sentiment and regime awareness is materially more robust than any isolated predictor. The hybrid-AI overlay is itself a calibratable on/off gate—if the optimizer selects it off, the rule-based stack carries the decision.

9.1 Directional model — gradient-boosted trees (XGBoost)

The directional learner is an XGBoost classifier that predicts whether the forward move is favorable. The reference design [9] frames this as a binary next-period-direction problem (label 1 if the next-period return is positive) over a compact technical feature set, with ~200 boosting rounds, tree depth 6, and a 0.05 learning rate on a 70/30 split. S4 adapts that recipe to its short-horizon, per-(asset, timeframe) setting: labels are the forward 10-bar return thresholded at ±0.5% (favorable/unfavorable), and the feature set is *direction-conditional*—separate long and short framings (roughly a dozen technical and microstructure features plus an is-long indicator and the ζ-Field velocity/jerk terms) are computed so one model scores both sides. Mechanically, XGBoost is an additive ensemble of shallow regression trees fit by stagewise gradient boosting on a regularized logistic objective: each new tree fits the negative gradient of the loss given the current ensemble, scaled by a shrinkage rate, with row/column subsampling and L1/L2 leaf-weight penalties; the ensemble outputs a calibrated probability. Two disciplines distinguish S4's use from the reference: the model is trained *in-cycle* on the current cycle's own accumulated trades (no stale prior-cycle model), and it is admitted only if it clears a per-direction discriminability floor (AUC ≥ 0.55 long *and* short) and an adversarial leakage bound (adversarial AUC < 0.85). The noisier short-horizon microstructure of natural gas and copper typically yields weaker discriminability than crude or gold, which the per-direction floor screens honestly rather than papering over. Its entry threshold is itself a calibratable parameter.

9.2 Sentiment model — FinBERT

The sentiment channel is FinBERT [10], a BERT-family transformer fine-tuned on financial text. Each news item is tokenized and contextualized through stacked self-attention layers and mapped by a classification head to probabilities over {positive, neutral, negative}; domain fine-tuning is what lets it parse negation, hedging, and finance-specific jargon that generic sentiment models miss. Per asset per day the article-level scores are aggregated into a normalized polarity,

$$S_t = (1/N)\ \Sigma_i\ (\ P_{positive}^{(i)} - P_{negative}^{(i)}\ ),\quad S_t \in [-1, 1].$$

Following [9], S4 consumes S_t as a *risk filter and soft threshold skew, never as a standalone entry signal*: a sufficiently negative S_t (below a calibrated floor) suspends new entries and can flatten exposure, shielding core positions against adverse-news shocks—OPEC headlines, inventory surprises, weather-driven gas spikes, macro-policy actions—that purely technical logic is blind to. *Honesty note:* in a historical backtest without a time-aligned, point-in-time news feed for the evaluated window, this channel is run neutral under a signed no-feed exemption (honest-empty) rather than fabricating sentiment; the exemption is recorded in the cycle dossier, and train=serve is preserved by deploying the same neutral configuration that was calibrated.

9.3 The tandem and the regime channel

The third model is the market-regime classifier—the Λ-Vol router (Section 6) at the strategy level, and a regime label at the overlay level—which conditions both the gate and the exit geometry. The three combine as a tandem: the directional model proposes a probability, the sentiment channel can veto or attenuate it, and the regime label selects the operating envelope. This is the multi-modal fusion that [9] shows outperforms any single isolated predictive model across changing regimes. The decisive engineering choice is that all three are computed in-cycle from the same data the system trades—the overlay calibrated is the overlay deployed, so the gate cannot be a flattering in-sample artifact of a model the live system never runs.

10. Contract-Roll Discontinuities and Data Integrity

Commodity micros expire and roll—energy contracts (MCL, MNG) on a monthly cycle, metals (MGC, MHG) on the COMEX schedule—and any continuous series stitched across expiries is a *synthetic* construction. Back-adjustment introduces a price discontinuity at each roll boundary that no trader could have captured: a naive engine reads it as a phantom breakout or a phantom adverse excursion, and a naive drawdown accountant marks a roll gap as real loss. S4 treats this as a first-class data-integrity control, not an afterthought. Roll boundaries are located from each instrument's expiry calendar, and a bar-over-bar gap that exceeds a volatility-scaled threshold coincident with a roll date is flagged as a roll-stitch discontinuity and neutralized: no engine may fire on it, no exit leg may be triggered by it, and it is excluded from the mark-to-market drawdown path. The discipline is symmetric—the control removes a spurious edge as readily as a spurious loss—and it is explicitly *not* an alpha source. We flag two honest hazards: back-adjustment can leak look-ahead if applied with full-history knowledge, and aggressive gap-filling can erase genuine overnight moves; S4's neutralize-don't-trade-through rule is chosen precisely because it is the conservative side of both errors.

11. Calibration and Acceptance

Each configuration is accepted into a cycle only on a simultaneous gate—win rate (≥ 55%, net basis), gross profit factor (≥ 1.10), trade frequency (≥ 80 per year, an S4-specific override reflecting the higher base rate of commodity intraday signals), drawdown (a soft, logged criterion with a floor that prevents the remediation loop from rescuing a config by truncating trade flow), and a commission-to-gross-profit bound (a hard block; Section 12)—evaluated on the hybrid-AI-gated system, with the hybrid-AI discriminability bounds (per-direction AUC ≥ 0.55, adversarial AUC < 0.85) enforced separately. Cycle eligibility additionally requires full 16/16 coverage with no blocked configuration, positive aggregate result, and in-cycle model freshness. Promotion toward live deployment requires consecutive consistent cycles with stable parameters, specification-to-implementation parity across Pine, Python, and NinjaScript C# within tolerance (trade count ±2%, net P&L ±5%), an authoritative TradingView backtest match, and a paper-trade qualification period—the governance described in the program's pipeline whitepaper.

12. Risk and Venue

Risk is enforced preventively at the sizing layer: positions are sized so that a worst-case adverse move, measured from the monotonically ratcheting equity peak, cannot carry the account past the hard drawdown ceiling (a structural 24% constant identical across the specification, the calibrator, and the executor). Because commodity micros gap across the overnight and weekend session breaks, each instrument carries a per-instrument *gap reserve*—calibrated from its own historical gap distribution, largest for natural gas—that the preventive sizer subtracts from available headroom so that a gap, not merely an intrabar move, cannot breach the ceiling. Order-flow-aware leverage attenuation (Section 5) further reduces self-impact on the thin books. The execution venue is CME via NinjaTrader; commissions use a conservative flat dollar-per-contract round-trip schedule that varies materially across the four micros (crude the highest, gold the lowest), applied per trade in the optimizer objective, with a commission-to-gross bound that fails closed when commissions consume more than 40% of gross profit on a net-losing config—the per-configuration contract cap is the lever calibration uses to hold that bound on the short-horizon crude and gas configs where it concentrates.

13. Limitations

The order-flow results that motivate the leverage map are strongest with a live depth feed; under a bar-derived order-flow proxy the depth normalization is approximate, and the architecture treats the live-depth path as an enhancement, not a precondition. The cross-asset network-momentum literature [4, 5] indicates that a learned energy/metals spillover graph—crude, gold, and copper are linked through the inflation channel—supplies orthogonal alpha and is the principled replacement for any fixed cross-asset heuristic; this is a roadmap item, not a present claim. The roll-integrity control is conservative by construction and will occasionally neutralize a bar that carried genuine information; the regime classifier is deterministic and can saturate, most visibly on natural gas; the hybrid-AI gate constrains but does not certify causal or durable structure; and execution, liquidity, leverage, gap, and venue/regulatory risks are not eliminated by architectural integrity. As with every Q7 deliverable, no realized-performance representation is made, and the acceptance criteria above are calibration gates, not outcomes.

References

[1] R. Cont, A. Kukanov, S. Stoikov. *The Price Impact of Order Book Events.* 2011.

[2] R. Cont, A. de Larrard. *Price Dynamics in a Markovian Limit Order Market.* 2011.

[3] M. Gould et al. *Limit Order Books* (survey). Quantitative Finance, 2013.

[4] *Follow the Leader: Enhancing Systematic Trend-Following Using Network Momentum.*

[5] *Network Momentum across Asset Classes.*

[6] *A Primer on the Signature Method in Machine Learning.*

[7] R. Cont, J. da Fonseca. *Dynamics of Implied Volatility Surfaces.* Quantitative Finance, 2002.

[8] *Quantum-Probability-Theoretic Asset Return Modeling* (multi-modal return distributions at high-volume regimes). 2024–2025.

[9] V. N. K. Pillai, A. Ajith, S. K. J. *Generating Alpha: A Hybrid AI-Driven Trading System Integrating Technical Analysis, Machine Learning and Financial Sentiment for Regime-Adaptive Equity Strategies.* ComSIA 2026 (Springer LNNS).

[10] D. Araci. *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.* 2019.

Methodology disclaimer. This document describes a research and trading-system architecture. It is not investment advice, an offer, or a solicitation. Nothing herein represents realized trading results or guarantees future performance; calibration acceptance criteria are gates, not outcomes, and any backtested figures are research artifacts subject to the limitations inherent in historical simulation. Material intended for public distribution should be reviewed by qualified counsel prior to publication.