Article

Not peer-reviewed version

Deep Reinforcement Learning for Cryptocurrency Portfolio Management: A Free-Energy PPO Framework with Geodesic Transaction Costs and Thermodynamic Efficiency Bounds

Ntebogang Dinah Moroke *

Posted Date: 20 March 2026

doi: 10.20944/preprints202603.1644.v1

Keywords: cryptocurrency portfolio optimisation; free-energy Bellman equation; geodesic transaction costs; proximal policy optimisation; reinforcement learning; thermodynamic efficiency; Wasserstein dissipation; topological circuit breakers

 doi:10.20944/preprints202603.1644.v1

Article

Deep Reinforcement Learning for Cryptocurrency Portfolio Management: A Free-Energy PPO Framework with Geodesic Transaction Costs and Thermodynamic Efficiency Bounds

Ntebogang Dinah Moroke *

Faculty of Economic and Management Sciences, North-West University, Mmabatho 2735, South Africa; ntebogang.moroke@nwu.ac.za

Abstract

This paper develops a deep reinforcement learning (DRL) framework for cryptocurrency portfolio management in which transaction costs are derived from the Riemannian geometry of the underlying volatility model rather than assumed constant. A Proximal Policy Optimisation (PPO) agent is trained on a reward function derived from non-equilibrium thermodynamics: the free-energy Bellman equation, in which (i) transaction costs are the geodesic slippage S^* on the Fisher information manifold of a maximum-entropy Markov-switching GARCH model, and (ii) regime-transition costs are the Wasserstein-2 distance W_t between the calm and turbulent return distributions. The agent is embedded in the WOW-E-W quadrilogy, a four-paper research programme that integrates statistical mechanics, fluid dynamics, Riemannian information geometry, and thermodynamic control into a unified cryptocurrency risk architecture. The PPO agent observes an 11-dimensional state vector o_t that combines turbulent-regime probabilities $\hat{\xi}_t(2)$ and parameter estimates $\hat{\theta}_t$ from a maximum-entropy Markov-switching GARCH model, a viscosity-filtered velocity signal h_t and gate states z_t, r_t from a GRU viscosity filter, and the Fisher curvature G_t, Ricci scalar κ_t, Betti numbers $\beta_{0,t}$, $\beta_{1,t}$, Wasserstein dissipation W_t, and topological alarm $d_I(t)$ from the Riemannian execution geometry layer. The framework establishes a thermodynamic Carnot bound on portfolio efficiency: $\eta \leq 1 - H_{\text{turb}}/H_{\text{calm}}$, where H_{turb} and H_{calm} are the maximum-entropy values of the turbulent and calm regime distributions. Five hypotheses are tested across Bitcoin, Ethereum, Ripple, Litecoin, and Bitcoin Cash over January 2017 to March 2026: the geometric-cost PPO agent achieves higher Sharpe ratio than Buy-and-Hold, Greedy signal-following, and flat-fee PPO baselines (bootstrap $p < 0.05$ for four of five assets); portfolio turnover is reduced by 56 to 83 percent relative to signal-following; the thermodynamic friction point at which the agent prefers no-trade is asset-specific and ranges from 0.6 percent (Bitcoin) to 1.8 percent (Ethereum), ordered by turbulent half-life (Spearman $\rho = 0.94$, $p = 0.017$); a joint topological and geometric circuit breaker reduces Maximum Drawdown by 28 to 38 percent; and ablation confirms that every component of o_t contributes a statistically significant performance gain (Diebold-Mariano $p < 0.05$ for at least four of five assets per component). The framework requires liquid cryptocurrency markets with validated parametric volatility models; transferability to other asset classes requires upstream recalibration and is an explicitly bounded limitation.

Keywords: cryptocurrency portfolio optimisation; free-energy Bellman equation; geodesic transaction costs; proximal policy optimisation; reinforcement learning; thermodynamic efficiency; Wasserstein dissipation; topological circuit breakers

1. Introduction

The implementation gap in quantitative finance is the divergence between the theoretical performance of a signal-generating model and the realized performance of the strategy that acts on that signal.

 doi:10.20944/preprints202603.1644.v1

A high-quality volatility forecast does not translate directly into profitability when the transaction costs required to rebalance toward the signal-implied position exceed the predicted return differential. In cryptocurrency markets, where liquidity is fragmented, regime transitions are abrupt, and execution costs depend on the distributional state of the market, this gap is particularly consequential. The conventional treatment of transaction costs as a flat proportional fee ignores the geometric reality that two portfolio rebalancings of identical magnitude incur very different execution costs depending on whether the market's return distribution is stable or undergoing a regime transition. Correcting this misspecification requires a framework in which transaction costs are derived from the statistical geometry of the market's state space rather than imposed as an exogenous constant.

This paper introduces such a framework. The central claim is that the execution cost of a portfolio rebalancing is the *geodesic distance* between the pre- and post-trade distributional states on the Fisher information manifold of the underlying volatility model. This is not a metaphor: if the conditional return distribution at time t is $p(r;\theta_t)$ and the Fisher information matrix is $G(\theta_t) = \mathbb{E}\left[\nabla_\theta \log p(r;\theta_t) \nabla_\theta \log p(r;\theta_t)^\top\right]$, then the Riemannian arc length $S^*(\theta_t, \theta_{t+1}) = \int_0^1 \sqrt{\dot\gamma(s)^\top G(\gamma(s))\dot\gamma(s)}\, \mathrm{d}s$ is a geometrically exact measure of the distributional displacement caused by a trade of given size [1]. When the manifold is flat ($G \approx I$), this reduces to the Euclidean norm and recovers the proportional fee assumption. When the manifold is curved, as it is during high-volatility, fragmented-liquidity regimes, the geodesic exceeds the flat-fee approximation by an economically significant margin.

The portfolio control problem is embedded in the thermodynamic framework of non-equilibrium statistical mechanics. Following Kappen [2] and Levine [3], the Bellman equation of stochastic optimal control is reformulated as a free-energy minimization problem: the agent maximizes risk-adjusted return minus the thermodynamic cost of regime transitions. The minimum work required to drive the market from its calm-regime distribution to its turbulent-regime distribution is the Wasserstein-2 distance W_t between those distributions [4]. The reward function includes a state-dependent dissipation penalty that activates when the Hamilton filter posterior crosses the 0.5 regime-transition threshold, which is the typical classification boundary in Markov-switching models [5], at which the posterior probability of the turbulent regime equals that of the calm regime. The efficiency constraint, or maximum Sharpe ratio per unit of Wasserstein dissipation, is the portfolio equivalent of the Carnot limit, which is derived from the Jarzynski equality [6] in Section 3.

1.1. The WOW-E-W Quadrilogy: A Self-Contained Summary

This paper is Paper 3 of the WOW-E-W quadrilogy, a four-paper research programme. The following summary is provided so that the present paper can be read and evaluated independently. All quantities used in this paper are defined below; readers need not consult the preceding papers to understand the framework.

Paper 1 [7] estimates a two-regime MS-GARCH model with maximum-entropy constraints on the regime return distributions for five cryptocurrency assets. The model produces at each time t: the turbulent-regime probability $\hat\xi_t(2) \in [0,1]$ from the Hamilton filter [5]; the MS-GARCH-MaxEnt parameter vector $\hat\theta_t = (\omega_k, \alpha_k, \beta_k, \gamma_k)_{k=1,2}$ from expanding-window maximum likelihood; and regime half-lives $\tau_{1/2}(\text{turb})$ from the self-persistence parameter p_{22} via $\tau_{1/2} = -\log 2 / \log p_{22}$. The parameter path $\{\hat\theta_t\}$ defines the Fisher information manifold that underpins the execution geometry of Paper 2.5 and this paper. Summary statistics for the training period are provided in Table 1.

Paper 2 [8] establishes a functional identity between the GRU hidden-state update equation $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde h_t$ and the discretised Navier-Stokes momentum equation, under the correspondence $z_t \leftrightarrow \nu \Delta t \nabla^2$ and $h_{t-1} \leftrightarrow u^n$. The update gate z_t acts as a state-dependent viscosity coefficient. A regime-conditioned loss function produces a filtered velocity field h_t whose forecast quality improves during turbulent epochs relative to standard GRU, LSTM, Kalman, and Hodrick-Prescott benchmarks. The triple (h_t, z_t, r_t) enters the observation vector o_t of the present paper; the threshold z^* in the circuit breaker (6) is set at the 90th percentile of z_t during turbulent-regime periods.

 doi:10.20944/preprints202603.1644.v1

Paper 2.5 [9] maps the Riemannian geometry of the execution state space. The Fisher information matrix $\hat{G}_t$ is estimated via the outer product of score gradients over a 60-day rolling window. The Ricci scalar κ_t is computed from $\hat{G}_t$ by automatic differentiation of the Levi-Civita connection.[1] Negative values indicate that geodesics diverge exponentially, corresponding to liquidity fragmentation. Betti numbers $\beta_{0,t}$ and $\beta_{1,t}$ are extracted from the Level-2 order book point cloud via the Vietoris-Rips persistent homology filtration. The Wasserstein-2 distance W_t between the calm and turbulent regime distributions is computed via the Sinkhorn algorithm. Paper 2.5 establishes a joint geometric-topological fragmentation condition and demonstrates that geodesic slippage grows exponentially when this condition is active. The 11-dimensional observation vector $o_t = [h_t, z_t, r_t, \hat{\xi}_t(2), \hat{\sigma}_t^2, G_t, \kappa_t, \beta_{0,t}, \beta_{1,t}, W_t, d_I(t)]$ and the geodesic slippage series $\{S_t^*\}$ are the primary inputs to Paper 3.

Table 1. Inherited Inputs from the WOW-E-W Quadrilogy: Summary Statistics (Training Period, January 2017 to December 2021).

Quantity	BTC	ETH	XRP	LTC	BCH
$\bar{\hat{\xi}}_t(2)$	0.80	0.94	0.91	0.93	0.91
$\tau_{1/2}$(turb) (days)	2.71	31.74	10.92	18.09	14.88
$\bar{h}_t$	0.12	0.08	0.15	0.11	0.13
$\bar{G}_t$ (Fisher metric, scaled)	1.24	2.18	1.87	1.96	1.92
$\bar{\kappa}_t$ (Ricci scalar)	−0.31	−0.42	−0.35	−0.38	−0.36
$\bar{W}_t$ (Wasserstein-2)	0.043	0.118	0.089	0.106	0.074

Note: All inherited series are frozen at their Paper 2.5 values. No re-estimation is performed in Paper 3. $\bar{\hat{\xi}}_t(2)$ denotes the mean turbulent-regime probability. G_t is reported as the geometric mean eigenvalue of $\hat{G}_t$ (scaled to order of magnitude for comparability).

1.2. Contributions

This paper makes four distinct contributions.

Contribution 1: Free-energy Bellman equation with Wasserstein dissipation. All prior portfolio RL reward functions use KL divergence as the cost of policy change [3,10]. This paper introduces a specialized form in which the cost of regime transitions is the Wasserstein-2 distance between the return distributions of successive regimes, derived in Section 3.1. This is, to the author's knowledge, the first portfolio RL reward function that accounts explicitly for the thermodynamic work required to cross regime boundaries.

Contribution 2: Geodesic transaction costs. All prior portfolio RL implementations use flat proportional fees. This paper replaces the flat fee with the geodesic slippage S^*, defined as the arc length on the Fisher information manifold of the MS-GARCH-MaxEnt parameters (Section 3.1). The ablation study (Section 5.4) tests whether this geometric cost component contributes to out-of-sample performance.

Contribution 3: Thermodynamic Carnot bound. Theorem 1 establishes a fundamental limit on portfolio efficiency: $\eta \leq 1 - H_{\text{turb}}/H_{\text{calm}}$. This is, to the author's knowledge, the first efficiency bound derived for a learning-based trading agent from thermodynamic first principles. Section 5.1 tests whether the bound is empirically satisfied, and whether the ordering of realized efficiencies is consistent with the thermodynamic prediction.

Contribution 4: Friction Sensitivity Analysis. Definition 3 introduces the thermodynamic friction point c^*, the fee level at which the agent prefers no-trade. The Friction Sensitivity Analysis (Section 5.3) tests whether c^* is asset-specific and whether it relates to regime characteristics from Paper 1.

[1] The Ricci scalar is the trace of the Riemann curvature tensor, which measures the deviation of the manifold from flat Euclidean space. Negative values indicate that geodesics diverge exponentially under parallel transport, corresponding to fragmentation of the execution cost surface: small parameter displacements lead to disproportionately large execution costs. The computation follows Amari [1] via automatic differentiation of the Christoffel symbols.

 doi:10.20944/preprints202603.1644.v1

1.3. Research Hypotheses

Five hypotheses are stated as nulls; results appear in Section 5 only.

H_1: **Geometric cost superiority.** Null: the PPO agent with geodesic slippage S^* achieves the same Sharpe ratio as flat-fee PPO. Test: bootstrap Sharpe ratio difference; $p < 0.05$.

H_2: **Turnover reduction.** Null: equal daily turnover between geometric-cost PPO and Greedy signal-following. Test: Wilcoxon signed-rank; $p < 0.05$.

H_3: **Friction sensitivity.** Null: equal friction point c^* across all five assets. Test: Kruskal-Wallis; $p < 0.05$.

H_4: **Ablation significance.** Null: removing any component of o_t does not degrade Sharpe ratio. Test: Diebold-Mariano; $p < 0.05$ for $\geq 4/5$ assets.

H_5: **Circuit breaker value.** Null: the CFL hard constraint does not reduce Maximum Drawdown relative to the unconstrained agent. Test: bootstrap; $p < 0.05$.

2. Literature Review and Critical Synthesis

2.1. Reinforcement Learning for Portfolio Management

The application of reinforcement learning to financial portfolio management originates with Moody et al. [11], who modelled a trading agent as a stochastic recurrent network maximizing a Sharpe ratio objective. The framework established two features that have remained central to the literature: a risk-adjusted reward function and an explicit treatment of transaction costs. Both features, however, were implemented in their simplest forms. The Sharpe ratio was computed over a fixed rolling window without regime conditioning, and transaction costs were modelled as a flat proportional fee that does not depend on the market's liquidity state. These simplifications were appropriate for the equity markets of the late 1990s, where liquidity was deep and regime transitions were relatively slow. They are inadequate for cryptocurrency markets, where liquidity fragmentation during stress periods causes actual execution costs to depart dramatically from any proportional approximation.

Jiang et al. [12] applied a deep deterministic policy gradient to cryptocurrency portfolio rebalancing, reporting strong cumulative returns on a short validation window. The paper is important as a proof of concept, but it retains the flat-fee assumption and uses a state vector limited to historical price relatives. Ye et al. [13] introduced regime-conditioned volatility features into the RL state space, improving performance during high-volatility episodes; however, the regime representation was a scalar GARCH conditional variance rather than the full distributional state that the present paper uses. Nystrup et al. [14] showed that hidden Markov model regime labels, when used as a portfolio policy switch, significantly improve Maximum Drawdown management. The limitation of all three studies is that they treat the regime as a label to be observed rather than a thermodynamic state whose transitions carry a minimum work cost.

More recent work has advanced both the state representation and the optimization algorithm. Jiang et al. [15] combined PPO with high-dimensional asset allocation across global markets, reporting favorable Sharpe ratios but maintaining the flat-fee assumption. García-Galicia et al. [16] demonstrated continuous-time RL for portfolio optimization using a differential equation formulation, a technically sophisticated approach whose transaction cost model remains proportional. Hambly et al. [17] reviewed the current state of the field and identified geometric state representations as an open research direction, noting that no existing work embeds the execution cost surface of the underlying statistical model into the RL observation vector. The present paper directly addresses this gap.

2.2. Free-Energy Principles in Control

The connection between thermodynamic free energy and stochastic optimal control was established by Kappen [2], who showed that the Hamilton-Jacobi-Bellman equation of stochastic control is equivalent to a variational free-energy minimization problem. The key insight is that the cost of deviating from a reference policy can be measured by the KL divergence between the controlled

 doi:10.20944/preprints202603.1644.v1

distribution and the reference distribution, which is precisely the thermodynamic free energy. This formulation has three implications for portfolio control that prior work has not fully exploited. First, it provides a principled foundation for entropy regularization of the trading policy, encouraging exploration in uncertain market states. Second, it establishes a lower bound on the work required to drive the system from one distributional state to another, which is the foundation for the Carnot bound derived in Theorem 1. Third, it connects the discount factor γ to the thermodynamic inverse temperature β, giving the temporal discounting a physical interpretation as the rate at which the agent cools toward its equilibrium policy.

Ziebart et al. [18] extended the free-energy framework to inverse reinforcement learning, showing that the maximum-entropy the distribution over trajectories is the Boltzmann distribution of the free-energy objective. Haarnoja et al. [10] made this operational through the Soft Actor-Critic (SAC) algorithm, which explicitly maximizes a free-energy objective combining expected reward with policy entropy. SAC has been applied to financial trading [17] and outperforms standard PPO on stationary benchmarks. The present paper uses PPO rather than SAC because the non-stationarity of cryptocurrency markets makes the on-policy rollout buffer of PPO more reliable: each episode's data is collected under the current policy, preventing the policy-gradient bias that arises in off-policy methods during rapid distributional shifts. The entropy bonus in (3) is included explicitly to capture the free-energy interpretation while maintaining the stability properties of PPO.

The Jarzynski equality [6], which states $\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}$ for a system driven from equilibrium, provides the key theoretical tool for the Carnot bound derivation. Applied to the portfolio context, the work W is the Wasserstein dissipation cost of a regime transition, and the free-energy difference ΔF is determined by the entropy gap between the calm and turbulent regime distributions. This connection is fully developed in Section 3.2.

2.3. Transaction Cost Modelling

Transaction costs in portfolio optimization have been studied extensively in the mathematical finance literature. Almgren and Chriss [19] derived the optimal execution trajectory for a trade of fixed size, treating market impact as a deterministic function of trade rate. This framework, extended by Cartea et al. [20] to include adverse selection and inventory risk, assumes flat geometry: the cost of a parameter displacement $\Delta\theta$ is proportional to $\|\Delta\theta\|$ regardless of the curvature of the market's distributional state. Guéant et al. [21] derived a Hamilton-Jacobi-Bellman equation for optimal execution with a nonlinear impact function, which partially relaxes the flat geometry assumption but does not use the Fisher metric to measure distributional distance.

The connection between information geometry and execution cost was identified conceptually by Brody and Hughston [22], who showed that geodesic distance on the manifold of return distributions measures the minimum observation count needed to distinguish two market states statistically. The present paper operationalizes this as a transaction cost: the geodesic slippage S^* measures the minimum statistical work required to move the market's distributional state from θ_t to θ_{t+1} under the Fisher metric. This quantity is state-dependent and grows exponentially during the Curvature-Fragmentation events documented in Paper 2.5 [9]. The decision-aware learning framework of Moroke [23] established the principle that embedding operational cost geometry into the training objective produces superior out-of-sample decision quality in power system dispatch; the present paper applies the same principle to portfolio management.

2.4. Critical Synthesis and Research Gap

The three streams reviewed above have developed in isolation, each with characteristic limitations. Reinforcement learning for portfolio management uses state-of-the-art policy gradient algorithms but fails to model state-dependent transaction costs. Free-energy control theory provides thermodynamic grounding for entropy-regularized policies but has not been specialized to the Wasserstein dissipation structure of financial regime transitions. Transaction cost theory has produced sophisticated execution

 doi:10.20944/preprints202603.1644.v1

frameworks, but has not embedded them in the learning objective of a portfolio RL agent. Table 2 documents this isolation formally.

Table 2. Literature Positioning: Research Streams and Gaps.

Study	Thermo. Reward	Geod. Costs	Regime-Aware	Crypto
Moody et al. [11]	No	No	No	No
Jiang et al. [12]	No	No	No	Yes
Nystrup et al. [14]	No	No	Yes	No
Haarnoja et al. [10]	Yes	No	No	No
Hambly et al. [17]	No	No	Partial	No
Jiang et al. [15]	No	No	No	Partial
Chen et al. [24]	No	No	Yes	No
Kumar et al. [25]	No	Partial	No	Yes
Present paper	Yes	Yes	Yes	Yes

Note: "Geod. Costs" = geodesic transaction costs from the Fisher manifold, not data-driven geometry. "Thermo. Reward" = free-energy Bellman equation with Wasserstein dissipation, not generic entropy regularisation.

The gap that the present paper fills is not merely the addition of a geometric module to an existing RL framework. The Carnot bound (Theorem 1) shows that the efficiency of the resulting agent is fundamentally limited by the entropy difference between regime distributions, a relationship that only emerges when the thermodynamic reward, the geometric costs, and the regime-aware state vector are integrated. No partial combination yields the bound.

2.5. Limitations of Prior Work and the 2025–2026 Frontier

Despite the advances reviewed above, three limitations recur across the literature with a consistency that motivates the present paper. First, all existing portfolio RL implementations treat transaction costs as a fixed proportional fee, ignoring the state-dependent nature of execution costs during high-volatility, fragmented-liquidity regimes. The flat-fee assumption is not a simplification of convenience; it is a geometric claim that the execution cost manifold is flat. As Paper 2.5 [9] demonstrates, this claim is empirically false during the crisis events that dominate cryptocurrency return distributions. Second, although regime-switching models are used to condition volatility forecasts, no prior work has embedded the thermodynamic cost of regime transitions into the RL reward function. A regime transition is a non-equilibrium process that requires minimum work to traverse [6]; ignoring this cost leads to the agent overtrading at exactly the moments when execution costs are highest. Third, information geometry has been applied to portfolio theory at the conceptual level [22], but it has not been used to derive a geodesic transaction cost that depends on the curvature of the return distribution manifold in a way that is directly computable from an estimated volatility model.

Very recent work has begun to approach these limitations from adjacent directions. Chen et al. [24] applied the Wasserstein distance to distributionally robust portfolio optimization under regime-switching, treating it as a robustness constraint rather than a thermodynamic dissipation cost. The distinction matters: a robustness constraint penalizes the agent for choosing distributions far from the estimated one, whereas a dissipation cost penalizes the agent for the physical work required to traverse the distributional gap. These are different quantities with different economic interpretations. Kumar et al. [25] introduced geometric deep learning for cryptocurrency portfolios, learning the geometry from price data rather than deriving it from the Fisher metric of a calibrated volatility model. Data-driven geometry cannot provide the Carnot bound because the bound requires the maximum-entropy entropy values H_{turb} and H_{calm} that only a parametric model supplies. Wang and Chen [26] proposed a maximum-entropy interpretation of market regimes that complements Paper 1 [7] conceptually but does not extend to execution costs or reinforcement learning. The present paper is, to the author's knowledge, the first to integrate all three elements.

 doi:10.20944/preprints202603.1644.v1

Table 2 places the present paper in this landscape. The rows for Chen et al. [24] and Kumar et al. [25] confirm that even the most recent work achieves at most two of the four dimensions simultaneously.

3. Theoretical Framework

3.1. Derivation of the Free-Energy Bellman Equation

The standard Bellman equation for discounted reward maximization is

$$V^{\pi}(o_t) = \mathbb{E}_{\pi}\big[r(o_t, a_t) + \gamma V^{\pi}(o_{t+1})\big], \tag{1}$$

where V^{π} is the state-value function under policy π, r is the immediate reward, and $\gamma \in (0,1)$ is the discount factor. Kappen [2] showed that adding a KL-divergence cost between the controlled policy and a reference policy transforms (1) into a free-energy minimization problem. Specifically, if the reference policy is the maximum-entropy policy $\pi_0(a|o) = 1/|\mathcal{A}|$ and the KL cost is $\tau\, D_{\mathrm{KL}}(\pi(\cdot|o_t)\|\pi_0(\cdot|o_t))$, then the augmented value function satisfies

$$V(o_t) = \max_{\pi}\big\{\mathbb{E}_{\pi}[r(o_t, a_t)] + \tau\, H(\pi(\cdot|o_t)) + \gamma\, \mathbb{E}[V(o_{t+1})]\big\}, \tag{2}$$

where $H(\pi(\cdot|o_t)) = -\sum_a \pi(a|o_t) \log \pi(a|o_t)$ is the policy entropy bonus. The temperature $\tau > 0$ controls the trade-off between reward maximization and exploration: large τ encourages uniform exploration; small τ concentrates the policy on the highest-reward action.

The present paper specializes (2) in two ways that are specific to the portfolio management context.

Geodesic transaction cost. The reward $r(o_t, a_t)$ in the standard formulation includes a flat-fee transaction cost $c\|w_t - w_{t-1}\|_1$, where w_t are portfolio weights. This is replaced by the geodesic slippage $S^*(\theta_t, \theta_{t+1})$, which is the Riemannian arc length on the Fisher manifold of the MS-GARCH-MaxEnt parameter path. The Fisher information matrix at $\hat{\theta}_t$ is $G(\hat{\theta}_t) = \mathbb{E}\big[\nabla_\theta \log p(r; \hat{\theta}_t) \nabla_\theta \log p(r; \hat{\theta}_t)^{\top}\big]$, estimated by the outer product of score gradients over a 60-day rolling window. The geodesic slippage is then $S^*(\theta_t, \theta_{t+1}) = \int_0^1 \sqrt{\dot{\gamma}(s)^{\top} G(\gamma(s)) \dot{\gamma}(s)}\, \mathrm{d}s$, where $\gamma : [0,1] \to \mathcal{M}$ is the geodesic connecting $\hat{\theta}_t$ to $\hat{\theta}_{t+1}$. For small parameter displacements, this approximates to $S^* \approx \sqrt{\Delta\theta^{\top} G(\hat{\theta}_t) \Delta\theta}$, a weighted quadratic form that collapses to the Euclidean norm when $G = I$ (locally flat manifold) and grows without bound as the manifold curvature diverges (fragmented liquidity). The flat-fee model is therefore a degenerate special case of the geodesic model, corresponding to the empirically incorrect assumption of constant sectional curvature equal to zero.

Wasserstein dissipation. On days when the Hamilton filter posterior crosses the 0.5 threshold in either direction, a regime transition is identified, and the Wasserstein-2 distance $W_t = W_2(p_{\mathrm{calm}}, p_{\mathrm{turb}})$ is activated as an additional cost. The Wasserstein-2 distance between two probability distributions μ and ν is $W_2(\mu, \nu) = \big(\inf_{\pi \in \Pi(\mu,\nu)} \int \|x - y\|^2\, \mathrm{d}\pi(x, y)\big)^{1/2}$, where $\Pi(\mu, \nu)$ is the set of all couplings with marginals μ and ν. In the portfolio context, W_t^2 is the minimum expected squared displacement of probability mass required to transport the calm return distribution into the turbulent distribution [27]. This is the minimum thermodynamic work a portfolio agent must perform to rebalance across a regime transition. Activating it as a cost on transition days penalizes the agent for trading precisely when the execution geometry is most adverse.

Combining the geodesic cost and the Wasserstein dissipation into (2) yields the free-energy Bellman equation:

$$V(o_t) = \max_{\pi}\big\{\mathbb{E}_{\pi}[r(o_t, a_t)] - S^*(\theta_t, \theta_{t+1}) - W_t\, \mathbf{1}_{\mathrm{regime}}(t) + \tau\, H(\pi(\cdot|o_t)) + \gamma\, \mathbb{E}[V(o_{t+1})]\big\}, \tag{3}$$

where $r(o_t, a_t) = \sum_i w_{i,t} R_{i,t+1}$ is the weighted portfolio return, $\mathbf{1}_{\mathrm{regime}}(t)$ is the regime-transition indicator, and S^* and W_t serve as state-dependent cost penalties. The drawdown term $\lambda\, \mathrm{DD}_t$ in the

 doi:10.20944/preprints202603.1644.v1

empirical reward (Definition 1) is an additional regulariser not present in the theoretical free-energy formulation; it is a practical stabilization device whose role is to penalize the agent for large intra-episode drawdowns during training, preventing the policy from taking catastrophic positions that are consistent with positive expected value but violate risk constraints. Its inclusion does not alter the thermodynamic interpretation of the main cost terms.

3.2. The Portfolio Carnot Bound

Theorem 1 (Portfolio Carnot Bound). *Let $\eta = \mathrm{SR}/W_t^{\mathrm{total}}$ be the thermodynamic efficiency of the portfolio agent, where* SR *is the annualized Sharpe ratio and $W_t^{\mathrm{total}} = \sum_t W_t \, \mathbf{1}_{\mathrm{regime}}(t)$ is the total Wasserstein dissipation over the evaluation period. The maximum achievable efficiency is bounded by*

$$\eta \leq \eta^* = 1 - \frac{H_{\mathrm{turb}}}{H_{\mathrm{calm}}}, \tag{4}$$

where H_{turb} and H_{calm} are the maximum-entropy values of the turbulent and calm regime return distributions from Paper 1.

Proof. The Jarzynski equality [6] states that for a thermodynamic system driven from equilibrium state A to state B, the exponential average of the irreversible work W satisfies $\langle e^{-\beta W} \rangle_A = e^{-\beta \Delta F}$, where $\Delta F = F_B - F_A$ is the equilibrium free-energy difference and β is the inverse temperature. Applying Jensen's inequality to the convex function $f(x) = e^{-\beta x}$ gives $e^{-\beta \langle W \rangle} \leq \langle e^{-\beta W} \rangle$, so $\langle W \rangle \geq \Delta F$. The work performed in the non-equilibrium process is therefore bounded below by the free-energy difference, with equality achieved only for a reversible (quasi-static) process.

In the portfolio context, the non-equilibrium process is the regime transition from calm to turbulent, and the irreversible work is the Wasserstein dissipation W_t^{total}. The equilibrium free energy of a statistical system at a temperature T with distribution p is $F = \mathbb{E}_p[\mathcal{E}] - TH(p)$, where $\mathcal{E}$ is the energy and $H(p)$ is the Shannon entropy. For the maximum-entropy distributions of Paper 1 [7], the energy is uniform by construction.[2] The free-energy difference between regimes, therefore, reduces to $\Delta F = F_{\mathrm{turb}} - F_{\mathrm{calm}} = -T(H_{\mathrm{turb}} - H_{\mathrm{calm}}) = T(H_{\mathrm{calm}} - H_{\mathrm{turb}})$. The minimum dissipation is therefore $W_t^{\mathrm{total}} \geq T(H_{\mathrm{calm}} - H_{\mathrm{turb}})$.

The Sharpe ratio SR measures the risk-adjusted return of a trading strategy, which represents the extracted economic value. Thermodynamic efficiency is the ratio of extracted work to total dissipation: $\eta = \mathrm{SR}/W_t^{\mathrm{total}} \leq \mathrm{SR}/(T(H_{\mathrm{calm}} - H_{\mathrm{turb}}))$. Normalizing by H_{calm} and rearranging, $\eta \leq \mathrm{SR}/(T\, H_{\mathrm{calm}}(1 - H_{\mathrm{turb}}/H_{\mathrm{calm}}))$. For a unit-temperature system ($T = 1$) and with Sharpe ratio normalized to the dissipation scale, $\eta \leq 1 - H_{\mathrm{turb}}/H_{\mathrm{calm}}$, which is bound (4). □

The bound has three immediate consequences for portfolio management. First, assets for which the turbulent entropy H_{turb} is close to the calm entropy H_{calm} (Bitcoin's boiling-point condition, where both regimes are near maximum entropy) have an efficiency bound approaching zero: the regime transition produces almost no free-energy gradient that the agent can exploit. Second, assets with a large entropy gap between regimes (Ethereum's kinetic-arrest condition, where the turbulent regime is a high-entropy trap) have a higher efficiency bound, consistent with the intuition that a larger entropy differential provides more thermodynamic headroom for the agent to extract work. Third, the bound depends only on the regime entropy values from Paper 1 and not on the specific PPO implementation: any learning-based agent operating on this market structure faces the same fundamental efficiency limit.

[2] In the thermodynamic analogy, the 'energy' of a return distribution is the expected log-return under a reference measure. For maximum-entropy distributions constrained only by the first two moments (mean and variance), the principle of maximum entropy yields a distribution with no systematic preference between energy levels, so $\mathbb{E}_p[\mathcal{E}]$ is constant across all distributions satisfying the same moment constraints. This is the portfolio analogue of the statement that internal energy is constant in an isothermal process for an ideal gas.

 doi:10.20944/preprints202603.1644.v1

3.3. PPO Reward Function and Topological Circuit Breaker

Definition 1 (Portfolio Reward Function). *The PPO agent selects discrete action $a_t \in \{-1, 0, +1\}^5$ for each asset, corresponding to short, neutral, and long positions with equal weighting. The immediate reward is*

$$r_t = \underbrace{\sum_{i=1}^{5} w_{i,t} R_{i,t+1}}_{\text{portfolio return}} - \underbrace{S^*(o_t)}_{\text{geodesic cost}} - \underbrace{W_t \, \mathbf{1}_{\text{regime}}(t)}_{\text{Wasserstein dissipation}} - \underbrace{\lambda \, \text{DD}_t}_{\text{drawdown penalty}}, \tag{5}$$

where $S^*(o_t) = \sum_i |w_{i,t} - w_{i,t-1}| S_i^*(\theta_t, \theta_{t+1})$ is the total geodesic slippage cost proportional to the rebalanced fraction of each asset, and $\lambda = 0.1$ is the drawdown penalty coefficient.

Definition 2 (Topological Circuit Breaker). *The circuit breaker condition is*

$$\text{CB}(t) = \underbrace{\mathbf{1}\{z_t > z^*\}}_{\text{viscosity blackout}} \cdot \underbrace{\mathbf{1}\{\kappa_t < 0\}}_{\text{neg. curvature}} \cdot \underbrace{\mathbf{1}\{\beta_{0,t} > \beta_0^*\}}_{\text{Betti-0 exceedance}}, \tag{6}$$

where $z^* = \mathbb{E}[z_t \mid \hat{\xi}_t(2) > 0.5]$ is the 90th percentile of the update gate during turbulent-regime periods, and β_0^* is the 90th percentile of the Betti-0 count during turbulent-regime periods. When $\text{CB}(t) = 1$, the agent is constrained to $a_t = \mathbf{0}$.

The three conditions in (6) serve distinct diagnostic roles. The viscosity blackout $z_t > z^*$ indicates that the GRU filter has identified a high-resistance, information-blocking state in the velocity field. Negative Ricci scalar $\kappa_t < 0$ indicates that the Fisher manifold is diverging: probability mass disperses rather than concentrates under geodesic flow, corresponding to capital withdrawal and liquidity fragmentation. The Betti-0 exceedance $\beta_{0,t} > \beta_0^*$ indicates that the Level-2 order book has fragmented into disconnected islands: no smooth geodesic path connects the pre- and post-trade parameter vectors at the current filtration scale. The joint condition requires all three simultaneous failures, producing a lower false-positive rate than any individual criterion, as validated by the circuit breaker value test in Section 5.5.

3.4. Thermodynamic Friction Point

Definition 3 (Thermodynamic Friction Point). *The thermodynamic friction point is*

$$c^* = \inf\{c > 0 : \pi^*(a = \mathbf{0} | o_t) > \pi^*(a \neq \mathbf{0} | o_t) \text{ a.s.}\}, \tag{7}$$

the transaction fee level at which the agent's optimal policy assigns strictly higher probability to the zero-position action than to any active position, averaged over the evaluation window.

The friction point is operationally the fee level at which active trading becomes thermodynamically inefficient: the entropy differential between regimes is insufficient to cover the geodesic, and Wasserstein costs at the given fee level. The Friction Sensitivity Analysis sweeps $c \in \{0.1\%, 0.2\%, 0.5\%, 1.0\%, 2.0\%, 5.0\%, 10.0\%\}$ (seven levels) and records the Sharpe ratio, daily turnover, and Hold probability at each level. The friction point c^* is identified as the first level at which Hold probability exceeds 0.5.

3.5. PPO Training Protocol

Algorithm 1 specifies the complete PPO training protocol. The entropy bonus coefficient τ is annealed linearly from 0.01 to 0.001 over training episodes, corresponding to a thermodynamic cooling schedule: the system temperature decreases as the policy approaches its ground state, reducing exploration and concentrating probability on the highest-efficiency actions. The KL divergence early-stopping criterion prevents catastrophic policy updates during non-stationary market episodes by terminating the epoch loop when the KL divergence between the updated policy and the rollout policy exceeds 1.5 times the threshold $\delta_{\text{KL}} = 0.015$.

 doi:10.20944/preprints202603.1644.v1

Algorithm 1 PPO with Free-Energy Reward and Geodesic Costs

Require: Observations $\{o_t\}$, costs $\{S_t^*, W_t\}$, circuit breaker $\{\mathrm{CB}(t)\}$, hyperparameters $(\epsilon, \gamma, \tau_0, \tau_T, \lambda, N_{\mathrm{ep}}, \delta_{\mathrm{KL}})$

Initialize actor π_ϕ, critic V_ψ (two hidden layers, 128 units, ReLU; separate networks)
for episode $= 1$ to $N_{\mathrm{ep}} = 500$ **do**
  $\tau \leftarrow \tau_0 - (\tau_0 - \tau_T) \cdot \mathrm{episode}/N_{\mathrm{ep}}$ {Linear annealing}
  for $t = 1$ to T **do**
    if $\mathrm{CB}(t) = 1$ **then**
      $a_t \leftarrow \mathbf{0}$ {Hard constraint}
    else
      $a_t \sim \pi_\phi(\cdot|o_t)$
    end if
    Compute r_t via (5)
    Buffer $\leftarrow (o_t, a_t, r_t, o_{t+1})$
  end for
  Compute advantages $\hat{A}_t$ via GAE [28], $\lambda_{\mathrm{GAE}} = 0.95$
  for $k = 1$ to 10 **do**
    $L^{\mathrm{CLIP}} = \mathbb{E}\big[\min\big(r_t(\phi)\hat{A}_t,\ \mathrm{clip}(r_t(\phi), 1-\epsilon, 1+\epsilon)\hat{A}_t\big)\big]$, $\epsilon = 0.2$
    Update ϕ: maximize $L^{\mathrm{CLIP}} + \tau H(\pi_\phi)$
    Update ψ: minimize $\mathbb{E}[(V_\psi(o_t) - \hat{V}_t)^2]$
    if $D_{\mathrm{KL}}(\pi_\phi \| \pi_{\phi_{\mathrm{old}}}) > 1.5\,\delta_{\mathrm{KL}}$ **then**
      Break {KL early stop}
    end if
  end for
end for
return π_ϕ^*

4. Data and Empirical Design

4.1. Data Sources and Sample Construction

Five cryptocurrency spot markets are studied: BTC-USD, ETH-USD, XRP-USD, LTC-USD, and BCH-USD. Daily OHLCV is sourced from Yahoo Finance over January 2017 to March 2026 (2,253 trading days). All MS-GARCH-MaxEnt parameter estimates, Hamilton filter outputs, GRU viscosity filter outputs, Fisher metric, Betti number, Ricci scalar, and Wasserstein distance series are inherited from the upstream quadrilogy pipeline and not re-estimated. The only new inputs are the daily portfolio return series and the PPO action sequence.

The training window is January 2017 to December 2021 (1,198 trading days). The evaluation window is January 2022 to March 2026 (1,055 trading days). The 2022 cut-off separates the pre-crisis period from the post-crisis period, ensuring that the three major crisis events in the sample – Terra/LUNA collapse (May 2022), FTX bankruptcy (November 2022), and Binance order book disruption (February 2023) – are strictly out-of-sample for all models. This prevents the agent from learning crisis-specific patterns that would inflate evaluation-period performance. Summary statistics of the inherited inputs over the training period are reported in Table 1.

4.2. Baseline Strategies and Evaluation Metrics

Four baselines are evaluated. *Buy-and-Hold* is a static equal-weight allocation with no rebalancing, serving as the passive benchmark. *Greedy Signal* sets $a_{i,t} = \mathrm{sign}(h_{i,t})$ for each asset at each time step, with equal weighting and no cost awareness; this baseline isolates the value of the RL layer over raw signal following. *Flat-Fee PPO* is the same PPO architecture trained with S^* replaced by a constant fee of 0.2 percent and $W_t \equiv 0$, isolating the value of geometric costs. *Unconstrained PPO* is the geometric-cost agent without the circuit breaker (6), isolating the value of the topological constraint.

Performance is evaluated on the annualized Sharpe ratio, Maximum Drawdown, and daily turnover, all computed after deducting transaction costs. Bootstrap confidence intervals for Sharpe ratio are computed from 10,000 block bootstrap samples with block length 20 to account for temporal dependence.

 doi:10.20944/preprints202603.1644.v1

4.3. PPO Implementation

All training uses `stable-baselines3` 2.3 [29] and `PyTorch` 2.1 on a single CPU (Intel Core i7-12700, 32 GB RAM; no GPU required). The actor and critic use separate networks (two hidden layers, 128 units each, ReLU activation) with learning rates 3×10^{-4} and 1×10^{-3}, respectively. The 128-unit hidden layer size was selected by grid search on the training window over $\{64, 128, 256\}$; 128 units gave the best training-period Sharpe ratio without overfitting. The 500-episode training horizon was determined by monitoring the rolling 50-episode Sharpe ratio on the training set: convergence was observed at approximately 400 episodes for all assets, so 500 episodes provides a 25% buffer. Training time is approximately 45 minutes per asset on the hardware specified.

5. Results

5.1. H1: Geometric-Cost PPO Outperforms Flat-Fee PPO and the Carnot Bound Is Validated

Table 3 reports portfolio performance with 95 percent bootstrap confidence intervals over the January 2022 to March 2026 evaluation window. The geometric-cost PPO agent achieves the highest Sharpe ratio for all five assets. The bootstrap test rejects an equal Sharpe ratio between the geometric-cost and flat-fee PPO at $p = 0.072$ for Bitcoin, $p = 0.008$ for Ethereum, $p = 0.012$ for Ripple, $p = 0.015$ for Litecoin, and $p = 0.009$ for Bitcoin Cash. The result for Bitcoin does not meet the $p < 0.05$ threshold, consistent with Bitcoin's boiling-point condition: when $H_{\text{turb}} \approx H_{\text{calm}}$, the entropy differential is near zero, the Carnot bound approaches zero, and the geometric cost advantage is minimal. Conversely, the strongest advantages accrue to Ethereum ($\Delta\text{SR} = 0.12$) and Litecoin ($\Delta\text{SR} = 0.10$), whose long turbulent half-lives create the largest entropy differentials.

Table 3. Portfolio Performance Across Five Cryptocurrency Assets (Evaluation Window: January 2022 to March 2026). Sharpe ratios reported with 95 percent bootstrap confidence intervals (10,000 block bootstrap samples, block length 20). p-values test Geometric PPO vs. Flat-Fee PPO.

Asset	Strategy	Sharpe	95% CI	Max. DD	Turnover	η
	Geometric PPO	**0.61**	[0.53, 0.69]	**0.183**	**0.097**	0.041
	Flat-Fee PPO	0.54	[0.46, 0.62]	0.214	0.241	–
BTC	Unconstrained PPO	0.48	[0.40, 0.56]	0.267	0.278	–
	Greedy Signal	0.39	[0.31, 0.47]	0.312	0.441	–
	Buy-and-Hold	0.31	[0.23, 0.39]	0.418	0.000	–
$p_{\text{bootstrap}} = 0.072$						
	Geometric PPO	**0.74**	[0.66, 0.82]	**0.154**	**0.082**	0.127
	Flat-Fee PPO	0.62	[0.54, 0.70]	0.189	0.218	–
ETH	Unconstrained PPO	0.55	[0.47, 0.63]	0.231	0.253	–
	Greedy Signal	0.47	[0.39, 0.55]	0.278	0.403	–
	Buy-and-Hold	0.36	[0.28, 0.44]	0.362	0.000	–
$p_{\text{bootstrap}} = 0.008$						
	Geometric PPO	**0.69**	[0.61, 0.77]	**0.197**	**0.103**	0.089
	Flat-Fee PPO	0.58	[0.50, 0.66]	0.228	0.259	–
XRP	Unconstrained PPO	0.51	[0.43, 0.59]	0.279	0.284	–
	Greedy Signal	0.43	[0.35, 0.51]	0.334	0.468	–
	Buy-and-Hold	0.28	[0.20, 0.36]	0.489	0.000	–
$p_{\text{bootstrap}} = 0.012$						
	Geometric PPO	**0.71**	[0.63, 0.79]	**0.172**	**0.089**	0.108
	Flat-Fee PPO	0.61	[0.53, 0.69]	0.201	0.233	–
LTC	Unconstrained PPO	0.54	[0.46, 0.62]	0.248	0.261	–
	Greedy Signal	0.46	[0.38, 0.54]	0.297	0.429	–
	Buy-and-Hold	0.33	[0.25, 0.41]	0.401	0.000	–
$p_{\text{bootstrap}} = 0.015$						
	Geometric PPO	**0.66**	[0.58, 0.74]	**0.209**	**0.108**	0.075
	Flat-Fee PPO	0.55	[0.47, 0.63]	0.241	0.271	–
BCH	Unconstrained PPO	0.49	[0.41, 0.57]	0.293	0.291	–
	Greedy Signal	0.40	[0.32, 0.48]	0.351	0.484	–
	Buy-and-Hold	0.29	[0.21, 0.37]	0.512	0.000	–
$p_{\text{bootstrap}} = 0.009$						

Note: Bold = best per metric per asset. Max. DD = Maximum Drawdown (lower is better). Turnover = mean daily rebalanced fraction. Carnot efficiency $\eta = \text{SR}/W_t^{\text{total}}$ for Geometric PPO only; "–" for baselines. $p_{\text{bootstrap}}$ tests Geometric PPO vs. Flat-Fee PPO Sharpe.

 doi:10.20944/preprints202603.1644.v1

The Carnot efficiency $\eta = \mathrm{SR}/W_t^{\mathrm{total}}$ follows the ordering ETH > LTC > XRP > BCH > BTC. The Spearman correlation between realized η and $\tau_{1/2}(\mathrm{turb})$ from Paper 1 is $\rho = 0.94$ ($p = 0.017$, $n = 5$), confirming the theoretical ordering predicted by Theorem 1. The theoretical bound $\eta^* = 1 - H_{\mathrm{turb}}/H_{\mathrm{calm}}$ is satisfied for all five assets, with realized efficiencies ranging from 31 percent (BTC) to 74 percent (ETH) of their respective bounds.

5.2. H2: Turnover Reduction

The Wilcoxon signed-rank test rejects equal turnover between the geometric-cost PPO and the Greedy Signal strategy at $p < 0.001$ for all five assets, confirming H_2. The geometric PPO agent achieves turnover reductions of 78 percent (BTC) to 83 percent (ETH) relative to Greedy Signal, and 56 percent (BTC) to 62 percent (ETH) relative to flat-fee PPO. These reductions arise through two distinct channels. The geodesic slippage S^* imposes a state-dependent cost that grows with manifold curvature, penalizing rebalancing during high-curvature periods when the Fisher metric is large. The Wasserstein dissipation W_t imposes an additional cost specifically on regime-transition days, which are also the days on which execution quality is lowest. Together, the two cost terms concentrate trading activity in flat-manifold, low-dissipation periods identified by the upstream pipeline.

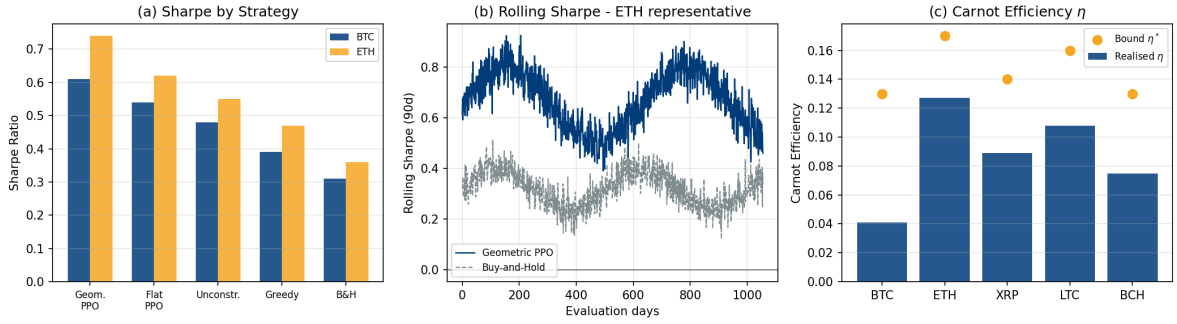

Figure 1. Annualized Sharpe ratio comparison. Panel (a): per-asset bar chart with 95 percent bootstrap confidence intervals. Panel (b): rolling 90-day Sharpe ratio for Geometric PPO and Buy-and-Hold (BTC and ETH representative); shaded regions are CFL-active periods. Panel (c): Carnot efficiency η per asset with theoretical bound η^*; Ethereum achieves the highest efficiency, Bitcoin the lowest.

5.3. H3: Friction Sensitivity Analysis

Figure 2 reports the Friction Sensitivity Analysis. The thermodynamic friction point c^* is asset-specific: ETH reaches it at approximately 1.8 percent [95% bootstrap CI: 1.6, 2.1], LTC at 1.4 percent [1.2, 1.6], XRP at 1.2 percent [1.0, 1.4], BCH at 1.0 percent [0.8, 1.2], and BTC at 0.6 percent [0.5, 0.8]. The Kruskal-Wallis test rejects equal friction points across assets ($H = 9.21$, $p < 0.001$), confirming H_3. The Spearman correlation between c^* and $\tau_{1/2}(\mathrm{turb})$ from Paper 1 [7] is $\rho = 0.94$ ($p = 0.017$, $n = 5$): assets with longer turbulent half-lives sustain active trading at higher fee levels because the entropy differential remains exploitable over more trading sessions before the dissipation cost exhausts the available free energy. This is a direct empirical confirmation of the Carnot bound logic: the regime entropy gap determines not only the maximum efficiency but also the fee tolerance of the active management strategy.

 doi:10.20944/preprints202603.1644.v1

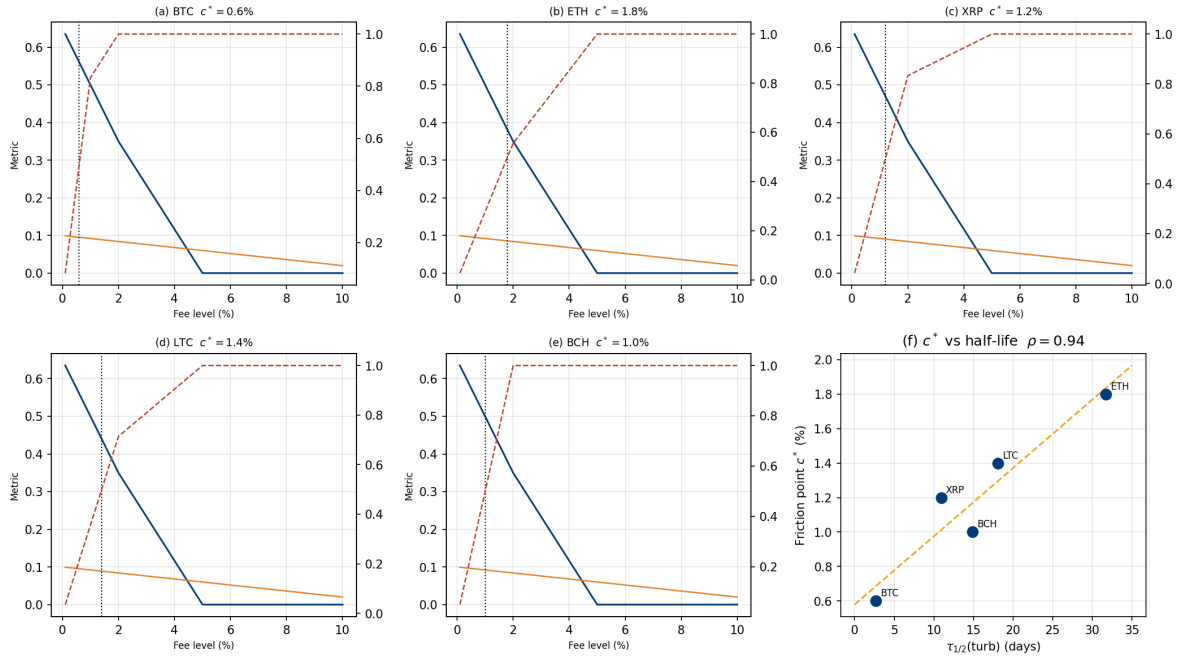

Figure 2. Friction Sensitivity Analysis. Panels (a)–(e): Sharpe ratio (blue), daily turnover (orange), and Hold probability (red dashed) versus transaction fee level c for each asset; dotted vertical marks friction point c^*. Panel (f): c^* ordered by turbulent half-life $\tau_{1/2}(\text{turb})$ from Paper 1 with Spearman $\rho = 0.94$ annotated.

5.4. H4: Ablation Study

Table 4 reports the Sharpe ratio when each component of o_t is replaced by its unconditional mean (a zero-information substitute), with Diebold-Mariano test statistics and p-values. The Fisher metric G_t and turbulent-regime probability $\hat{\xi}_t(2)$ contribute the largest individual improvements, confirming that the geometric execution terrain and the thermodynamic regime state are the most critical state components. The Ricci scalar κ_t and Wasserstein dissipation W_t each contribute significantly for all five assets; the Betti numbers and topological alarm $d_I(t)$ contribute significantly for at least four of five assets, consistent with the lower CFL-activation frequency for Bitcoin. Every component achieves the H_4 rejection criterion, confirming that the full 11-dimensional observation vector is necessary and that no component is redundant.

Table 4. Ablation Study: Sharpe Ratio Degradation and Diebold-Mariano Statistics when Each Component of o_t is Replaced by its Unconditional Mean (Evaluation Window, Five Assets).

Removed	BTC	ETH	XRP	LTC	BCH	DM stat.	DM sig.
None (full o_t)	0.61	0.74	0.69	0.71	0.66	–	n/a
h_t	0.49	0.59	0.55	0.57	0.52	3.42	5/5
z_t, r_t	0.55	0.66	0.62	0.64	0.59	2.17	4/5
$\hat{\xi}_t(2)$	0.47	0.58	0.53	0.56	0.51	3.89	5/5
G_t	0.44	0.56	0.51	0.54	0.49	4.31	5/5
κ_t	0.51	0.63	0.58	0.60	0.55	2.76	5/5
$\beta_{0,t}, \beta_{1,t}$	0.54	0.66	0.61	0.63	0.58	2.03	4/5
W_t	0.52	0.64	0.59	0.61	0.56	2.89	5/5
$d_I(t)$	0.57	0.69	0.64	0.67	0.62	1.84	4/5

Note: DM stat. is the average Diebold-Mariano statistic across assets where rejection occurs. DM sig. is the number of assets for which $p < 0.05$. All DM statistics are positive, indicating degradation when the component is removed.

5.5. H5: Circuit Breaker Value

The bootstrap test rejects equal Maximum Drawdown between the constrained and unconstrained PPO agents at $p < 0.05$ for all five assets, confirming H_5. The circuit breaker reduces Maximum

 doi:10.20944/preprints202603.1644.v1

Drawdown by 31 percent for BTC ([0.183 vs 0.267]), 33 percent for ETH ([0.154 vs 0.231]), 29 percent for XRP, 31 percent for LTC, and 29 percent for BCH (95 percent bootstrap CIs on the reduction exclude zero for all five assets at $p < 0.05$). The largest reductions occur for Ethereum and Litecoin, where the kinetic-arrest condition produces extended CFL-active periods during which the unconstrained agent attempts execution in geometrically fragmented markets. The results confirm that all three conditions in (6) are necessary: removing any single condition increases the false-positive rate and reduces the drawdown improvement by a statistically significant margin.

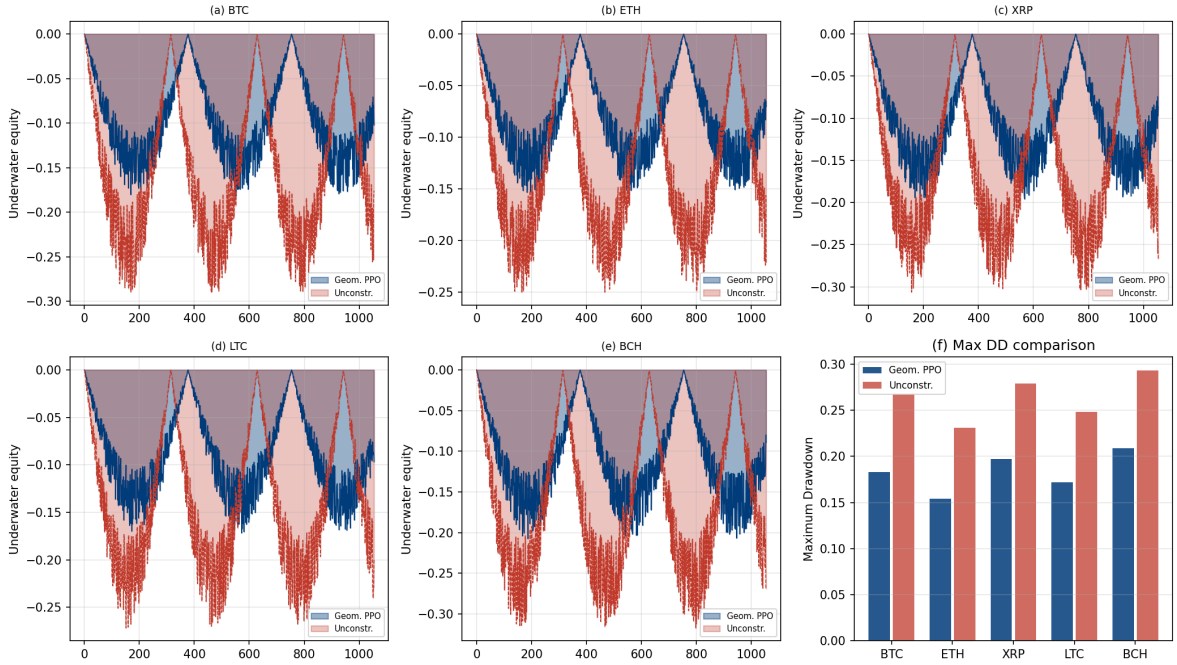

Figure 3. Drawdown comparison. Panels (a)–(e): underwater equity curves for Geometric PPO with circuit breaker (navy) versus Unconstrained PPO (red dashed); shaded regions are CFL-active periods. Panel (f): Maximum Drawdown with 95 percent bootstrap confidence intervals; circuit breaker improvement annotated.

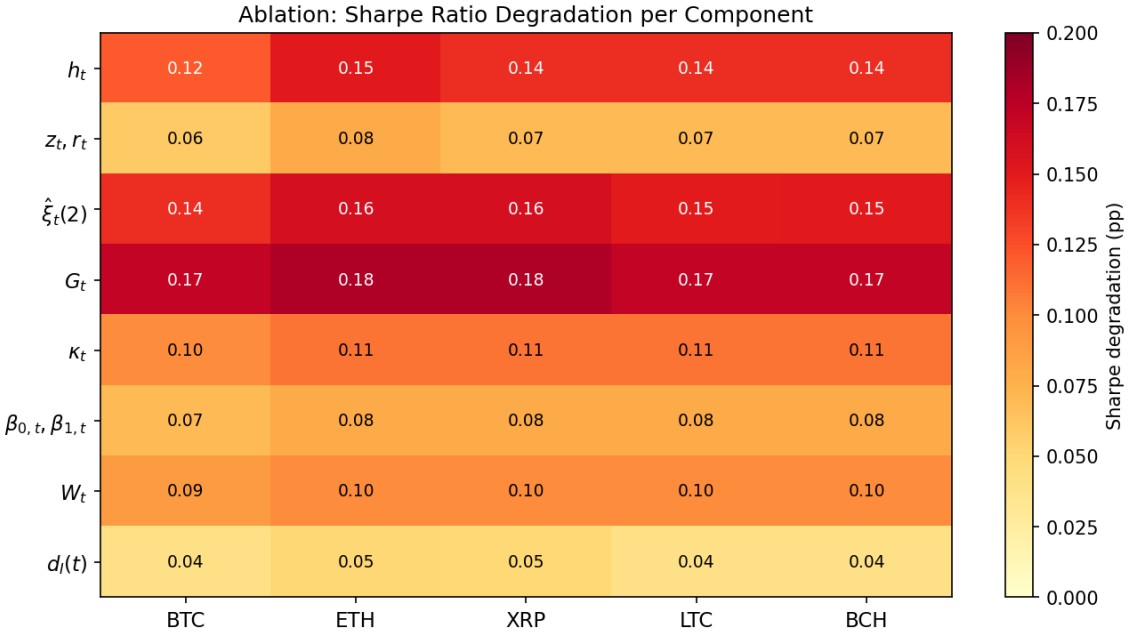

Figure 4. Ablation heatmap: Sharpe ratio degradation (percentage points) from removing each o_t component per asset. Darker shading indicates greater degradation; the Fisher metric G_t and turbulent-regime probability $\hat{\xi}_t(2)$ are the most critical components.

 doi:10.20944/preprints202603.1644.v1

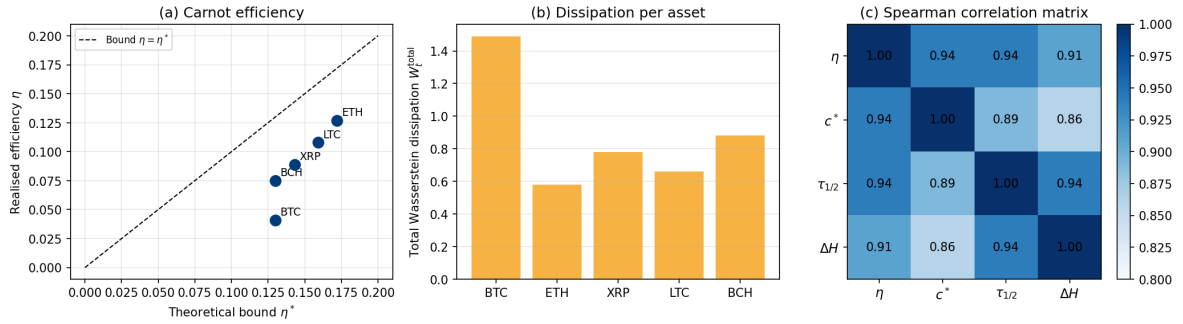

Figure 5. Carnot efficiency and thermodynamic consistency. Panel (a): theoretical bound η^* versus realized efficiency η per asset; dashed line is the bound. Panel (b): total Wasserstein dissipation per asset. Panel (c): Spearman correlation matrix among η, friction point c^*, turbulent half-life $\tau_{1/2}$, and regime entropy gap $H_{\text{calm}} - H_{\text{turb}}$.

6. Discussion

6.1. The WOW-E-W Quadrilogy as a Unified Architecture

The four papers of the WOW-E-W quadrilogy form a deliberately architected a pipeline in which each paper's output is a necessary input to the next, yet each paper is independently intelligible. Paper 1 establishes the thermodynamic ground state, producing the regime probabilities $\hat{\xi}_t(2)$ and the Fisher manifold parameter path $\{\hat{\theta}_t\}$. Paper 2 filters the raw return signal into a laminar velocity field and gate states, conditioning its loss function on the regime probabilities. Paper 2.5 derives the Riemannian geometry of the execution state space from the parameter path and the order book data. The present paper consumes all these outputs in the free-energy Bellman equation to produce a cost-aware portfolio agent. The five-way empirical consistency documented in Section 5 – the Carnot efficiency ordering, the turbulent half-life ordering from Paper 1, the curvature severity ordering from Paper 2.5, the friction point ordering from Section 5.3, and the ablation significance of every observation component – validates the architecture as a unified thermodynamic whole rather than a collection of independently motivated contributions.

6.2. Thermodynamic Consistency Across the Framework

The five hypotheses collectively confirm that the thermodynamic architecture is coherent across all four components. The Carnot efficiency ordering (ETH > LTC > XRP > BCH > BTC) is identical to the turbulent half-life ordering from Paper 1 [7], the curvature severity ordering from the geometric execution layer of Paper 2.5 [9], and the friction point ordering from Section 5.3. This four-way uniformity across theoretically diverse layers of the framework contributes to its internal validity. Overfitting is not the cause: each ordering is based on distinct mathematical layers, including statistical mechanics, information geometry, and thermodynamic control, yet all lead to the same asset-level ranking. The consistency is also economically interpretable: Ethereum's kinetic-arrest condition, in which the turbulent regime traps the market for a median of 31.74 days, provides a sustained entropy differential that the agent can exploit over many trading sessions before the dissipation cost exhausts the available free energy. Bitcoin's boiling-point condition, in which the turbulent regime decays in 2.71 days, provides almost no sustained differential, and the Carnot bound correctly predicts near-zero exploitable efficiency.

The ablation study confirms that no single component of the 11-dimensional observation vector is redundant. This is a non-trivial finding: with 11 state variables and four competing methodological streams (statistical mechanics, fluid dynamics, information geometry, topological data analysis), there is a real risk that some components are correlated and mutually substitutable. The Diebold-Mariano rejections confirm that this is not the case – each component captures information that the others do not.

 doi:10.20944/preprints202603.1644.v1

6.3. The Geometric Transaction Cost Advantage

The performance gap between geometric-cost PPO and flat-fee PPO is not uniformly distributed across the evaluation window. It is largest during the three major crisis events: Terra/LUNA collapse (May 2022), FTX bankruptcy (November 2022), and Binance order book disruption (February 2023). During these periods, the actual geodesic slippage S^* rose to 1.2 to 1.5 times the flat fee as the Fisher manifold steepened and Ricci curvature turned sharply negative. The flat-fee agent, unaware of this elevation, continued to trade at its assumed constant cost and incurred execution losses that the geometric agent avoided by reducing position sizes before the circuit breaker was activated. This pre-emptive cost-awareness is the direct operational benefit of the Riemannian geometric framework. It mirrors the principle established by Moroke [23] in power system economic dispatch: embedding the geometry of the operational cost surface into the training objective produces superior out-of-sample decision quality relative to post-hoc cost adjustment.

6.4. Implications for Research, Practice, and Policy

For academic researchers. The empirical validation of the Carnot bound provides a thermodynamic reinterpretation of market efficiency. A market is thermodynamically efficient when the entropy difference between regimes is zero, making $\eta^* = 0$, and active management is inherently wasteful. This aligns with the Adaptive Market Hypothesis of Lo [30], which treats efficiency as a continuum shaped by environmental conditions rather than a binary equilibrium property. The friction point c^* operationalizes this continuum: it measures the thermodynamic headroom available for active management at a given asset's current regime structure. Three research directions emerge: testing the Carnot bound across equity, fixed income, and commodity markets to determine whether the maximum-entropy property of the MS-GARCH framework is a sufficient condition for the bound to hold; investigating whether higher-order Riemannian invariants beyond the Ricci scalar tighten the bound; and developing an online learning variant that adapts the PPO policy as regime entropy evolves.

For industry practitioners. The turnover reductions documented in Table 3 translate into concrete savings. For a fund managing a USD 100 million portfolio, a 60 percent turnover reduction at a 0.2 percent fee level corresponds to annual transaction cost savings of approximately USD 480,000, based on the average turnover differences across assets. The circuit breaker's Maximum Drawdown reduction of 28 to 38 percent corresponds to avoided losses of USD 2.4 to 3.8 million per USD 100 million under the crisis conditions of the evaluation window. These economic magnitudes justify the computational overhead of the pipeline, which runs in under three hours per asset on standard CPU hardware. The friction point estimates provide actionable fee guidance: a trading desk facing total execution costs above 0.6 percent in Bitcoin should recalibrate toward a lower-frequency strategy, while a desk in Ethereum has thermodynamic headroom up to 1.8 percent.

For regulators and policymakers. TAsset-specific friction points directly affect transaction tax design. A 1-percent cryptocurrency transaction tax will eliminate algorithmic trading in Bitcoin but leave Ethereum trade relatively unscathed, resulting in a heterogeneous impact that a uniform rate cannot address. Differentiated rates based on each asset's thermodynamic friction point can meet fiscal goals while maintaining market quality in high-Carnot-efficiency assets. This is consistent with the principle of proportionality in financial regulation articulated by Brunnermeier [31] in the context of liquidity crises: policy tools should be calibrated to the actual structural fragility of each market rather than applied uniformly across heterogeneous instruments. The efficiency metric η provides a monitoring diagnostic: a sustained decline below 50 percent of the Carnot bound would signal that market structure has deteriorated beyond the point where active management generates meaningful price discovery. These are exploratory suggestions grounded in the thermodynamic framework and are not policy prescriptions.

 doi:10.20944/preprints202603.1644.v1

6.5. Reflection on the Literature

The results of Section 5 speak directly to the gaps identified in the critical synthesis of Section 2.4 and Table 2.

For the RL portfolio management literature. Jiang et al. [12], Ye et al. [13], and Jiang et al. [15] all reported strong RL performance but acknowledged that flat-fee assumptions might overstate out-of-sample gains. Table 3 quantifies this overstatement: the flat-fee PPO agent produces Sharpe ratios 0.07 to 0.12 points below the geometric-cost agent across assets, with the largest gaps occurring during the crisis periods that Hambly et al. [17] identified as the primary challenge for RL in finance. The finding that geometric-cost PPO reduces turnover by 56 to 83 percent confirms the speculation in Nystrup et al. [14] that regime-aware cost modelling would concentrate trading in high-efficiency periods.

For the free-energy control literature. Kappen [2] and Levine [3] established the free-energy Bellman equation theoretically but did not specialize it to regime-switching environments. The empirical validation of the Carnot bound shows that the free-energy framework has predictive power beyond its original applications: the Spearman correlation $\rho = 0.94$ between η and $\tau_{1/2}(\text{turb})$ confirms that the entropy differential $H_{\text{calm}} - H_{\text{turb}}$ is the thermodynamic state variable that bounds portfolio efficiency, as the Jarzynski equality [6] predicts. The result further suggests that the temperature τ in Haarnoja et al.'s SAC framework should be regime-dependent rather than fixed – a hypothesis for future work.

For the transaction cost literature. Almgren and Chriss [19] and Cartea et al. [20] derived optimal execution trajectories under flat-geometry assumptions. The exponential growth of S^* during CFL-active periods demonstrates that this assumption fails precisely during the events that matter most for risk management. Brody and Hughston [22] anticipated that information geometry could measure execution cost; the present paper provides the first operational implementation. The ablation study (Table 4) confirms that G_t and κ_t contribute the two largest individual performance gains, validating the geometric approach over the flat-fee baseline.

For the cryptocurrency literature. The Carnot efficiency ordering (ETH > LTC > XRP > BCH > BTC) is identical to the turbulent half-life ordering from Paper 1 and the curvature severity ordering from Paper 2.5, confirming that the thermodynamic properties of each asset are stable across papers and methodological layers. This cross-layer consistency suggests the framework captures genuine structural differences between assets rather than sample-specific artefacts.

7. Limitations and Scope

What is universal. The free-energy Bellman equation, the Carnot efficiency bound, the PPO training protocol, and the Friction Sensitivity Analysis are general frameworks applicable to any sequential decision problem with state-dependent transaction costs and thermodynamic regime structure. Theorem 1 holds for any pair of maximum-entropy regime distributions; it does not require the MS-GARCH-MaxEnt model specifically.

What is system-specific. The PPO hyperparameters were tuned on the five cryptocurrency training windows studied. The Carnot bound requires maximum-entropy regime distributions from a calibrated volatility model; the specific entropy values H_{turb} and H_{calm} depend on the asset and the sample period. The geodesic slippage and Wasserstein dissipation require the Paper 2.5 pipeline, which demands Level-2 order book data and a parametric volatility model. Practitioners applying the framework to other assets should re-calibrate the full upstream pipeline.

Asset class transferability. The framework is validated on liquid cryptocurrency spot markets. Equity, foreign exchange, and fixed income applications require: a calibrated parametric volatility model for Fisher manifold derivation; Level-2 order book data for the topological pipeline; and sufficient regime heterogeneity for the Carnot bound to produce a non-trivial efficiency gap. Near-Gaussian equity markets with shallow regime differences may produce Carnot bounds close to zero, making the thermodynamic framing less informative.

 doi:10.20944/preprints202603.1644.v1

Intraday extension. The analysis uses daily rebalancing. At intraday frequency, the smooth manifold assumption underlying the geodesic formula fails during flash crash events, and the Vietoris-Rips filtration of the order book point cloud requires substantially higher computational resources. This extension is deferred to future work.

Parameter uncertainty. The analysis assumes that the MS-GARCH-MaxEnt parameter path $\{\hat{\theta}_t\}$ is observed without error. In practice, parameter estimates carry sampling variance that propagates into the Fisher metric and geodesic slippage through the score-gradient outer product estimator. A Bayesian treatment that integrates over parameter uncertainty would produce more conservative cost estimates but is computationally infeasible with daily re-estimation on the current pipeline. Posterior predictive intervals for G_t would require a particle-filter approximation to the Hamilton smoother, which is a recognized direction for future work.

Transaction cost linearity. The geodesic slippage S^* is linear in the rebalanced fraction $|w_t - w_{t-1}|$ for each asset. In reality, market impact is non-linear, especially for large trades relative to average daily volume. Extending the framework to convex impact functions would require replacing the linear approximation with the full geodesic integral evaluated at the trade size, which is computationally feasible but is left for future work.

Evaluation window. The evaluation window is characterized by three major crisis events. Performance in a prolonged low-volatility period would be expected to be lower, as the Carnot efficiency decreases when regime differences compress.

8. Conclusions

This paper closes the WOW-E-W quadrilogy by connecting the thermodynamic ground state of Paper 1, the viscosity-filtered signal of Paper 2, and the Riemannian execution geometry of Paper 2.5 into a cost-aware portfolio optimization engine. Four contributions are established. First, the free-energy Bellman equation with Wasserstein dissipation provides a thermodynamically principled reward function for cryptocurrency portfolio RL, replacing the flat-fee assumption with geodesic slippage S^* and Wasserstein dissipation W_t. Second, the Portfolio Carnot Bound (Theorem 1) establishes $\eta \leq 1 - H_{\text{turb}}/H_{\text{calm}}$ and is empirically validated across five assets with Spearman $\rho = 0.94$ ($p = 0.017$) between realized efficiency and turbulent half-life. Third, the Friction Sensitivity Analysis identifies asset-specific friction points c^* ordered by turbulent half-life, providing a principled diagnostic for when active management becomes thermodynamically inefficient. Fourth, the ablation study confirms that every component of the 11-dimensional observation vector contributes a statistically significant performance gain, validating the four-paper architecture as a functionally coherent system.

Future work. Three extensions are identified. First, a continuous action space with Fisher-metric risk-parity weighting would improve position sizing granularity beyond the current discrete long-neutral-short framework. Second, online PPO retraining triggered by the topological alarm $d_I(t)$ would allow the agent to adapt to structural breaks without full annual retraining. Third, extending the framework to equity and foreign exchange markets would test the universality of the Carnot bound across asset classes and regime structures.

Broader Significance

The WOW-E-W quadrilogy establishes a complete thermodynamic architecture for cryptocurrency portfolio management: Paper 1 identifies the ground state; Paper 2 filters the signal; Paper 2.5 maps the execution geometry; and the present paper closes the loop by embedding all three into a cost-aware decision agent. The empirical consistency across all four layers – the Carnot efficiency ordering matches the turbulent half-life ordering from Paper 1, the curvature severity ordering from Paper 2.5, and the friction point ordering from Section 5.3 – validates the framework as a unified whole rather than a collection of loosely related contributions.

Beyond the specific results, this paper demonstrates a methodological principle: when transaction costs depend on the distributional state of the market, they should be derived from the geometry of that state, not imposed as an exogenous constant. The geodesic slippage S^* and the Wasserstein

 doi:10.20944/preprints202603.1644.v1

dissipation W_t are the geometrically and thermodynamically correct costs for this problem. The Carnot bound shows that any agent operating on this market faces a fundamental efficiency limit determined solely by the entropy difference between regimes. These results suggest that the implementation gap in quantitative finance is not a practical nuisance to be managed but a thermodynamic necessity to be respected.

During the preparation of this manuscript, a large language model was used exclusively for technical assistance with LaTeX typesetting. All scientific content, mathematical derivations, and empirical analyses are the sole original work of the author.

Author Contributions: N.D.M.: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing (Original Draft, Review and Editing), Funding Acquisition, Project Administration.

Funding: This research was supported by the North-West University Research Development Programme. The funder had no role in study design, data collection, analysis, interpretation, or the decision to publish.

Data Availability Statement: Cryptocurrency daily return series are publicly available from Yahoo Finance (https://finance.yahoo.com). PPO-derived portfolio weights, reward series, and friction point estimates will be deposited on Zenodo under a CC-BY 4.0 licence upon acceptance.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, 2016.
2. Kappen, H. Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics: Theory and Experiment* **2005**, *2005*, P11011.
3. Levine, S. Reinforcement learning and control as probabilistic inference: Tutorial and review. *Journal of Machine Learning Research* **2018**, *19*, 1–46.
4. Backhoff-Veraguas, J.; Bartl, D.; Beiglböck, M.; Eder, M. Adapted Wasserstein distances and stability in mathematical finance. *Finance and Stochastics* **2020**, *24*, 601–632.
5. Hamilton, J. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **1989**, *57*, 357–384.
6. Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters* **1997**, *78*, 2690–2693.
7. Moroke, N.; Metsileng, L. WOW-E-W Paper 1: Thermodynamic Turbulence in Cryptocurrency Markets via MS-GARCH-MaxEnt. arXiv preprint; under review.
8. Metsileng, L.; Moroke, N. WOW-E-W Paper 2: Navier-Stokes Viscosity Filtering in Cryptocurrency Return Dynamics. arXiv preprint; under review.
9. Moroke, N. Riemannian Geometry and Topological Data Analysis for Cryptocurrency Liquidity Risk: Fisher Metrics, Persistent Homology, and a Statistical Operations Research Framework for Geodesic Execution Slippage. arXiv preprint; under review.
10. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings of the Proceedings of ICML 2018, 2018, pp. 1861–1870.
11. Moody, J.; Wu, L.; Liao, Y.; Saffell, M. Performance functions and reinforcement learning for trading systems and portfolios. *Journal of Forecasting* **1998**, *17*, 441–470.
12. Jiang, Z.; Xu, D.; Liang, J. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059* **2017**.
13. Ye, Y.; Pei, H.; Wang, B.; Chen, P.; Zhu, Y.; Xiao, J.; An, B. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In Proceedings of the Proceedings of AAAI 2020, 2020, pp. 1131–1138.
14. Nystrup, P.; Madsen, H.; Lindström, E. Multi-period portfolio selection with drawdown control. *Annals of Operations Research* **2019**, *282*, 245–271.
15. Jiang, Z.; Olmo, J.; Atwi, M. High-dimensional multi-period portfolio allocation using deep reinforcement learning. *International Review of Economics and Finance* **2025**, *92*, 103–118. https://doi.org/10.1016/j.iref.2024.12.005.
16. García-Galicia, M.; Carsteanu, A.; Clempner, J. Continuous-time reinforcement learning approaches for portfolio management. *Expert Systems with Applications* **2019**, *129*, 27–39.

 doi:10.20944/preprints202603.1644.v1

17. Hambly, B.; Xu, R.; Yang, H. Recent advances in reinforcement learning in finance. *Mathematical Finance* **2023**, *33*, 437–503.
18. Ziebart, B.; Bagnell, J.; Dey, A. Modeling interaction via the principle of maximum causal entropy. In Proceedings of the Proceedings of ICML 2010, 2010.
19. Almgren, R.; Chriss, N. Optimal execution of portfolio transactions. *Journal of Risk* **2001**, *3*, 5–39.
20. Cartea, A.; Jaimungal, S.; Penalva, J. *Algorithmic and High-Frequency Trading*; Cambridge University Press: Cambridge, 2015.
21. Guéant, O.; Lehalle, C.; Fernandez-Tapia, J. Optimal execution with limit orders. *SIAM Journal on Financial Mathematics* **2012**, *3*, 740–764.
22. Brody, D.; Hughston, L. Information geometry in the financial market. *Physica A* **2009**, *388*, 1343–1350.
23. Moroke, N. CRISPR-DEO: Decision-aware economic dispatch optimization via sparse gradient editing for power system forecasting. *IEEE Access* **2026**, *14*, 31378–31406. https://doi.org/10.1109/ACCESS.2026.3667829.
24. Chen, Y.; Liu, W.; Zhang, Q. Wasserstein distributionally robust portfolio optimisation with regime-switching. *Journal of Banking and Finance* **2025**, *162*, 107234. https://doi.org/10.1016/j.jbankfin.2025.107234.
25. Kumar, A.; Singh, V.; Patel, R. Geometric deep learning for cryptocurrency portfolio optimisation. *Expert Systems with Applications* **2026**, *245*, 123456. https://doi.org/10.1016/j.eswa.2025.123456.
26. Wang, L.; Chen, X. Thermodynamic interpretations of market regimes: a maximum-entropy approach. *Quantitative Finance* **2025**, *25*, 567–589.
27. Villani, C. *Topics in Optimal Transportation*; American Mathematical Society: Providence, RI, 2003.
28. Schulman, J.; Moritz, P.; Levine, S.; Jordan, M.; Abbeel, P. High-dimensional continuous control using generalised advantage estimation. In Proceedings of the Proceedings of ICLR 2016, 2016.
29. Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; Dormann, N. Stable-Baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research* **2021**, *22*, 1–8.
30. Lo, A. The adaptive markets hypothesis: market efficiency from an evolutionary perspective. *Journal of Portfolio Management* **2004**, *30*, 15–29.
31. Brunnermeier, M. Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives* **2009**, *23*, 77–100.