Information-Geometric Ergodic Theory: Trajectory Fisher Information, Quantitative Mixing Bounds, and Computational Detection

D. Christian[1, *]

[1] *The Axioms of Pattern Ontology (APO) Initiative*

(Dated: March 7, 2026)

We develop a systematic information-geometric theory of ergodic properties in dynamical systems. The central construction is the *trajectory Fisher Information Matrix* (tFIM), a Riemannian metric on initial conditions defined via kernel-smoothed empirical measures of trajectories. The tFIM measures how sensitively the long-run statistics of a trajectory depend on the initial condition—precisely the quantity that ergodic properties control.

We prove three characterization theorems. First, ergodicity is equivalent to trajectory FIM collapse: a system is ergodic if and only if the tFIM converges to zero for almost every initial condition (Theorem IV.1). Second, mixing is equivalent to temporal FIM factorization: the cross-information between past and future observations decays to zero if and only if the system is mixing (Theorem V.3). Third, for hyperbolic systems, the Hopf argument becomes directional FIM degeneration: ergodicity is equivalent to FIM collapse along both stable and unstable manifolds (Theorem VI.2).

Beyond these qualitative equivalences, we establish quantitative bounds. For exponentially mixing systems with rate α, the tFIM decays as $\|g^{\mathrm{traj}}(T)\| = \mathcal{O}(T^{-1})$, with constants depending explicitly on α, the Lyapunov spectrum, and the phase space dimension (Theorem VII.1). For polynomially mixing systems with rate $t^{-\beta}$, we obtain $\|g^{\mathrm{traj}}(T)\| = \mathcal{O}(T^{-2\beta/(2\beta+d+2)})$ (Theorem VII.3). These bounds complement classical entropy-based characterizations (Kolmogorov–Sinai) by providing geometric rates for a Riemannian quantity rather than a scalar one.

We introduce the *FIM mixing estimator*, an algorithm that estimates the mixing rate of a system from a single trajectory using the tFIM, with convergence guarantees (Theorem VIII.3). The estimator requires no prior knowledge of the system dynamics and achieves minimax-optimal rates for the class of exponentially mixing systems.

Applications to geodesic flows on negatively curved manifolds, hyperbolic toral automorphisms, cellular automata, and Navier–Stokes flows illustrate the scope of the theory. We extend the framework to partially hyperbolic and infinite-dimensional settings.

I. INTRODUCTION

A. Ergodic Properties as Distinguishability Statements

Ergodic theory classifies dynamical systems by their long-run statistical properties. The central hierarchy—ergodicity, weak mixing, strong mixing, Kolmogorov, Bernoulli—describes progressively stronger forms of statistical regularity. Each level can be stated as a statement about *distinguishability*:

- **Ergodicity**: a single trajectory becomes *indistinguishable* from random sampling of the invariant measure.
- **Mixing**: past and future observations become *indistinguishable* from independent samples.
- **Chaos**: nearby trajectories become *distinguishable* exponentially fast, but boundedly.
- **Blow-up** (in PDE systems): distinguishability diverges in finite time.

Information geometry provides the natural language for distinguishability. The Fisher Information Matrix (FIM) [1] is the canonical Riemannian metric on spaces of probability distributions, defined as the expected outer product of the score function. Its eigenvalues measure the precision with which parameters can be estimated from observations—large FIM means high distinguishability; degenerate FIM means indistinguishability.

The core thesis of this paper is that these two frameworks—the ergodic hierarchy and the FIM—are not merely analogous but formally equivalent. We make this precise by constructing the *trajectory Fisher Information Matrix* (tFIM), a new geometric object that bridges the two theories.

* daniel@freereason.org

B. The Trajectory Fisher Information Matrix

The tFIM measures how sensitively the long-run statistics of a trajectory depend on the choice of initial condition.

Consider a dynamical system $\phi_t : X \to X$ preserving a measure μ, and a parameterized family of initial conditions $\theta \mapsto x(\theta) \in X$. The trajectory from $x(\theta)$ generates an empirical density, which we smooth with a kernel to obtain a well-defined probability density $\hat{p}_T^\theta$. The tFIM is the Fisher information of this kernel-smoothed empirical density with respect to θ:

$$g_{ij}^{\text{traj}}(\theta, T) = \int \frac{\partial_{\theta_i} \hat{p}_T^\theta(y) \cdot \partial_{\theta_j} \hat{p}_T^\theta(y)}{\hat{p}_T^\theta(y)}\, dy. \tag{1}$$

For ergodic systems, $\hat{p}_T^\theta$ converges to the (smoothed) invariant density, which does not depend on θ. Therefore the tFIM converges to zero. This convergence is the information-geometric content of ergodicity.

For non-ergodic systems, trajectories from different initial conditions generate persistently distinguishable statistics, and the tFIM remains bounded below.

The tFIM thus provides a Riemannian metric on the space of initial conditions whose collapse is exactly ergodicity. This is a stronger statement than classical equivalences, because the FIM is not a scalar but a matrix: it carries directional information. This directional structure is precisely what makes the Hopf argument natural in FIM language.

C. What This Paper Does

1. **Defines the trajectory FIM** (Section III): We give a rigorous definition using kernel density estimation from trajectory data, prove well-definedness, and establish a convergence theorem relating the tFIM to classical ergodic properties.

2. **Characterizes ergodicity** (Section IV): Ergodicity $\Leftrightarrow g^{\text{traj}}(T) \to 0$ for μ-a.e. initial condition.

3. **Characterizes mixing** (Section V): Mixing $\Leftrightarrow$ temporal FIM factorization (cross-information between (x_0, x_t) decays).

4. **Reframes the Hopf argument** (Section VI): Ergodicity of hyperbolic systems $\Leftrightarrow$ directional FIM degeneration along stable and unstable foliations.

5. **Proves quantitative FIM-mixing bounds** (Section VII): Explicit decay rates for the tFIM under exponential and polynomial mixing, with dependence on Lyapunov exponents, dimension, and mixing rate.

6. **Introduces a mixing estimator** (Section VIII): An algorithm to estimate the mixing rate from trajectory data using the tFIM, with convergence guarantees.

7. **Applies to multiple domains** (Section IX): Geodesic flows, toral automorphisms, cellular automata, Navier–Stokes, and statistical mechanics.

8. **Extends to harder settings** (Section X): Partially hyperbolic systems, infinite-dimensional systems, and non-invertible maps.

D. What This Paper Does Not Do

We do not claim priority for the observation that ergodic properties involve information loss. This is well-known and encoded in the Kolmogorov–Sinai entropy, transfer entropy [21], and related concepts. What we offer is a *Riemannian* (matrix-valued, directional) formalization rather than a scalar one, together with the first quantitative bounds and a computational estimator.

We also do not claim that FIM methods supersede classical techniques for proving ergodicity. In concrete cases, the Hopf argument, spectral methods, and entropy calculations remain more efficient. The value of the FIM perspective is in settings where classical methods are difficult (infinite-dimensional systems, singularities, computation from data) and in providing a unified geometric framework.

E. Related Work

Information geometry and dynamical systems. The Fisher information metric appears in statistical mechanics as the Ruppeiner and Weinhold metrics [19]. Green [6] connected Fisher information to Lyapunov exponents for differentiable dynamical systems. Liang [13] developed information flow as a rigorous notion of causality in dynamical systems. Our work differs in that we define a new geometric object (the tFIM) and prove equivalences with the full ergodic hierarchy, rather than studying existing information-theoretic quantities in dynamical settings.

Kernel density estimation for dynamical systems. Hang et al. [7] established consistency and convergence rates for kernel density estimators applied to trajectories of ergodic dynamical systems, using the $\mathcal{C}$-mixing framework. Maume-Deschamps [15] obtained pointwise convergence using Höffding-type inequalities. We build on this foundation by studying the Fisher information *of* the kernel density estimator, which is a second-order quantity requiring derivative estimates, not merely density estimates.

Mixing time estimation. Wolfer [22] estimated mixing times of Markov chains from a single trajectory using contraction methods. Hsu et al. [**?**] initiated the program of mixing time estimation from trajectory data. Our FIM-based estimator provides a geometric alternative with explicit dependence on the Lyapunov spectrum, applicable to continuous-time continuous-state systems beyond the Markov chain setting.

Empirical measure convergence. The convergence of empirical measures to invariant measures in Wasserstein distance is studied in [12, 20]. Our trajectory divergence is related but operates in the Fisher metric (a second-order quantity) rather than the Wasserstein metric (a first-order transport distance).

II. PRELIMINARIES

A. Measure-Preserving Dynamical Systems

Let $(X, \mathcal{B}, \mu)$ be a probability space with $X \subseteq \mathbb{R}^d$ a smooth manifold (or $\mathbb{R}^d$ itself).

Definition II.1 (Measure-preserving system)**.** *A* measure-preserving transformation *is a measurable map* $T : X \to X$ *with* $\mu(T^{-1}A) = \mu(A)$ *for all* $A \in \mathcal{B}$*. A* measure-preserving flow *is a one-parameter family* $\phi_t : X \to X$, $t \in \mathbb{R}$, *of measure-preserving transformations with* $\phi_0 = \mathrm{id}$ *and* $\phi_{s+t} = \phi_s \circ \phi_t$.

We state results primarily for discrete time (iterates of T) and note the continuous-time analogues.

Definition II.2 (Ergodic hierarchy)**.** *Let* $(X, \mathcal{B}, \mu, T)$ *be a measure-preserving system.*

(E) T *is* ergodic *if every* T*-invariant set has measure* 0 *or* 1.

(WM) T *is* weakly mixing *if for all* $A, B \in \mathcal{B}$*:* $\frac{1}{N}\sum_{n=0}^{N-1} |\mu(A \cap T^{-n}B) - \mu(A)\mu(B)| \to 0$.

(SM) T *is* (strongly) mixing *if for all* $A, B \in \mathcal{B}$*:* $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$.

(K) T *is* Kolmogorov *(K-mixing) if it has completely positive entropy.*

(B) T *is* Bernoulli *if it is isomorphic to a Bernoulli shift.*

These form the ergodic hierarchy*:* $(B) \subset (K) \subset (SM) \subset (WM) \subset (E)$.

Definition II.3 (Exponential and polynomial mixing)**.** *A mixing system has* exponential mixing rate $\alpha > 0$ *(for a class* $\mathcal{F}$ *of observables) if*

$$\left| \int f \cdot (g \circ T^n)\, d\mu - \int f\, d\mu \int g\, d\mu \right| \le C_{f,g}\, e^{-\alpha n} \tag{2}$$

for all $f, g \in \mathcal{F}$ *and all* $n \ge 0$*. It has* polynomial mixing rate $\beta > 0$ *if the bound is* $C_{f,g}\, n^{-\beta}$.

B. Information Geometry

We recall the essentials of Fisher information geometry [1].

Definition II.4 (Statistical model and score)**.** *A* statistical model *is a family of probability densities* $\{p_\theta : \theta \in \Theta\}$ *on a measurable space* $(Y, \mathcal{C})$*, parameterized by an open subset* $\Theta \subseteq \mathbb{R}^m$*. The* score function *is* $s_i(\theta, y) = \partial_{\theta_i} \ln p_\theta(y)$.

Definition II.5 (Fisher Information Matrix)**.** *The* Fisher Information Matrix *(FIM) at parameter θ is*

$$g_{ij}(\theta) = \int s_i(\theta, y)\, s_j(\theta, y)\, p_\theta(y)\, dy = \mathbb{E}_{p_\theta}\big[s_i\, s_j\big]. \tag{3}$$

Under standard regularity conditions (differentiability under the integral sign), $g_{ij}(\theta) = -\mathbb{E}_{p_\theta}\big[\partial_{\theta_i}\partial_{\theta_j} \ln p_\theta\big]$.

The FIM is a positive semidefinite symmetric matrix. It defines a Riemannian metric on Θ (the Fisher–Rao metric), under which the geodesic distance between parameters θ and θ' measures the statistical distinguishability of p_θ and $p_{\theta'}$.

Definition II.6 (Spectral gap of the FIM)**.** *The* FIM spectral gap *is the smallest positive eigenvalue:*

$$\lambda_1(\theta) = \inf_{v \neq 0,\, v \perp \ker g} \frac{g_{ij} v^i v^j}{|v|^2}. \tag{4}$$

When g is positive definite, λ_1 is the smallest eigenvalue. When $\lambda_1 = 0$, the FIM is degenerate and there exist directions along which nearby parameters are indistinguishable.

C. Kernel Density Estimation

We recall the kernel density estimator (KDE) as the tool that converts trajectory data into smooth densities.

Definition II.7 (Kernel and KDE)**.** *A* kernel *is a nonnegative function $K : \mathbb{R}^d \to [0, \infty)$ with $\int K(y)\, dy = 1$, $\int y\, K(y)\, dy = 0$, and $\sigma_K^2 := \int \|y\|^2 K(y)\, dy < \infty$. The* scaled kernel *is $K_h(y) = h^{-d} K(y/h)$ for bandwidth $h > 0$.*

Given observations $x_1, \ldots, x_N \in \mathbb{R}^d$, the kernel density estimator *is*

$$\hat{p}_h(y) = \frac{1}{N} \sum_{i=1}^{N} K_h(y - x_i). \tag{5}$$

For a continuous trajectory $\{\phi_t(x)\}_{0 \le t \le T}$, the continuous-time KDE is

$$\hat{p}^x_{T,h}(y) = \frac{1}{T} \int_0^T K_h(y - \phi_t(x))\, dt. \tag{6}$$

The KDE turns the singular empirical measure (a sum of Dirac masses or a measure on a curve) into a smooth probability density. For observations from an ergodic dynamical system, $\hat{p}^x_{T,h} \to K_h * p_\mu$ as $T \to \infty$ for μ-a.e. x, where p_μ is the density of the invariant measure (consistency follows from the Birkhoff ergodic theorem applied to the function $y \mapsto K_h(y - \cdot)$).

III. THE TRAJECTORY FISHER INFORMATION MATRIX

A. Definition

We now define the central object of this paper.

Let $(X, \mathcal{B}, \mu, \phi_t)$ be a measure-preserving flow on $X \subseteq \mathbb{R}^d$, with μ admitting a continuous density p_μ with respect to Lebesgue measure. Let $\Theta \subseteq \mathbb{R}^m$ be an open parameter space and $\theta \mapsto x(\theta) \in X$ a smooth embedding.

Definition III.1 (Trajectory Fisher Information Matrix)**.** *For bandwidth $h > 0$ and observation time $T > 0$, the* trajectory Fisher Information Matrix *(tFIM) is:*

$$g^{\mathrm{traj}}_{ij}(\theta, T, h) = \int_{\mathbb{R}^d} \frac{\partial_{\theta_i} \hat{p}^\theta_{T,h}(y) \cdot \partial_{\theta_j} \hat{p}^\theta_{T,h}(y)}{\hat{p}^\theta_{T,h}(y)}\, dy, \tag{7}$$

where $\hat{p}^\theta_{T,h} := \hat{p}^{x(\theta)}_{T,h}$ is the trajectory KDE from (6) started at $x(\theta)$.

Remark III.2 (Interpretation)**.** *g_{ij}^{traj} measures the sensitivity of the smoothed empirical density to the initial condition parameter θ. It is the FIM of the statistical model $\{\hat{p}^{\theta}_{T,h} : \theta \in \Theta\}$. Large g^{traj} means that the trajectory statistics from $x(\theta)$ and $x(\theta+\delta\theta)$ are easily distinguished by a statistician observing the smoothed empirical density. Small g^{traj} means the two trajectories produce nearly identical long-run statistics.*

Remark III.3 (Discrete-time version)**.** *For a discrete-time system $T : X \to X$, the trajectory KDE is $\hat{p}^{\theta}_{N,h}(y) = \frac{1}{N}\sum_{n=0}^{N-1} K_h(y - T^n(x(\theta)))$, and the tFIM is defined by the same formula (7) with T replaced by N.*

B. Well-Definedness

Proposition III.4 (Well-definedness of the tFIM)**.** *Suppose that:*

(a) *The kernel K is C^1 with $\|\nabla K\|_\infty < \infty$ and $\int \|\nabla K\|^2 / K\, dy < \infty$.*

(b) *The flow ϕ_t is C^1 in the initial condition, uniformly on compact subsets of $\Theta \times [0, T]$.*

(c) *The bandwidth satisfies $h > 0$.*

Then $g_{ij}^{\mathrm{traj}}(\theta, T, h)$ is finite for all $\theta \in \Theta$ and all $T, h > 0$.

Proof. The trajectory KDE $\hat{p}^{\theta}_{T,h}(y)$ is a time-average of $K_h(y - \phi_t(x(\theta)))$, which is C^1 in θ by hypothesis (b) and the chain rule. Explicitly:

$$\partial_{\theta_i}\hat{p}^{\theta}_{T,h}(y) = -\frac{1}{T}\int_0^T (\nabla K_h)(y - \phi_t(x(\theta))) \cdot \frac{\partial \phi_t(x(\theta))}{\partial \theta_i}\, dt. \tag{8}$$

Since ∇K_h is bounded by $h^{-d-1}\|\nabla K\|_\infty$ and $\partial_{\theta_i}\phi_t$ is bounded on $[0, T]$ by hypothesis (b), the integrand is bounded. Therefore $\partial_{\theta_i}\hat{p}^{\theta}_{T,h}(y)$ is bounded uniformly in y.

The denominator $\hat{p}^{\theta}_{T,h}(y)$ is bounded below by $h^{-d}\inf K \cdot \mu_T^{x(\theta)}(B(y,h))$, where μ_T^x is the empirical measure. For fixed T and h, the set $\{y : \hat{p}^{\theta}_{T,h}(y) > 0\}$ is an open neighborhood of the trajectory, and the ratio $|\partial_\theta \hat{p}|^2/\hat{p}$ is integrable by hypothesis (a). □

C. The Derivative Formula

The key to the tFIM is equation (8), which we now develop further.

Denote the *sensitivity matrix* $Y_i(t) := \partial_{\theta_i}\phi_t(x(\theta)) \in \mathbb{R}^d$, the derivative of the flow map with respect to the ith parameter. For a flow generated by the ODE $\dot{x} = F(x)$, Y_i satisfies the variational equation:

$$\dot{Y}_i = DF(\phi_t(x)) \cdot Y_i, \qquad Y_i(0) = \partial_{\theta_i}x(\theta). \tag{9}$$

The tFIM can then be written as:

$$g_{ij}^{\mathrm{traj}}(\theta, T, h) = \int_{\mathbb{R}^d} \frac{A_i(y)\cdot A_j(y)}{\hat{p}^{\theta}_{T,h}(y)}\, dy, \tag{10}$$

where

$$A_i(y) = \frac{1}{T}\int_0^T (\nabla K_h)(y - \phi_t(x)) \cdot Y_i(t)\, dt. \tag{11}$$

The vector $A_i(y)$ is a time-averaged, kernel-weighted measurement of the trajectory's sensitivity to parameter θ_i, observed at spatial point y.

Remark III.5 (The sensitivity–ergodicity tension)**.** *For a hyperbolic system, the sensitivity $\|Y_i(t)\|$ grows exponentially as $e^{\lambda t}$, where λ is the maximal Lyapunov exponent. Naively, this suggests A_i should grow, making g^{traj} diverge. But the growth of $\|Y_i\|$ is in a fixed* direction *(the unstable direction), while the* spatial weighting *by $\nabla K_h(y - \phi_t(x))$ selects only the component of Y_i relevant to the neighborhood of y. In a mixing system, the direction of Y_i when the trajectory revisits a neighborhood of y is effectively randomized by the dynamics. The resulting cancellation in the time average is precisely what drives $g^{\mathrm{traj}} \to 0$. This tension between pointwise sensitivity growth and statistical averaging is the dynamical content of the tFIM.*

D. Convergence Theorem

Theorem III.6 (tFIM convergence)**.** *Let (X, μ, ϕ_t) be a measure-preserving flow with μ admitting a C^2 density $p_\mu > 0$. Let K be a C^1 kernel with compact support and $h = h(T) \to 0$ with $Th^{d+2} \to \infty$ as $T \to \infty$. Then for μ-a.e. initial condition $x(\theta)$:*

(i) $\hat{p}^{\theta}_{T,h} \to K_h * p_\mu$ *in* L^1 *as* $T \to \infty$.

(ii) *If* ϕ_t *is ergodic, then* $g^{\text{traj}}_{ij}(\theta, T, h) \to 0$ *as* $T \to \infty$.

(iii) *If* ϕ_t *is not ergodic, there exist* θ, θ' *and a constant* $c > 0$ *such that* $g^{\text{traj}}(\theta, T, h) \geq c > 0$ *for all sufficiently large* T *and sufficiently small* h.

Proof. Part (i). By the Birkhoff ergodic theorem, for μ-a.e. x and every bounded continuous function f: $\frac{1}{T}\int_0^T f(\phi_t(x))\,dt \to \int f\,d\mu$. Taking $f(z) = K_h(y - z)$ for fixed y and h:

$$\hat{p}^{x}_{T,h}(y) \to \int K_h(y - z)\, p_\mu(z)\, dz = (K_h * p_\mu)(y). \tag{12}$$

The convergence is pointwise in y for a.e. x. Since K_h is bounded, dominated convergence extends this to L^1 convergence.

Part (ii). If ϕ_t is ergodic, the limit $K_h * p_\mu$ does not depend on the initial condition $x(\theta)$. Therefore $\partial_{\theta_i}(K_h * p_\mu) = 0$, and the score $s_i = \partial_{\theta_i} \ln \hat{p}^{\theta}_{T,h}$ converges to zero in $L^2(\hat{p}^{\theta}_{T,h})$.

More precisely, we bound:

$$g^{\text{traj}}_{ij} = \int \frac{(\partial_i \hat{p})(\partial_j \hat{p})}{\hat{p}}\, dy \leq \left\|\hat{p}^{-1}\right\|_\infty \|\partial_i \hat{p}\|_{L^2} \|\partial_j \hat{p}\|_{L^2}. \tag{13}$$

The norm $\left\|\hat{p}^{-1}\right\|_\infty$ is bounded for T large enough (since $\hat{p} \to K_h * p_\mu > 0$ on the support of p_μ, and $p_\mu > 0$ by hypothesis). The norm $\|\partial_i \hat{p}\|_{L^2}$ converges to $\|\partial_i(K_h * p_\mu)\|_{L^2} = 0$. Therefore $g^{\text{traj}}_{ij} \to 0$.

Part (iii). If ϕ_t is not ergodic, there exist ergodic components X_1, X_2 with $\mu(X_1), \mu(X_2) > 0$ and distinct conditional invariant measures $\mu_1 \neq \mu_2$. Choose θ, θ' with $x(\theta) \in X_1$ and $x(\theta') \in X_2$. Then $\hat{p}^{\theta}_{T,h} \to K_h * p_{\mu_1}$ and $\hat{p}^{\theta'}_{T,h} \to K_h * p_{\mu_2}$. Since $\mu_1 \neq \mu_2$, the densities $K_h * p_{\mu_1}$ and $K_h * p_{\mu_2}$ are distinct (for small h). The FIM of the parameterized family $\epsilon \mapsto (1-\epsilon)\hat{p}^{\theta} + \epsilon \hat{p}^{\theta'}$ at $\epsilon = 0$ is bounded below by the squared L^2 distance between the score functions, which is positive. A continuous parameterization $\theta \mapsto x(\theta)$ crossing between X_1 and X_2 then has $g^{\text{traj}} \geq c > 0$. □

IV. ERGODICITY AS TRAJECTORY FIM COLLAPSE

Theorem III.6 gives a one-directional statement: ergodicity implies tFIM collapse. We now sharpen this to a full equivalence.

Theorem IV.1 (Ergodicity ⇔ tFIM collapse)**.** *Let (X, μ, ϕ_t) be a measure-preserving flow with μ admitting a C^2 positive density. Let $\Theta \subseteq \mathbb{R}^m$ and $\theta \mapsto x(\theta)$ be a smooth embedding with $x(\Theta)$ intersecting every ergodic component of positive measure. Then the following are equivalent:*

(i) ϕ_t *is ergodic with respect to* μ.

(ii) *For* μ*-a.e.* $x(\theta)$ *and every bandwidth schedule* $h(T) \to 0$ *with* $Th^{d+2} \to \infty$*:*

$$g^{\text{traj}}(\theta, T, h(T)) \to 0 \quad \text{as } T \to \infty. \tag{14}$$

(iii) *(Trajectory divergence collapse.) For* μ*-a.e.* x*, the empirical measure* μ^x_T *converges weakly to* μ.

Proof. (i)⇔(iii) is the Birkhoff ergodic theorem. (i)⇒(ii) is Theorem III.6(ii). (ii)⇒(i): Contrapositive. If ϕ_t is not ergodic, there exist distinct ergodic components, and by the hypothesis on $x(\Theta)$, there exist θ, θ' mapping into different components. By Theorem III.6(iii), the tFIM remains bounded below. □

A. Rate of Convergence: Ergodic but Not Mixing

For ergodic but non-mixing systems, the tFIM converges but slowly.

Proposition IV.2 (Ergodic rate). *If (X, μ, T) is ergodic with $T : X \to X$ a measure-preserving map, then for the discrete-time tFIM with bandwidth $h = h(N) \sim N^{-1/(d+4)}$:*

$$g^{\text{traj}}(\theta, N, h) = \mathcal{O}(N^{-2/(d+4)}) \tag{15}$$

for μ-a.e. $x(\theta)$. Under exponential mixing with rate α, the improved bound $g^{\text{traj}} = \mathcal{O}(N^{-1})$ holds (Theorem VII.1 below).

Proof. This follows from the standard bias-variance decomposition of the KDE. The squared bias of $\hat{p}^{\theta}_{N,h}$ is $\mathcal{O}(h^4)$ (from the second-order moment of K), and the variance is $\mathcal{O}(1/(Nh^d))$ (from the CLT for dependent sequences; the ergodic theorem gives the LLN, and the CLT for mixing sequences [18] bounds the fluctuations).

For the derivative $\partial_\theta \hat{p}$, the bias is still $\mathcal{O}(h^4)$ (the limit $\partial_\theta(K_h * p_\mu) = 0$), and the variance is $\mathcal{O}(1/(Nh^{d+2}))$ (the derivative of K_h is $h^{-d-1}K'(y/h)$, contributing an extra h^{-2} to the variance).

The tFIM involves $(\partial_\theta \hat{p})^2/\hat{p}$. Since $\hat{p} \to K_h * p_\mu \geq c_h > 0$ on the support, the tFIM is bounded by $\mathcal{O}(1/(Nh^{d+2}))$. Optimizing h for the combined bias-variance tradeoff gives $h \sim N^{-1/(d+4)}$ and $g^{\text{traj}} = \mathcal{O}(N^{-2/(d+4)})$. □

V. MIXING AS TEMPORAL FIM FACTORIZATION

A. The Temporal FIM

Mixing is the statement that past and future become statistically independent. We formalize this using the FIM of the joint distribution of past and future observations.

Definition V.1 (Temporal FIM). *Let μ_θ be a parameterized family of initial distributions with $\mu_0 = \mu$ (the invariant measure). The* temporal FIM at lag t *is the FIM of the joint distribution of (x_0, x_t), where $x_0 \sim \mu_\theta$ and $x_t = \phi_t(x_0)$:*

$$g^{\text{temp}}_{ij}(t) = \int\int s^{(t)}_i(x_0, x_t)\, s^{(t)}_j(x_0, x_t)\, p^{(t)}(x_0, x_t; \theta)\, dx_0\, dx_t \Big|_{\theta=0}, \tag{16}$$

where $s^{(t)}_i = \partial_{\theta_i} \ln p^{(t)}$ is the score of the joint density $p^{(t)}(x_0, x_t; \theta)$.

The temporal FIM decomposes naturally. Denote the marginal score of x_0 alone as $s^{(0)}_i(x_0) = \partial_{\theta_i} \ln p_\theta(x_0)$, and the conditional score of x_t given x_0 as $s^{(t|0)}_i(x_t|x_0)$. Since $x_t = \phi_t(x_0)$ is deterministic given x_0, the "conditional score" is absorbed into the marginal. However, from the *observer's* perspective, if one observes x_t without knowing x_0, the information about θ from x_t depends on the statistical coupling between x_0 and x_t.

Definition V.2 (Cross-information). *The* cross-information *at lag t is the matrix:*

$$C_{ij}(t) = g^{\text{temp}}_{ij}(t) - g^{(0)}_{ij} - g^{(t)}_{ij}, \tag{17}$$

where $g^{(0)}$ and $g^{(t)}$ are the marginal FIMs of x_0 and x_t respectively. When the joint distribution factorizes ($x_0 \perp x_t$), we have $C(t) = 0$.

Since μ is invariant, $g^{(t)} = g^{(0)}$ for all t. The cross-information $C(t)$ captures the excess Fisher information in the joint observation (x_0, x_t) beyond what the marginals provide.

Theorem V.3 (Mixing ⇔ temporal FIM factorization). *Let (X, μ, ϕ_t) be a measure-preserving system with μ admitting a smooth positive density. The following are equivalent:*

(i) *ϕ_t is mixing.*

(ii) *For every smooth parameterized family μ_θ with $\mu_0 = \mu$: $C_{ij}(t) \to 0$ as $t \to \infty$.*

(iii) *The joint density $p^{(t)}(x_0, x_t)$ converges weakly to $p_\mu(x_0)\, p_\mu(x_t)$ as $t \to \infty$.*

Proof. (i)⇔(iii): This is the definition of mixing. For a measure-preserving system, $\mu(A \cap T^{-n}B) \to \mu(A)\mu(B)$ for all measurable A, B if and only if the joint distribution of (x_0, x_n) converges weakly to the product measure $\mu \otimes \mu$. (The measure-theoretic statement is equivalent to weak convergence by testing against indicator functions, then extending by density.)

(iii)⇒(ii): If the joint density factorizes asymptotically, $p^{(t)}(x_0, x_t; \theta) \to p_\theta(x_0)\, p_\theta(x_t)$, then the joint score becomes $s_i^{(t)} \to s_i^{(0)}(x_0) + s_i^{(0)}(x_t)$ (since $s_i^{(0)} = s_i^{(t)}$ by stationarity). The joint FIM becomes $g^{\text{temp}}(t) \to g^{(0)} + g^{(0)} = 2g^{(0)}$, so $C(t) = g^{\text{temp}}(t) - 2g^{(0)} \to 0$.

(ii)⇒(iii): If the cross-information vanishes for every parameterized family μ_θ with $\mu_0 = \mu$, then for every smooth direction v, the joint distribution (x_0, x_t) carries no more parametric information about θ than the marginals separately. By the data processing inequality, this implies asymptotic independence of x_0 and x_t: if (x_0, x_t) were not asymptotically independent, one could construct a family μ_θ for which the joint carries strictly more information than the marginals. □

B. Quantitative Temporal FIM Decay

Proposition V.4 (Exponential mixing implies exponential FIM factorization). *If (X, μ, ϕ_t) is exponentially mixing with rate α for C^1 observables, then the cross-information decays exponentially:*

$$\|C(t)\| \leq C_0\, e^{-\alpha t}, \tag{18}$$

where C_0 depends on the parameterized family μ_θ and the smoothness of p_μ.

Proof. The cross-information $C_{ij}(t)$ can be expressed as a correlation of score functions:

$$\begin{aligned} C_{ij}(t) = &\int s_i^{(0)}(x_0)\, s_j^{(0)}(x_t)\, p^{(t)}(x_0, x_t)\, dx_0\, dx_t \\ &- \int s_i^{(0)}(x_0)\, p_\mu(x_0)\, dx_0 \int s_j^{(0)}(x_t)\, p_\mu(x_t)\, dx_t. \end{aligned} \tag{19}$$

The second integral vanishes (the score has zero mean: $\mathbb{E}[s_i] = 0$). The first integral is the correlation $\text{Cor}(s_i^{(0)}(x_0),\, s_j^{(0)}(\phi_t(x_0)))$ under μ. For exponentially mixing systems, this correlation decays as $e^{-\alpha t}$ for C^1 observables, giving the stated bound. □

Remark V.5 (Weak mixing and Cesàro FIM decay). *Weak mixing is equivalent to $\frac{1}{N}\sum_{n=0}^{N-1} |C(n)| \to 0$ rather than $C(n) \to 0$. In FIM terms: the cross-information need not converge to zero pointwise in time, but its time average does. This corresponds to information about the past being "mostly" destroyed, with occasional recurrences that do not persist.*

VI. THE HOPF ARGUMENT AS DIRECTIONAL FIM DEGENERATION

A. Classical Hopf Argument

The Hopf argument [9] is the standard technique for proving ergodicity of hyperbolic systems. We recall the structure:

1. Any T-invariant function f is constant along stable manifolds $W^s(x)$, since points on the same stable leaf share the same future and hence the same forward time average.

2. By time reversal, f is constant along unstable manifolds $W^u(x)$.

3. If W^s and W^u are transverse and jointly span the tangent space (hyperbolicity), then f is locally constant, hence constant μ-a.e.

4. Therefore T is ergodic.

B. Directional FIM

We reformulate the Hopf argument in FIM terms.

Definition VI.1 (Directional tFIM)**.** *Let W be a smooth foliation of X. For $x = x(\theta)$, let $\pi_W(\theta) : T_\theta\Theta \to T_x W(x)$ be the projection onto the tangent space of the leaf through x. The* directional tFIM *along W is:*

$$g^W_{ij}(\theta, T, h) = (\pi_W)^k_i \, (\pi_W)^\ell_j \, g^{\mathrm{traj}}_{k\ell}(\theta, T, h). \tag{20}$$

This measures the sensitivity of trajectory statistics to perturbations of the initial condition along the foliation W.

C. Hopf–FIM Equivalence

Theorem VI.2 (Hopf argument as FIM degeneration)**.** *Let ϕ_t be an Anosov flow (uniformly hyperbolic) on a compact Riemannian manifold (X, μ), with stable and unstable foliations W^s, W^u. Then the following are equivalent:*

(i) *ϕ_t is ergodic.*

(ii) *The directional tFIM along W^s collapses: $g^{W^s}(\theta, T, h) \to 0$ as $T \to +\infty$; and along W^u: $g^{W^u}(\theta, T, h) \to 0$ as $T \to -\infty$ (backward time).*

(iii) *The full tFIM g^{traj} collapses as $|T| \to \infty$.*

Proof. (i)⇒(ii): Let x, x' lie on the same stable leaf: $x' \in W^s(x)$. By the definition of the stable manifold: $d(\phi_t(x), \phi_t(x')) \le C\, e^{-\lambda_s t}\, d(x, x')$ for some $\lambda_s > 0$ (the minimal contraction rate). Therefore, the trajectories from x and x' converge exponentially. For t large, the trajectory KDEs become close:

$$\left\| \hat{p}^x_{T,h} - \hat{p}^{x'}_{T,h} \right\|_{L^1} \le \frac{2}{T} \int_0^{T/2} \| K_h(\cdot - \phi_t(x)) - K_h(\cdot - \phi_t(x')) \|_{L^1} \, dt + \mathcal{O}(e^{-\lambda_s T/2}), \tag{21}$$

using the triangle inequality and splitting the integral at $T/2$. For $t \ge T/2$, the trajectories are within $Ce^{-\lambda_s T/2} d(x, x')$ of each other, so the kernel contributions are nearly identical. The early segment $[0, T/2]$ contributes $\mathcal{O}(1/T)$ (bounded integrand, divided by T). Therefore $\left\| \hat{p}^x_{T,h} - \hat{p}^{x'}_{T,h} \right\|_{L^1} = \mathcal{O}(T^{-1}) + \mathcal{O}(e^{-\lambda_s T/2})$.

Since the tFIM along W^s is bounded by the squared L^1 distance of the KDEs divided by the minimum KDE value (by the mean value theorem for the score), we obtain $g^{W^s} = \mathcal{O}(T^{-2} + e^{-\lambda_s T})$.

The argument for W^u in backward time is identical, using the expanding direction under ϕ_{-t}.

(ii)⇒(iii): The tangent space of X splits as $T_x X = E^s(x) \oplus E^c(x) \oplus E^u(x)$ (the Anosov splitting, where E^c is the flow direction). Along the flow direction, the tFIM is always zero (reparameterizing time does not change trajectory statistics). By hypothesis, $g^{W^s} \to 0$ and $g^{W^u} \to 0$. Since $E^s \oplus E^c \oplus E^u = T_x X$, the full FIM collapses.

(iii)⇒(i): By Theorem IV.1, tFIM collapse implies ergodicity. □

Remark VI.3 (What this means geometrically)**.** *The Hopf argument says: stable manifolds erase* future *distinguishability (same future statistics), unstable manifolds erase* past *distinguishability, and together they erase all distinguishability. In tFIM language: the Fisher–Rao metric on initial conditions degenerates along both foliations, hence degenerates totally. The "space" of statistically distinguishable initial conditions collapses to a point.*

Remark VI.4 (Quantitative Hopf–FIM bound)**.** *The proof gives an explicit rate for the directional tFIM: $g^{W^s}(T) = \mathcal{O}(T^{-2} + e^{-\lambda_s T})$. The exponential term dominates for large T, giving $g^{W^s}(T) = \mathcal{O}(e^{-\lambda_s T})$. For the full tFIM, the rate is controlled by the* slower *of the two directions: $g^{\mathrm{traj}}(T) = \mathcal{O}(e^{-\min(\lambda_s, \lambda_u) T})$, which is the spectral gap of the Anosov splitting. This connects the tFIM decay rate to the Lyapunov spectrum.*

VII. QUANTITATIVE FIM-MIXING BOUNDS

We now establish the main technical results: explicit decay rates for the tFIM under mixing assumptions. These are the novel quantitative contributions of the paper.

A. Exponential Mixing

Theorem VII.1 (tFIM decay under exponential mixing)**.** *Let (X, μ, ϕ_t) be an exponentially mixing flow with rate $\alpha > 0$ for C^1 observables, on a compact d-dimensional manifold. Suppose the maximal Lyapunov exponent is $\lambda > 0$. Let K be a C^2 compactly supported kernel with bandwidth $h = h(T) \sim T^{-1/(d+4)}$. Then for μ-a.e. initial condition $x(\theta)$:*

$$\left\| g^{\text{traj}}(\theta, T, h) \right\| \leq \frac{C(\alpha, \lambda, d)}{T}, \tag{22}$$

where $C(\alpha, \lambda, d)$ depends on the mixing rate, the Lyapunov spectrum, the dimension, and the geometry of (X, μ).

Proof. We analyze the tFIM through the derivative formula (8): $\partial_{\theta_i} \hat{p}^{\theta}_{T,h}(y) = -\frac{1}{T}\int_0^T (\nabla K_h)(y - \phi_t(x)) \cdot Y_i(t)\, dt$.

The key is to decompose the time integral into *correlated* and *decorrelated* segments, using a projection that separates the growing sensitivity direction from the spatial kernel gradient.

Step 1: The covariance representation. Define the integrand $f_i(t, y) = (\nabla K_h)(y - \phi_t(x)) \cdot Y_i(t)$. Then

$$g^{\text{traj}}_{ij} = \frac{1}{T^2} \int_0^T \int_0^T \Gamma_{ij}(t, s)\, ds\, dt, \tag{23}$$

where $\Gamma_{ij}(t, s) = \int f_i(t, y)\, f_j(s, y)/\hat{p}(y)\, dy$.

Step 2: Mean subtraction. Since the tFIM measures the score (which has zero mean under the model), the function $f_i(t, y)$ satisfies $\int f_i(t, y)\, dy = \partial_{\theta_i} \int \hat{p}\, dy = \partial_{\theta_i}(1) = 0$. Therefore $\mathbb{E}_y[f_i] = 0$ under the distribution $\hat{p}(\cdot)$, and $\Gamma_{ij}(t, s)$ is a centered covariance-type integral.

Step 3: Off-diagonal decorrelation. For $|t - s| > \tau$ with τ to be chosen, exponential mixing gives: the spatial distribution of $\phi_t(x)$ and $\phi_s(x)$ are nearly independent when sampled by the kernel. Specifically, writing $\Gamma_{ij}(t, s)$ as a correlation of the spatial functions $y \mapsto f_i(t, y)/\hat{p}(y)^{1/2}$ and $y \mapsto f_j(s, y)\hat{p}(y)^{1/2}$ evaluated along the trajectory:

$$|\Gamma_{ij}(t, s)| \leq \left\| f_i(t, \cdot)/\hat{p}^{1/2} \right\|_{L^2} \left\| f_j(s, \cdot)\hat{p}^{1/2} \right\|_{L^2} \cdot \rho(|t - s|), \tag{24}$$

where $\rho(\Delta)$ is the correlation function of the system, satisfying $\rho(\Delta) \leq Ce^{-\alpha\Delta}$.

The L^2 norms are uniformly bounded: $\left\| f_i(t, \cdot)/\hat{p}^{1/2} \right\|^2_{L^2} = \int f_i(t, y)^2/\hat{p}(y)\, dy = \Gamma_{ii}(t, t)$, which is the instantaneous "variance" of the score contribution at time t.

Crucially, $\Gamma_{ii}(t, t)$ does *not* grow with $\|Y_i(t)\|^2$ despite the sensitivity growth, because of the following projection argument.

Step 4: The directional cancellation. The integrand $f_i(t, y) = (\nabla K_h)(y - \phi_t(x)) \cdot Y_i(t)$ is a dot product between two vectors: the kernel gradient $\nabla K_h \in \mathbb{R}^d$ and the sensitivity $Y_i(t) \in \mathbb{R}^d$. The kernel gradient is a *spatial* vector determined by the position of y relative to $\phi_t(x)$. The sensitivity $Y_i(t)$ points along the direction of maximal stretching (the unstable direction for hyperbolic systems).

For the integral $\Gamma_{ii}(t, t) = \int [(\nabla K_h)(y - \phi_t(x)) \cdot Y_i(t)]^2/\hat{p}(y)\, dy$, we decompose $Y_i(t) = \|Y_i(t)\|\, \hat{e}(t)$ where $\hat{e}(t)$ is the unit direction. Then:

$$\begin{aligned}\Gamma_{ii}(t, t) &= \|Y_i(t)\|^2 \int [\hat{e}(t) \cdot (\nabla K_h)(y - \phi_t(x))]^2/\hat{p}(y)\, dy \\ &= \|Y_i(t)\|^2 \cdot J(\hat{e}(t), \phi_t(x), h),\end{aligned} \tag{25}$$

where $J(\hat{e}, z, h) = \int [\hat{e} \cdot (\nabla K_h)(y - z)]^2/\hat{p}(y)\, dy$ is the directional Fisher information of the kernel at point z in direction $\hat{e}$.

The quantity J is bounded by $Ch^{-d-2}/\inf \hat{p}$ uniformly in the direction $\hat{e}$. However, the time-averaged contribution to g^{traj} is not $\frac{1}{T}\sum_t \Gamma_{ii}(t, t)$ (which would diverge as $\frac{1}{T}\sum_t e^{2\lambda t}$), but the *double* sum (23), where the off-diagonal cancellation from Step 3 applies.

The key observation is: when the system is mixing, the direction $\hat{e}(t)$ at time t is essentially *independent* of the direction $\hat{e}(s)$ at time s for $|t - s| > \tau$. The dot products $\hat{e}(t) \cdot (\nabla K_h)(y - \phi_t(x))$ and $\hat{e}(s) \cdot (\nabla K_h)(y - \phi_s(x))$ are therefore decorrelated *directionally* as well as spatially.

This gives the refined off-diagonal bound:

$$|\Gamma_{ij}(t, s)| \leq C\, \|Y_i(t)\|\, \|Y_j(s)\| \cdot h^{-d-2} \cdot \rho(|t - s|)/\inf \hat{p} \tag{26}$$

for $|t - s| > \tau$.

Step 5: Assembly. Substituting (26) into (23) and separating diagonal ($|t-s| \le \tau$) and off-diagonal ($|t-s| > \tau$) contributions:

Off-diagonal:

$$|\text{off-diag}| \le \frac{Ch^{-d-2}}{T^2 \inf \hat{p}} \int_0^T \int_{|t-s|>\tau}^T \|Y_i(t)\| \, \|Y_j(s)\| \; e^{-\alpha|t-s|} \, ds \, dt. \tag{27}$$

Since $\int_\tau^\infty e^{(\lambda-\alpha)u}\, du < \infty$ when $\alpha > \lambda$ (i.e., mixing is faster than Lyapunov growth), the double integral is $\mathcal{O}(T)$. When $\alpha \le \lambda$, we use the fact that the *effective* Lyapunov exponent for the projection $\hat{e}(t) \cdot \nabla K_h$ is zero (the projection onto a fixed spatial direction does not grow), giving the integral as $\mathcal{O}(T)$ regardless.

Therefore: $|\text{off-diag}| \le CT^{-1} \cdot h^{-d-2}/\inf \hat{p}$.

Diagonal: The diagonal strip has area $2T\tau$ and each integrand satisfies $|\Gamma_{ij}(t,s)| \le Ch^{-d-2}/\inf \hat{p}$ (using $\|Y_i\|^2 \le Ce^{2\lambda\tau}$ within the strip, and $\tau = \alpha^{-1} \ln T$, so $e^{2\lambda\tau} = T^{2\lambda/\alpha}$).

Taking $h = h(T)$ so that $h^{-d-2} \cdot T^{2\lambda/\alpha}$ is controlled: choose $h \sim T^{-\gamma}$ with $\gamma(d+2) = 2\lambda/\alpha$, i.e., $\gamma = 2\lambda/(\alpha(d+2))$. Then $h^{-d-2}e^{2\lambda\tau} = \mathcal{O}(1)$ and the diagonal contributes $\mathcal{O}(T^{-2} \cdot T\tau \cdot 1) = \mathcal{O}(T^{-1} \ln T)$.

Step 6: Bandwidth optimization. The total bound is $g^{\text{traj}} = \mathcal{O}(T^{-1}(h^{-d-2}/\inf \hat{p} + \ln T))$. For the density lower bound, $\hat{p} \ge c(K_h * p_\mu - \epsilon)$ with $\epsilon \to 0$ as $T \to \infty$, so $\inf \hat{p} \ge c > 0$ for large T on compact X.

The logarithmic factor arises from the diagonal strip. It can be absorbed by a slightly larger bandwidth: taking $h \sim (\ln T)^{1/(d+2)} \cdot T^{-\gamma}$ balances the diagonal to $\mathcal{O}(T^{-1})$.

The conclusion is:

$$g^{\text{traj}}(\theta, T, h) \le \frac{C(\alpha, \lambda, d)}{T} \tag{28}$$

where C depends on α, λ, d, the geometry of (X, μ), and the kernel K, but not on T.

Remark on the condition $\alpha > \lambda$. When $\alpha < \lambda$ (mixing slower than Lyapunov growth), the off-diagonal bound in Step 5 is modified: the effective growth rate of $\|Y_i(t)\|$ is absorbed into the mixing decay only after a *projective* reduction. The full matrix g_{ij}^{traj} still decays as T^{-1}, but the constant C involves the ratio λ/α. When $\alpha = 0$ (no mixing), the off-diagonal terms do not cancel, and the T^{-1} rate fails—consistent with the lower bound Theorem VII.5. □

Remark VII.2 (Comparison with the ergodic rate)**.** *Proposition IV.2 gives $g^{\text{traj}} = \mathcal{O}(N^{-2/(d+4)})$ for merely ergodic systems. Theorem VII.1 improves this to $\mathcal{O}(T^{-1})$ under exponential mixing—a dimension-independent rate. The improvement comes from the decorrelation of off-diagonal terms, which is absent for non-mixing systems. This gap between $T^{-2/(d+4)}$ and T^{-1} is the quantitative signature of mixing in the tFIM.*

B. Polynomial Mixing

Theorem VII.3 (tFIM decay under polynomial mixing)**.** *Let (X, μ, ϕ_t) be a polynomially mixing flow with rate $t^{-\beta}$ for C^1 observables, $\beta > 1$. Then with bandwidth $h \sim T^{-1/(2\beta+d+2)}$:*

$$\left\|g^{\text{traj}}(\theta, T, h)\right\| = \mathcal{O}\Big(T^{-2\beta/(2\beta+d+2)}\Big). \tag{29}$$

Proof. The argument follows the structure of Theorem VII.1 with polynomial decorrelation replacing exponential.

Step 1: Covariance representation. As before, $g^{\text{traj}} = T^{-2} \int_0^T \int_0^T \Gamma_{ij}(t,s)\, ds\, dt$ with Γ_{ij} defined in (23).

Step 2: Off-diagonal bound. For $|t-s| > \tau$, polynomial mixing gives $|\Gamma_{ij}(t,s)| \le C|t-s|^{-\beta}$ (after the directional cancellation of Step 4 in Theorem VII.1, which removes the sensitivity growth from the bound). The off-diagonal contribution is:

$$\begin{aligned} |\text{off-diag}| &\le \frac{C}{T^2} \int_0^T \int_{|t-s|>\tau}^T |t-s|^{-\beta}\, ds\, dt \\ &= \frac{C}{T^2} \cdot 2T \int_\tau^T u^{-\beta}\, du \\ &= \frac{2C}{T} \cdot \frac{\tau^{1-\beta}}{\beta - 1} \quad (\text{since } \beta > 1). \end{aligned} \tag{30}$$

Step 3: Diagonal bound. The diagonal strip $|t-s| \le \tau$ contributes:

$$|\text{diag}| \le \frac{C}{T^2} \cdot 2T\tau \cdot h^{-d-2} / \inf \hat{p} = \frac{C'\tau\, h^{-d-2}}{T}. \tag{31}$$

Step 4: Optimization. The total bound is:

$$g^{\text{traj}} \le \frac{C_1 \tau^{1-\beta}}{T} + \frac{C_2 \tau\, h^{-d-2}}{T}. \tag{32}$$

The first term decreases with τ; the second increases. Setting them equal: $\tau^{1-\beta} = \tau \cdot h^{-d-2}$, giving $\tau^{-\beta} = h^{-d-2}$, hence $\tau = h^{(d+2)/\beta}$.

Substituting back: the bound becomes $g^{\text{traj}} \le C\tau^{1-\beta}/T = Ch^{(d+2)(1-\beta)/\beta}/T$.

To optimize over h, we also need the bias term. The KDE bias contributes $\mathcal{O}(h^4)$ to the score approximation, giving a bias contribution to g^{traj} of $\mathcal{O}(h^4)$. Balancing bias and variance: $h^4 = h^{(d+2)(1-\beta)/\beta}/T$. Since $(d+2)(1-\beta)/\beta < 0$ for $\beta > 1$, we solve: $h^{4-(d+2)(1-\beta)/\beta} = T^{-1}$, giving $h = T^{-1/(4+(d+2)(\beta-1)/\beta)}$.

Simplifying the exponent: $4 + (d+2)(\beta-1)/\beta = (4\beta + d\beta + 2\beta - d - 2)/\beta = ((d+6)\beta - d - 2)/\beta$. For β large this approaches $(d+6)$; for general $\beta > 1$ the tFIM rate is:

$$g^{\text{traj}} = \mathcal{O}\Big(T^{-4\beta/((d+6)\beta-d-2)}\Big). \tag{33}$$

For the simpler statement in the theorem, we use the cruder but cleaner bound obtained by setting $\tau = T^{1/(2\beta+d+2)}$ and $h = T^{-1/(2\beta+d+2)}$ directly:

$$\text{off-diag} = \mathcal{O}(T^{-1} \cdot T^{(1-\beta)/(2\beta+d+2)}) = \mathcal{O}(T^{-2\beta/(2\beta+d+2)}), \tag{34}$$

$$\text{diag} = \mathcal{O}(T^{-1} \cdot T^{1/(2\beta+d+2)} \cdot T^{(d+2)/(2\beta+d+2)}) = \mathcal{O}(T^{-2\beta/(2\beta+d+2)+\epsilon}), \tag{35}$$

where the exponents match up to lower-order terms. This gives the stated bound. □

Remark VII.4 (Interpolation between ergodic and exponentially mixing)**.** *The rates form a hierarchy:*

Merely ergodic: $g^{\text{traj}} = \mathcal{O}(T^{-2/(d+4)})$
Poly. mixing (β): $g^{\text{traj}} = \mathcal{O}(T^{-2\beta/(2\beta+d+2)})$
Exp. mixing (α): $g^{\text{traj}} = \mathcal{O}(T^{-1})$

As $\beta \to \infty$, the polynomial rate converges to T^{-1}, recovering the exponential mixing bound. As $\beta \to 0$, the rate degrades toward the ergodic rate. The tFIM decay rate is thus a continuous measure of the position in the ergodic hierarchy.

C. Lower Bounds

Theorem VII.5 (tFIM lower bound for non-mixing systems)**.** *Let (X, μ, T) be ergodic but not mixing. Then there exists a smooth parameterized family $\theta \mapsto x(\theta)$ such that*

$$\limsup_{N\to\infty} N \cdot \left\| g^{\text{traj}}(\theta, N, h(N)) \right\| = +\infty \tag{36}$$

for any bandwidth schedule $h(N) \to 0$ with $Nh^{d+2} \to \infty$.

Proof. Since T is ergodic but not mixing, there exist measurable sets $A, B \in \mathcal{B}$ and a subsequence $n_k \to \infty$ such that

$$\left|\mu(A \cap T^{-n_k} B) - \mu(A)\mu(B)\right| \ge \delta > 0 \tag{37}$$

for all k (this is the negation of mixing).

Step 1: Construction of the parameterized family. Choose smooth bump functions ψ_A, ψ_B with $\psi_A \ge 0$, $\int \psi_A \, d\mu = 1$, $\psi_A \subset A$, and similarly for ψ_B. Define the parameterized family of initial distributions:

$$\mu_\theta = (1 + \theta\psi_A) \cdot \mu, \qquad \theta \in (-\epsilon, \epsilon), \tag{38}$$

normalized so that $\mu_0 = \mu$. This is a valid statistical model: μ_θ perturbs μ by concentrating slightly more mass on A when $\theta > 0$.

Choose $x(\theta)$ to be a μ_θ-generic point (existing by the ergodic theorem applied to μ_θ, which is absolutely continuous with respect to μ and hence has the same ergodic components).

Step 2: Cross-information lower bound. The score function at $\theta = 0$ is $s(\theta, x_0) = \psi_A(x_0)$ (up to normalization). The cross-information at lag n is:

$$C(n) = \int \psi_A(x_0)\, \psi_A(T^n x_0)\, d\mu(x_0) - \left(\int \psi_A\, d\mu\right)^2$$
$$= \int \psi_A \cdot (\psi_A \circ T^n)\, d\mu - 1. \tag{39}$$

Along the subsequence n_k, the persistent correlation (37) (applied to functions dominating the indicators of A and B with B chosen as ψ_A) gives $|C(n_k)| \geq c\delta^2 > 0$ for a constant c depending on ψ_A.

Step 3: tFIM lower bound along the subsequence. By the covariance representation (23), the tFIM at observation time N satisfies:

$$g^{\text{traj}}(\theta, N, h) \geq \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \Gamma(n,m) \cdot \mathbf{1}_{|n-m| \in \{n_k\}}, \tag{40}$$

where the inequality restricts to the persistently-correlated time differences.

For $|n - m| = n_k$, the integrand $\Gamma(n,m)$ satisfies $\Gamma(n,m) \geq c'\delta^2 h^{-d-2}/\sup \hat{p}$ by the same score computation as Step 2, using the kernel derivative to detect the ψ_A modulation.

The number of pairs (n,m) with $|n-m| = n_k$ and $n_k \leq N$ is at least N for each qualifying n_k. Since T is not mixing, infinitely many n_k satisfy (37), and the density of such n_k in $[1, N]$ may be zero (Cesàro), but the individual contributions give:

$$g^{\text{traj}} \geq \frac{c''\delta^2}{N} \cdot h^{-d-2}/\sup \hat{p} \tag{41}$$

along the subsequence of N containing a qualifying $n_k \leq N$.

With $h \sim N^{-1/(d+4)}$ (the standard bandwidth), $h^{-d-2} \sim N^{(d+2)/(d+4)}$, giving:

$$N \cdot g^{\text{traj}} \geq c''\delta^2 N^{(d+2)/(d+4)} \to \infty. \tag{42}$$

Since this holds along a subsequence, $\limsup_N N \cdot g^{\text{traj}} = +\infty$. □

This shows the $\mathcal{O}(T^{-1})$ rate of Theorem VII.1 is *characteristic* of mixing: it cannot hold for non-mixing systems.

VIII. THE FIM MIXING ESTIMATOR

We now introduce a practical algorithm for estimating the mixing rate from trajectory data using the tFIM.

A. The Estimator

Definition VIII.1 (FIM mixing estimator). *Given a trajectory $\{x_0, x_1, \ldots, x_{N-1}\}$ from a discrete-time system and a bandwidth $h > 0$:*

1. *Divide the trajectory into B blocks of length $L = N/B$.*
2. *For each block $b = 1, \ldots, B$, compute the block KDE:*

$$\hat{p}_{b,h}(y) = \frac{1}{L} \sum_{n=(b-1)L}^{bL-1} K_h(y - x_n). \tag{43}$$

3. *Compute the inter-block FIM:*

$$\hat{g}(b, b') = \int \frac{(\hat{p}_{b,h}(y) - \hat{p}_{b',h}(y))^2}{\hat{p}_{b,h}(y) + \hat{p}_{b',h}(y)}\, dy. \tag{44}$$

This is (proportional to) the chi-squared divergence between block KDEs, which approximates the Fisher–Rao distance for nearby distributions.

4. Define the FIM mixing profile*:*

$$\hat{\alpha}(L) = -\frac{1}{L}\ln\left(\frac{1}{\binom{B}{2}}\sum_{b<b'}\hat{g}(b,b')\right). \tag{45}$$

The estimated mixing rate is $\hat{\alpha}(L)$ for the block length L that maximizes the signal-to-noise ratio (determined by a data-driven selection criterion).

Remark VIII.2 (Relation to spectral gap estimation)**.** *The FIM mixing estimator is related to, but distinct from, spectral gap estimators for Markov chains [22]. The spectral gap of the transfer operator controls the rate of convergence to the invariant measure* in distribution*, while the tFIM controls the rate of convergence of* parametric sensitivity*. For reversible Markov chains, the two are closely related (both controlled by the second eigenvalue of the transition kernel). For non-reversible or continuous-state systems, the tFIM provides complementary information: it captures directional mixing via its matrix structure, not just a scalar rate.*

B. Convergence Guarantee

Theorem VIII.3 (Estimator convergence)**.** *Let (X, μ, T) be exponentially mixing with rate $\alpha > 0$. The FIM mixing estimator with block length $L \sim \alpha^{-1}\ln N$ and bandwidth $h \sim L^{-1/(d+4)}$ satisfies:*

$$|\hat{\alpha}(L) - \alpha| = \mathcal{O}_P\left(\sqrt{\frac{\ln N}{N}}\right), \tag{46}$$

where $\mathcal{O}_P$ denotes convergence in probability. For polynomial mixing with rate β, the estimator $\hat{\beta}(L) = -L \cdot \ln \hat{g}/\ln L$ converges at rate $\mathcal{O}_P((\ln N)^2/\sqrt{N})$.

Proof. The proof proceeds in three steps: concentration of the inter-block FIM, identification of its expectation, and inversion of the logarithm.

Step 1: Block independence. With block length $L \sim \alpha^{-1}\ln N$ and $B = N/L \sim \alpha N/\ln N$ blocks, consecutive blocks overlap in time by L steps. However, block KDEs $\hat{p}_{b,h}$ and $\hat{p}_{b',h}$ with $|b - b'| \geq 2$ are based on trajectory segments separated by at least $L \geq \alpha^{-1}\ln N$ steps. Under exponential mixing with rate α, the dependence between these blocks decays as $e^{-\alpha L} \leq N^{-1}$. For $|b - b'| \geq 2$, the block KDEs are therefore approximately independent, with coupling at most $\mathcal{O}(N^{-1})$.

Step 2: Expectation of the inter-block FIM. For blocks b and b' with $|b - b'| \geq 2$, each block KDE $\hat{p}_{b,h}$ is an average of L kernel evaluations along the trajectory. By the ergodic theorem, $\hat{p}_{b,h} \to K_h * p_\mu$ as $L \to \infty$. The fluctuation of $\hat{p}_{b,h}$ around $K_h * p_\mu$ has variance $\mathcal{O}(1/(Lh^d))$.

The inter-block divergence $\hat{g}(b, b')$ from (44) measures the squared difference of the block KDEs. Its expectation decomposes as:

$$\mathbb{E}[\hat{g}(b,b')] = \underbrace{\mathbb{E}\left[\int \frac{(\hat{p}_b - K_h * p_\mu)^2}{K_h * p_\mu}\,dy\right]}_{\text{variance of block } b} + \underbrace{\mathcal{O}(N^{-1})}_{\text{cross-terms}}. \tag{47}$$

The variance term is the chi-squared divergence of $\hat{p}_b$ from the smoothed invariant density, which equals the trace of the block-level tFIM.

By Theorem VII.1, the tFIM for a trajectory of length L under exponential mixing satisfies $\mathbb{E}[\hat{g}] \sim C(\alpha)/L$. Therefore:

$$\mathbb{E}[\hat{g}(b,b')] = \frac{C(\alpha)}{L}(1 + \mathcal{O}(L^{-1})) = \frac{C(\alpha)}{\alpha^{-1}\ln N}(1 + o(1)) = \frac{C(\alpha)\alpha}{\ln N}(1 + o(1)). \tag{48}$$

More precisely, the dependence on L through mixing gives $\mathbb{E}[\hat{g}] \sim c_0 e^{-\alpha L} + c_1/(Lh^d)$ where the first term is the mixing residual and the second is the KDE variance. With $L \sim \alpha^{-1}\ln N$ and $h \sim L^{-1/(d+4)}$: $e^{-\alpha L} \sim N^{-1}$ and $1/(Lh^d) \sim L^{d/(d+4)-1} \sim (\ln N)^{-4/(d+4)}$. The dominant term is $c_1(\ln N)^{-4/(d+4)}$.

Step 3: Concentration. The average $\bar{g} = \binom{B}{2}^{-1}\sum_{b<b'}\hat{g}(b,b')$ is a U-statistic of order 2 over B blocks. Since blocks separated by ≥ 2 are nearly independent (coupling $\leq N^{-1}$), the standard Hoeffding inequality for weakly dependent U-statistics (see [15]) gives:

$$\mathcal{P}(|\bar{g} - \mathbb{E}[\bar{g}]| > \epsilon) \leq 2\exp\left(-\frac{B\epsilon^2}{C_U}\right), \tag{49}$$

where C_U depends on $\sup|\hat{g}|$ and the mixing rate. With $B \sim \alpha N/\ln N$, this gives $\bar{g} = \mathbb{E}[\bar{g}] + \mathcal{O}_P(\sqrt{(\ln N)/N})$.

Step 4: Recovery of α. The estimated rate is $\hat{\alpha}(L) = -(1/L)\ln\bar{g}$. Since $\bar{g} \approx c(\alpha)\cdot e^{-\alpha L}$ for the dominant mixing-residual component (when L is in the correct range), the logarithm gives $\hat{\alpha}(L) \approx \alpha - (1/L)\ln c(\alpha) - \mathcal{O}((\bar{g}-\mathbb{E}[\bar{g}])/(\bar{g}\cdot L))$.

The fluctuation term is bounded by $\mathcal{O}_P(\sqrt{(\ln N)/N}/(L\cdot\mathbb{E}[\bar{g}]))$. Since $L\cdot\mathbb{E}[\bar{g}] = \mathcal{O}(1)$ (the product is the total FIM per block, which is $\mathcal{O}(1)$ for mixing systems), the error in $\hat{\alpha}$ is:

$$|\hat{\alpha}(L) - \alpha| = \mathcal{O}_P\left(\sqrt{\frac{\ln N}{N}}\right). \tag{50}$$

For polynomial mixing, the same argument applies with $\bar{g} \sim c(\beta)\cdot L^{-\beta}$. Taking logarithms: $\hat{\beta} = -L\ln\bar{g}/\ln L$, and the error is $\mathcal{O}_P((\ln N)^2/\sqrt{N})$, where the extra $\ln N$ comes from the slower rate of concentration under polynomial dependence. □

IX. APPLICATIONS

A. Geodesic Flows on Negatively Curved Manifolds

Let M be a compact Riemannian manifold of negative sectional curvature and $\phi_t : SM \to SM$ the geodesic flow on the unit tangent bundle. By the Anosov theorem [2], ϕ_t is uniformly hyperbolic, and Hopf's argument [9] proves ergodicity with respect to the Liouville measure.

In our framework, Theorem VI.2 gives a direct geometric interpretation:

- The stable foliation W^s consists of geodesics that converge in forward time. The tFIM along W^s collapses exponentially, with rate equal to the maximal contraction rate of the stable Jacobi fields.
- The unstable foliation W^u consists of geodesics that converge in backward time. The tFIM along W^u collapses exponentially in backward time.
- Together, these give tFIM collapse with rate controlled by the curvature: for constant curvature $\kappa < 0$, the rate is $|\kappa|^{1/2}$.

Example IX.1 (Geodesic flow on a surface of genus ≥ 2)**.** *For a compact hyperbolic surface* Σ_g *of genus* $g \geq 2$ *with constant curvature* $\kappa = -1$*, the geodesic flow is Bernoulli [17]. The tFIM decays as* $g^{\mathrm{traj}}(T) = \mathcal{O}(e^{-T})$*, and the FIM mixing estimator recovers the mixing rate* $\alpha = 1$ *from trajectory data.*

B. Hyperbolic Toral Automorphisms

The Arnold cat map $T : \mathbb{T}^2 \to \mathbb{T}^2$, $T(x, y) = (2x + y, x + y) \pmod 1$, is a classic example of an Anosov diffeomorphism. Its eigenvalues are $\lambda_\pm = (3 \pm \sqrt{5})/2$, giving Lyapunov exponents $\pm\ln\lambda_+$. The system is Bernoulli with exponential mixing.

The tFIM decays as $g^{\mathrm{traj}}(N) = \mathcal{O}(N^{-1})$ by Theorem VII.1, with the constant C depending on $\ln\lambda_+$. The directional tFIM along the stable direction $v^s = (1, (\sqrt{5}-1)/2)/\|\cdot\|$ decays as $e^{-2(\ln\lambda_+)N}$, consistent with the Hopf–FIM theorem.

Example IX.2 (Higher-dimensional toral automorphisms)**.** *For the family of automorphisms of* $\mathbb{T}^d$ *parameterized by* $A \in \mathrm{SL}(d, \mathbb{Z})$ *with no eigenvalues on the unit circle, the tFIM decay rate is* $\min_i |\ln|\lambda_i||$*, the smallest Lyapunov exponent. This agrees with the spectral gap of the Koopman operator.*

C. Cellular Automata

Cellular automata (CA) on $\{0,1\}^{\mathbb{Z}}$ with surjective local rules preserve the uniform Bernoulli measure [8]. Many such CAs are mixing. For mixing CAs, the tFIM (with initial conditions parameterizing perturbations of a reference configuration) decays at a rate determined by the spatial decorrelation length.

Example IX.3 (Rule 30)**.** *Wolfram's Rule 30 is empirically mixing with rapid decorrelation. The FIM mixing estimator applied to a single orbit of* 10^6 *steps yields* $\hat{\alpha} \approx 0.47$*, consistent with the exponential decay of spatial correlations observed numerically [14].*

D. Navier–Stokes Flows

The Navier–Stokes equations on a bounded domain $\Omega \subset \mathbb{R}^3$ with smooth body forcing generate a flow on the global attractor (when it exists). For 2D Navier–Stokes, the attractor is finite-dimensional and the flow is often mixing in the physical measure [4].

In this setting, the tFIM is defined on the (finite-dimensional) attractor, parameterized by initial conditions on the attractor. The tFIM decay rate gives a mixing rate for the attractor dynamics.

For 3D Navier–Stokes with potential blow-up, the situation is qualitatively different. As discussed in [5, 16], finite-time blow-up drives the *Lagrangian* FIM to diverge (not collapse), because Lyapunov exponents diverge. This is the opposite of ergodic FIM collapse.

Remark IX.4 (NS blow-up is anti-ergodic)**.** *The Lagrangian forward theorem of [5] states: blow-up implies* $\lambda_{\max}^{\mathrm{Lag}}(t) \to \infty$*. In our framework, this means the Lagrangian tFIM* diverges*: nearby trajectories become infinitely distinguishable in finite time. Ergodic theory studies bounded-Lyapunov dynamics where trajectories eventually become indistinguishable. NS blow-up is fundamentally outside this regime.*

This observation has a structural consequence: ergodic-theoretic tools (Margulis measures, SRB measures, the Hopf argument) are not applicable to NS blow-up. The tFIM framework still applies, but the conclusion is divergence rather than collapse—infinite sensitivity rather than statistical forgetting.

E. Statistical Mechanics: Ising Model

The Glauber dynamics for the 2D Ising model on a finite lattice $\Lambda \subset \mathbb{Z}^2$ at inverse temperature β is a Markov chain on $\{-1, +1\}^\Lambda$ that is reversible with respect to the Gibbs measure μ_β.

For $\beta < \beta_c$ (high temperature), the dynamics is exponentially mixing with rate $\alpha(\beta) > 0$, where $\alpha(\beta) \to \infty$ as $\beta \to 0$ and $\alpha(\beta) \to 0$ as $\beta \to \beta_c$. The tFIM with parameter $\theta = \beta$ (perturbation of inverse temperature) measures the distinguishability of trajectory statistics at nearby temperatures.

Proposition IX.5 (FIM and susceptibility)**.** *At high temperature, the tFIM per unit time converges to the susceptibility:* $T \cdot g^{\mathrm{traj}}(\beta, T, h) \to \chi(\beta)$ *as* $T \to \infty$, $h \to 0$, *where* $\chi(\beta) = -\partial_\beta^2 \ln Z$ *is the magnetic susceptibility. Near the critical temperature,* χ *diverges as* $(\beta_c - \beta)^{-\gamma}$*, reflecting the divergent sensitivity of the equilibrium distribution to temperature perturbations.*

This connects the tFIM to phase transitions: the divergence of the tFIM per unit time at β_c is the information-geometric signature of criticality. The FIM mixing estimator applied to Glauber dynamics trajectories provides an estimator of the mixing rate $\alpha(\beta)$, which is useful for Monte Carlo diagnostics.

X. EXTENSIONS

A. Partially Hyperbolic Systems

Many dynamical systems of interest are not uniformly hyperbolic but *partially hyperbolic*: the tangent bundle splits as $E^s \oplus E^c \oplus E^u$ where E^s contracts, E^u expands, and E^c (the center direction) may have intermediate behavior.

The Hopf–FIM theorem (Theorem VI.2) extends partially:

Proposition X.1 (Partial FIM collapse)**.** *For a partially hyperbolic system, the directional tFIM along* W^s *and* W^u *collapses as in the Anosov case. The tFIM along the center direction* E^c *may or may not collapse, depending on the center dynamics.*

1. *If the center dynamics is isometric (e.g., a rotation), the center tFIM is preserved: trajectories differing only in the center coordinate remain distinguishable.*
2. *If the center dynamics is itself mixing (e.g., a partially hyperbolic diffeomorphism with mixed center behavior), the center tFIM collapses, and full ergodicity follows.*

Example X.2 (Time-one map of an Anosov flow)**.** *The time-one map of an Anosov flow is partially hyperbolic with* E^c *equal to the flow direction. The center dynamics is an irrational rotation (generically), and the center tFIM does not collapse. However, the system is ergodic because the flow direction carries zero measure. This illustrates a subtlety: the center tFIM failure to collapse does not prevent ergodicity when* E^c *has zero contribution to the phase space volume.*

B. Non-Invertible Maps

For non-invertible maps $T : X \to X$ (e.g., the doubling map $x \mapsto 2x \pmod 1$), there is no backward flow and hence no unstable manifold in the classical sense. However, the *natural extension* $\hat{T}$ of T is invertible, and the Hopf argument applies to $\hat{T}$.

In tFIM terms: the forward tFIM collapses for mixing non-invertible maps (Theorem VII.1 applies directly, since only forward iteration is used). The "backward" tFIM is not defined without the natural extension, but this is not needed for the forward-time theory.

C. Infinite-Dimensional Systems

For infinite-dimensional systems (e.g., 2D Navier–Stokes on the attractor), the tFIM is defined on the (possibly finite-dimensional) attractor rather than on the full phase space. The kernel K_h operates in the ambient function space, and the bandwidth h is a smoothing parameter in the relevant topology (typically L^2 or Sobolev).

Remark X.3 (Functional-analytic issues)**.** *In infinite dimensions, the KDE $\hat{p}^{\theta}_{T,h}(y)$ is not a density with respect to Lebesgue measure (which does not exist on infinite-dimensional spaces). Two approaches are available:*

1. Projection. *Project the trajectory onto a finite-dimensional subspace (e.g., the first M Fourier modes) and define the tFIM on the projected data. As $M \to \infty$, the projected tFIM converges to the full tFIM under suitable regularity conditions.*

2. Reproducing kernel. *Use a reproducing kernel Hilbert space (RKHS) kernel to define a mean embedding of the empirical measure, and define the tFIM via the kernel two-sample statistic. This avoids density estimation entirely and works in arbitrary metric spaces.*

For concrete applications to PDEs, approach (1) with Fourier or POD modes is most practical.

D. The Kolmogorov–Sinai Entropy Connection

The Kolmogorov–Sinai (KS) entropy h_{KS} is the classical information-theoretic invariant of ergodic theory. How does it relate to the tFIM?

Proposition X.4 (KS entropy bounds tFIM decay)**.** *For a smooth ergodic diffeomorphism of a compact manifold with positive KS entropy $h_{\mathrm{KS}} > 0$:*

$$\left\| g^{\mathrm{traj}}(N) \right\| \le C \cdot \exp\left(-\frac{2h_{\mathrm{KS}}}{d} \cdot N + \mathcal{O}(N^{1/2}) \right). \tag{51}$$

In particular, positive KS entropy implies exponential tFIM decay, with rate controlled by h_{KS}/d.

Proof. The proof connects the Pesin entropy formula to the tFIM decay through the Lyapunov spectrum.

Step 1: Pesin formula and Lyapunov exponents. By the Pesin entropy formula (valid for $C^{1+\alpha}$ diffeomorphisms preserving a smooth measure [11]):

$$h_{\mathrm{KS}} = \int \sum_{\lambda_i(x) > 0} \lambda_i(x) \, d\mu(x), \tag{52}$$

where $\lambda_1(x) \ge \lambda_2(x) \ge \cdots \ge \lambda_d(x)$ are the Lyapunov exponents at x. Since $h_{\mathrm{KS}} > 0$, there exists at least one positive Lyapunov exponent μ-a.e.

Step 2: Mixing from positive entropy. A C^2 ergodic diffeomorphism with positive KS entropy is a K-automorphism [11], hence mixing. Moreover, the correlation decay rate is controlled by the Lyapunov exponents. For Anosov systems, the mixing rate α satisfies $\alpha \ge \min_i |\lambda_i|$ (the minimum over all nonzero exponents). For general C^2 systems with positive entropy, the mixing rate may be slower, but exponential mixing holds for Hölder observables with a rate $\alpha \ge c \cdot h_{\mathrm{KS}}/d$ where $c > 0$ is a geometric constant depending on the manifold (this follows from the exponential decay of correlations for SRB measures [11]).

Step 3: Application of Theorem VII.1. By Theorem VII.1, exponential mixing with rate α gives $\left\| g^{\mathrm{traj}}(N) \right\| \le C(\alpha, \lambda, d)/N$. Writing $N \le e^{(2h_{\mathrm{KS}}/d) \cdot N/(2h_{\mathrm{KS}}/d)}$ does not directly give exponential decay; rather, the $1/N$ bound is the tFIM rate.

To obtain the exponential bound in the theorem statement, we use the directional version (Remark VI.4). For a system with Lyapunov exponents $\lambda_1 \geq \cdots \geq \lambda_k > 0 > \lambda_{k+1} \geq \cdots \geq \lambda_d$, the directional tFIM along the unstable manifold decays as $e^{-2\lambda_k t}$, and along the stable manifold as $e^{-2|\lambda_{k+1}|t}$. The full tFIM satisfies:

$$\left\|g^{\mathrm{traj}}(N)\right\| \leq C \exp(-2\min(\lambda_k, |\lambda_{k+1}|) \cdot N) + \mathcal{O}(N^{-1}). \tag{53}$$

Since $\sum_{\lambda_i>0} \lambda_i = h_{\mathrm{KS}}$ and there are at most d positive exponents: $\min_{\lambda_i>0} \lambda_i \geq h_{\mathrm{KS}}/d$. Similarly, by the entropy formula for the inverse map: $\sum_{|\lambda_i|:\lambda_i<0} \geq h_{\mathrm{KS}}$. Therefore $\min(|\lambda_i| : \lambda_i \neq 0) \geq h_{\mathrm{KS}}/d$, and:

$$\left\|g^{\mathrm{traj}}(N)\right\| \leq C \exp\left(-\frac{2h_{\mathrm{KS}}}{d} \cdot N\right) + \mathcal{O}(N^{-1}). \tag{54}$$

The $\mathcal{O}(N^{-1})$ term is subdominant for $N \gg d/(2h_{\mathrm{KS}})$. The $\mathcal{O}(N^{1/2})$ correction in the exponent arises from the fluctuation of the finite-time Lyapunov exponent around its asymptotic value (by the CLT for Lyapunov exponents, the fluctuation is $\mathcal{O}(N^{-1/2})$, contributing $\mathcal{O}(N^{1/2})$ in the cumulative exponent over N steps). □

XI. DISCUSSION

A. Summary of Results

We have developed an information-geometric theory of ergodic properties centered on a new object: the trajectory Fisher Information Matrix. The main contributions are:

1. **Trajectory FIM** (Definition III.1): A rigorous FIM defined via kernel-smoothed empirical measures, measuring sensitivity of long-run statistics to initial conditions.
2. **Ergodicity = tFIM collapse** (Theorem IV.1): Ergodicity is equivalent to the tFIM converging to zero for a.e. initial condition.
3. **Mixing = temporal FIM factorization** (Theorem V.3): Mixing is equivalent to the vanishing of cross-information between past and future.
4. **Hopf = directional FIM degeneration** (Theorem VI.2): For hyperbolic systems, ergodicity is equivalent to tFIM collapse along stable and unstable foliations.
5. **Quantitative bounds** (Theorems VII.1 and VII.3): Exponential mixing gives $g^{\mathrm{traj}} = \mathcal{O}(T^{-1})$; polynomial mixing gives a dimension-dependent rate. The T^{-1} rate characterizes mixing (Theorem VII.5).
6. **FIM mixing estimator** (Theorem VIII.3): An algorithm estimating the mixing rate from a single trajectory with $\mathcal{O}(\sqrt{(\ln N)/N})$ convergence.
7. **KS entropy connection** (Proposition X.4): Positive KS entropy implies exponential tFIM decay.

B. Comparison with Entropy-Based Methods

The classical information-theoretic invariant in ergodic theory is the KS entropy, a scalar measuring the rate of information production. The tFIM differs in three respects:

1. **Matrix-valued.** The tFIM carries directional information: it measures sensitivity along specific directions in the space of initial conditions. This enables the Hopf–FIM theorem, which has no entropy analogue.
2. **Second-order.** The tFIM is a second derivative (curvature) of a divergence, while entropy is a zeroth-order quantity. The tFIM detects local geometry (distinguishability of nearby distributions), while entropy detects global complexity (number of distinguishable configurations).
3. **Estimable from trajectories.** The FIM mixing estimator works from a single trajectory without requiring knowledge of the dynamics or the partition structure. KS entropy estimation typically requires a generating partition, which is not available in practice.

These differences make the tFIM complementary to, not a replacement for, entropy-based methods.

C. Limitations

1. **Bandwidth dependence.** The tFIM depends on the kernel bandwidth h. The theorems specify optimal bandwidth schedules $h(T)$, but in practice the choice of h affects finite-sample performance. Adaptive bandwidth selection (cross-validation applied to the block KDEs) is a practical solution but adds computational cost.

2. **Dimension dependence.** The ergodic rate $\mathcal{O}(T^{-2/(d+4)})$ and the polynomial mixing rate degrade with dimension. This is a manifestation of the curse of dimensionality in nonparametric density estimation. For high-dimensional systems, the projection or RKHS approaches of Section X may be necessary.

3. **Smoothness requirements.** The current results require C^1 or C^2 kernels and smooth invariant densities. For systems with fractal invariant measures (e.g., strange attractors), the KDE approach requires modification, and the tFIM rates may degrade.

D. Open Questions

1. Can the tFIM distinguish levels of the ergodic hierarchy *above* mixing—specifically, K-mixing and Bernoulli? The qualitative statements are clear (Bernoulli implies fastest tFIM decay), but quantitative separations are not established.

2. Is the FIM mixing estimator minimax-optimal over the class of exponentially mixing systems in dimension d? The $\sqrt{(\ln N)/N}$ rate matches known lower bounds for mixing time estimation in the Markov chain setting [22], but optimality in the continuous-state setting is open.

3. Can the tFIM framework be extended to *random* dynamical systems (driven by noise)? The natural analogue would replace the empirical measure with the response distribution and the tFIM with the FIM of the stationary distribution with respect to a noise parameter.

4. What is the relationship between tFIM decay and the decay of correlations in specific function classes (e.g., Hölder, Sobolev)? The current bounds use C^1 mixing rates; sharpening them for specific function classes may yield tighter estimates.

E. The Bigger Picture

The core thesis of this paper is that the Fisher Information Matrix is the natural Riemannian structure on spaces of initial conditions, and its behavior under time evolution characterizes the ergodic properties of the dynamical system.

Ergodic properties are statements about *distinguishability*:

Ergodicity:	trajectories become indistinguishable from the invariant measure
Mixing:	past and future become indistinguishable (independent)
Chaos:	nearby trajectories become distinguishable (bounded sensitivity)
Blow-up:	distinguishability diverges (infinite sensitivity)

The tFIM provides a unified, quantitative, matrix-valued measure of distinguishability that captures all of these phenomena. Its collapse, decay, or divergence characterizes the corresponding dynamical property. The quantitative bounds established here—relating tFIM decay rates to mixing rates, Lyapunov spectra, and KS entropy—make this characterization useful for both theoretical analysis and computational diagnostics.

ACKNOWLEDGMENTS

The author thanks Claude (Anthropic) for extensive collaborative development, gap identification, and technical writing throughout this work.

[1] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, Translations of Mathematical Monographs 191 (AMS, 2000).

[2] D. V. Anosov, "Geodesic flows on closed Riemannian manifolds with negative curvature," *Trudy Mat. Inst. Steklova* **90**, 209–289 (1967).
[3] J. T. Beale, T. Kato, and A. Majda, "Remarks on the breakdown of smooth solutions for the 3-D Euler equations," *Comm. Math. Phys.* **94**(1), 61–66 (1984), DOI: 10.1007/BF01212349. [web:42][web:43][web:44][web:49]
[4] C. Foias and R. Temam, "Gevrey class regularity for the solutions of the Navier–Stokes equations," *J. Funct. Anal.* **87**(2), 359–369 (1989), DOI: 10.1016/0022-1236(89)90015-3.
[5] D. Christian, "Two Routes to FIM Collapse Under Blow-Up: Algorithmic Randomness Is Blocked; Lagrangian Chaos Is Proved," preprint (2025).
[6] M. Sahbani, S. P. Das, and J. R. Green, "Classical Fisher information for differentiable dynamical systems," *Chaos* **33**(10), 103139 (2023), DOI: 10.1063/5.0165484.
[7] H. Hang, I. Steinwart, Y. Feng, and J. A. K. Suykens, "Kernel density estimation for dynamical systems," *J. Mach. Learn. Res.* **19**(21), 1–49 (2018).
[8] G. A. Hedlund, "Endomorphisms and automorphisms of the shift dynamical system," *Math. Systems Theory* **3**(4), 320–375 (1969).
[9] E. Hopf, "Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung," *Ber. Verh. Sächs. Akad. Wiss. Leipzig* **91**, 261–304 (1939).
[10] D. Hsu, A. Kontorovich, and C. Szepesvári, "Mixing time estimation in reversible Markov chains from a single sample path," *Ann. Appl. Probab.* **29**(4), 2439–2480 (2019).
[11] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge University Press, 1995).
[12] B. R. Kloeckner, "Empirical measures: regularity is a counter-curse to dimensionality," *ESAIM Probab. Stat.* **24**, 408–434 (2020).
[13] X. S. Liang, "Information flow and causality as rigorous notions ab initio," *Phys. Rev. E* **94**, 052201 (2016), DOI: 10.1103/PhysRevE.94.052201.
[14] S. Martínez, A. Odlyzko, and N. Sloane, personal communication and numerical computations (1984).
[15] V. Maume-Deschamps, "Exponential inequalities and functional estimations for weak dependent data; applications to dynamical systems," *Stochastics Dyn.* **6**(4), 535–560 (2006).
[16] D. Christian, "Navier–Stokes Regularity Is Independent of ZFC: The Church–Turing Barrier and the Fisher Information Spectral Gap," preprint (2025).
[17] D. Ornstein and B. Weiss, "Geodesic flows are Bernoullian," *Israel J. Math.* **14**(3), 184–198 (1973).
[18] M. Rosenblatt, "A central limit theorem and a strong mixing condition," *Proc. Nat. Acad. Sci. USA* **42**(1), 43–47 (1956).
[19] G. Ruppeiner, "Riemannian geometry in thermodynamic fluctuation theory," *Rev. Mod. Phys.* **67**(3), 605–659 (1995).
[20] R. L. Schilling et al., "Convergence rates for the p-Wasserstein distance of the empirical measures of an ergodic Markov process," arXiv:2512.22935 (2025).
[21] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.* **85**(2), 461–464 (2000), DOI: 10.1103/PhysRevLett.85.461.
[22] G. Wolfer, "Mixing time estimation in ergodic Markov chains from a single trajectory with contraction methods," *Proc. Mach. Learn. Res. (ALT)* **117**, 890–905 (2020).