Supplemental Material of "Causal Fisher-Information Inequalities: Classical Causal Model Falsification and Metrological Advantage"

Jeongho Bang[1,2,*] and Su-Yong Lee[3,4,†]

[1]*Institute for Convergence Research and Education in Advanced Technology, Yonsei University, Seoul 03722, Republic of Korea*
[2]*Department of Quantum Information, Yonsei University, Incheon 21983, Republic of Korea*
[3]*Agency for Defense Development, Daejeon 34186, Republic of Korea*
[4]*Weapon Systems Engineering, ADD school, University of Science and Technology, Daejeon, 34060, Republic of Korea*

(Dated: June 8, 2026)

This supplemental material (SM) is written as a self-contained and proof-complete companion to our main submitting manuscript. It provides full derivations, rigorous proofs, and extended methodological details required for a complete understanding of our work.

CONTENTS

* jbang@yonsei.ac.kr
† suyong2@add.re.kr

SUPPLEMENTARY NOTE I. INTRODUCTION: THE WORK SUMMARY

Precision parameter estimation is central to modern science and engineering. In a metrological task, an unknown parameter θ (e.g., a time interval, phase shift, rotation angle, field strength, or concentration) is encoded into measurement statistics, and one aims to infer θ from observed data. A standard operational quantifier of statistical sensitivity is the Fisher information (FI) [1]. For a measurement outcome x with conditional probability $p(x|\theta)$, the FI is defined by

$$F(\theta) = \sum_x p(x|\theta) \left(\frac{\partial}{\partial\theta} \ln p(x|\theta) \right)^2 . \tag{1}$$

In the asymptotic limit, FI governs achievable precision through the Cramér–Rao bound [2, 3].

Quantum metrology seeks to exploit genuine quantum features to enhance estimation precision beyond what is achievable by "most classical" strategies. In the conventional framework, the ultimate bound is often characterized in terms of quantum Fisher information (QFI), which is defined by optimizing over all measurements [4, 5]. However, the measurement that attains the bound can be difficult to identify or experimentally infeasible to implement in realistic settings, motivating complementary approaches based on operational and experimentally accessible criteria. A particularly appealing direction is to employ FI-based criteria, including FI inequalities, as experimentally accessible witnesses of nonclassical resources and metrological advantage [6–9].

A representative example is a classical FI inequality that appears when one estimates a total parameter either directly or by dividing the process into intermediate segments. Let us consider a parameter decomposition as

$$\theta_{ab} = \theta_{ac} + \theta_{cb}, \tag{2}$$

which is naturally associated with an intermediate description of a process. Under the classical assumptions that amount to a causal-path structure together with statistical independence of the estimation parameters, the following inequality is satisfied:

$$F(\theta_{ab})^{-1} \geq F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}. \tag{3}$$

One may define the violation statistic

$$V := F(\theta_{ab})^{-1} - F(\theta_{ac})^{-1} - F(\theta_{cb})^{-1}, \tag{4}$$

and an observed value $V < 0$ indicates a violation of Eq. (3). Such violations arise, for instance, in simple quantum dynamics when one attempts to characterize the evolution with a specific sequence of intermediate states. Crucially, the correct interpretation of a violation is not "quantumness in general" but the failure of the underlying classical assumptions used to derive Eq. (3).

In this work, we elevate this observation into a general principle and propose a unified framework: FI inequalities of the type in Eq. (3) are most naturally understood as tests of classical causal models. The bridge is provided by Bayesian networks and conditional independence. A classical causal model is specified by a directed acyclic graph (DAG) (possibly with latent variables), which implies a family of conditional-independence relations [10]. When a metrological protocol is assumed to admit a description within such a model class, those conditional-independence relations translate into constraints on the achievable FI. Thus, violating an FI inequality is equivalent to falsifying the corresponding classical causal model class. To emphasize the generality, we introduce the notion of *causal Fisher-information inequalities* (CFIIs): given a classical causal model class $\mathcal{M}$ defined by a DAG-induced set of conditional independences, CFIIs are the FI constraints that must hold for any experiment whose statistics admit a realization within $\mathcal{M}$. The trajectory-based inequality in Eq. (3) is then a special case associated with a causal-path (Markov-chain-like) model class.

Theorem S1 (Causal Fisher-information inequalities (informal statement))**.** *Consider a family of parameterized experiments specified by conditional distributions $\{p(x|s,\theta)\}$, where s denotes an experimental context (measurement choice, intermediate intervention, segmentation strategy, or control setting). Let $\mathcal{M}$ be a classical causal model class defined by a DAG that imposes conditional-independence constraints among the relevant variables, together with regularity assumptions ensuring that the parameter dependence factorizes consistently with the model. Then, there exists a family of inequalities among the corresponding Fisher informations that must be satisfied by any data compatible with $\mathcal{M}$. Any experimental violation of these inequalities certifies that no model in the class $\mathcal{M}$ can reproduce the observed statistics.*

This causal-model perspective has two immediate consequences. First, it provides a systematic route to generalization: rather than focusing on a specific hypothesis such as a discrete trajectory decomposition, one may start from an arbitrary causal model class $\mathcal{M}$, derive its associated CFIIs, and use experimental data to test (and potentially falsify) $\mathcal{M}$. Second, it reframes nonclassical resources in metrology: what appears as contextuality, non-Markovianity, or incompatibility can be interpreted as the collapse of a classical causal description, and such collapse can be harnessed as an information-theoretic resource.

SUPPLEMENTARY NOTE II. GENERAL FRAMEWORK

Here we present a general operational framework that unifies (i) standard parameter estimation, (ii) experiments involving multiple contexts (measurement choices, interventions, segmentations), and (iii) classical causal model classes specified by conditional-independence constraints. Throughout, we focus on a single real parameter θ for clarity. The generalization to multi-parameter estimation can be formulated by replacing FI with the Fisher information matrix [11, 12].

A. Operational setting: parameterized experiments with contexts

We consider an experiment in which an unknown parameter θ is encoded into outcome statistics. The experimenter can choose an experimental context (or setting) $s \in \mathcal{S}$, where s may represent a measurement choice, a control operation, an intermediate intervention, or a segmentation strategy. For each context s, the experiment produces an outcome x in an outcome set $\mathcal{X}_s$ according to a conditional distribution

$$p(x|s,\theta). \tag{5}$$

We allow $\mathcal{X}_s$ to depend on s to incorporate, for example, different output alphabets for different measurement settings. For notational simplicity, we present the formulas for discrete outcomes; continuous outcomes are treated by replacing sums with integrals.

A data set consists of independent samples collected under possibly multiple contexts. Let $\mathcal{S}_D \subseteq \mathcal{S}$ denote the set of contexts used in the protocol. For each $s \in \mathcal{S}_D$, we collect N_s independent outcomes

$$x_{s,1}, x_{s,2}, \cdots, x_{s,N_s}. \tag{6}$$

The total number of samples is

$$N = \sum_{s\in\mathcal{S}_D} N_s. \tag{7}$$

Assuming conditional independence of samples given (s, θ), the likelihood function is

$$\mathcal{L}(\theta|D) = \prod_{s\in\mathcal{S}_D} \prod_{i=1}^{N_s} p(x_{s,i}|s,\theta), \tag{8}$$

and the log-likelihood is

$$\ell(\theta|D) = \sum_{s\in\mathcal{S}_D} \sum_{i=1}^{N_s} \ln p(x_{s,i}|s,\theta). \tag{9}$$

An estimator is a function of the data, denoted as $\hat{\theta} = \hat{\theta}(D)$.

B. Fisher information under multi-context sampling

For each context s, the Fisher information is defined by [1]

$$F_s(\theta) = \sum_{x\in\mathcal{X}_s} p(x|s,\theta) \left(\frac{\partial}{\partial\theta} \ln p(x|s,\theta) \right)^2. \tag{10}$$

The quantity $F_s(\theta)$ characterizes the local sensitivity of the outcome distribution for context s with respect to θ.

When data are collected under multiple contexts, the relevant statistical model is the product distribution associated with Eq. (8). The corresponding Fisher information of the full data set, denoted by $F_D(\theta)$, is obtained from the score $\partial_\theta \ell(\theta|D)$. A basic property of FI is the additivity for independent samples, which yields a simple decomposition across contexts [1].

Theorem S2 (Additivity of Fisher information across contexts)**.** *Assume that conditioned on (s, θ) all outcomes $\{x_{s,i}\}$ are independent and distributed as $p(x|s, \theta)$. Then, the Fisher information of the full data set D satisfies*

$$F_D(\theta) = \sum_{s\in\mathcal{S}_D} N_s F_s(\theta). \tag{11}$$

Proof. —From Eq. (9),

$$\frac{\partial}{\partial\theta} \ell(\theta|D) = \sum_{s\in\mathcal{S}_D} \sum_{i=1}^{N_s} \frac{\partial}{\partial\theta} \ln p(x_{s,i}|s,\theta). \tag{12}$$

Taking the expectation over the product distribution and using the independence, the cross terms vanish because the score has zero mean under standard regularity conditions. Thus, one obtains Eq. (11). □

Theorem S2 shows that the sample allocation $\{N_s\}$ determines the effective FI through a linear combination of $\{F_s(\theta)\}$. It is convenient to introduce weights $q_s := N_s/N$, so that

$$\frac{1}{N} F_D(\theta) = \sum_{s\in\mathcal{S}_D} q_s F_s(\theta). \tag{13}$$

In the asymptotic regime, an unbiased estimator $\hat{\theta}$ satisfies the Cramér–Rao bound [2, 3]

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{F_D(\theta)}. \tag{14}$$

Hence, for a fixed total budget N, the performance comparisons between protocols must specify the sample allocation across contexts.

C. Quantum realization of the conditional model

The operational model in Eq. (5) includes the quantum experiments as a special case. Let the probe be prepared in a quantum state $\hat{\rho}(\theta)$ that depends on θ. A context s specifies a measurement (and possibly additional control operations), represented by a POVM $\{\hat{M}_{x|s}\}_{x\in\mathcal{X}_s}$ satisfying $\hat{M}_{x|s} \geq 0$ and $\sum_{x\in\mathcal{X}_s} \hat{M}_{x|s} = \hat{\mathbb{1}}$. Here, the outcome statistics are given by the Born rule,

$$p(x|s,\theta) = \mathrm{Tr}\Big[\hat{M}_{x|s}\hat{\rho}(\theta)\Big]. \tag{15}$$

For a pure-state model, $\hat{\rho}(\theta) = |\psi(\theta)\rangle\langle\psi(\theta)|$.

Sequential or adaptive protocols can also be included by enlarging the context and outcome spaces. For example, a consecutive measurement producing an outcome pair (a, b) can be treated as a single outcome $x = (a, b)$ for a context s that specifies the corresponding measurement instrument. Thus, the multi-context framework in Sec. II A naturally accommodates both conventional single-setting metrology and protocols that integrate data from multiple settings.

D. Classical causal models with conditional independence

We next formalize the notion of a classical causal explanation of experimental statistics. A classical causal model is specified by a directed acyclic graph (DAG) G whose nodes correspond to random variables [10]. Let $\mathcal{V} = \{X_1, X_2, \cdots, X_n\}$ denote the set of model variables, and let $\mathrm{pa}(X_j)$ denote the set of parents of node X_j in G. The defining property of a Bayesian network is the factorization of the joint distribution:

$$p(x_1, x_2, \cdots, x_n) = \prod_{j=1}^{n} p(x_j|\mathrm{pa}(x_j)). \tag{16}$$

The factorization implies conditional-independence (CI) relations. For three sets of variables U, V and W, the CI statement $U \perp V|W$ means

$$p(u, v|w) = p(u|w)p(v|w), \tag{17}$$

for all values u, v and w with $p(w) > 0$. A DAG G induces a collection of CI constraints (e.g., via d-separation), and the set of distributions satisfying Eq. (16) is equivalently characterized by these constraints together with positivity conditions.

In metrology, the experimenter controls the context S and the parameter θ is unknown. To connect the causal models to the operational distributions $p(x|s,\theta)$, we treat S as an exogenous variable that can be set by the experimenter, and we allow latent variables Λ to represent unobserved degrees of freedom. A parameterized causal model generates a family of conditional distributions $p(v|s,\theta)$ over $\mathcal{V}$. In the most general form considered here, the model assumes the existence of Λ such that

$$p(v, \lambda|s,\theta) = p(\lambda) \prod_{j=1}^{n} p(x_j|\mathrm{pa}(x_j), s, \theta, \lambda), \tag{18}$$

where $v = (x_1, \cdots, x_n)$ and λ is a value of Λ. The operational distribution $p(x|s,\theta)$ of an observed outcome variable X is obtained by marginalization over unobserved variables.

Definition S1 (Causal model compatibility)**.** *Let X be the observed outcome variable and let $p(x|s,\theta)$ be an operational model. We say that $p(x|s,\theta)$ is compatible with a classical causal model class $\mathcal{M}(G)$ if there exist (possibly latent) variables $\mathcal{V} \setminus \{X\}$ and Λ and a family of joint distributions $p(v, \lambda|s,\theta)$ that factorize as in Eq. (18) for the DAG G, such that the induced marginal equals the operational model for all contexts and parameter values,*

$$p(x|s,\theta) = \sum_{v\setminus x} \sum_{\lambda} p(v, \lambda|s,\theta). \tag{19}$$

Definition S1 formalizes the statement that observed statistics admit a classical causal explanation within a specified model class. Different choices of G and different restrictions on how θ enters Eq. (18) define different notions of classicality [10, 13, 14]. In particular, trajectory-based (Markov-chain-like) models and context-free (noncontextual) models correspond to different CI structures and lead to different testable constraints.

E. Causal Fisher-information inequality framework

Given a causal model class $\mathcal{M}$, the set of compatible operational models $\{p(x|s,\theta)\}$ is constrained by the CI relations implied by $\mathcal{M}$. Since FI is a functional of $p(x|s,\theta)$, these constraints induce relations among the Fisher informations $\{F_s(\theta)\}$. This motivates the following general definition.

Definition S2 (Causal Fisher-information inequality)**.** *Fix a causal model class $\mathcal{M}$ and a finite set of contexts $\mathcal{S}_0 \subseteq \mathcal{S}$. A causal Fisher-information inequality (CFII) is an inequality of the form*

$$\mathcal{G}\left(\{F_s(\theta)\}_{s\in\mathcal{S}_0}\right) \geq 0 \tag{20}$$

that holds for all θ and for every operational model $p(x|s,\theta)$ compatible with $\mathcal{M}$ in the sense of Definition S1.

CFIIs are model-dependent: the function $\mathcal{G}$ depends on the causal assumptions encoded in $\mathcal{M}$. In the following, we derive explicit CFIIs for several broad classes of causal models. The central operational message is that if experimentally estimated Fisher informations violate a CFII, then the observed statistics are incompatible with the assumed classical causal model class. This provides a general route to interpret FI-inequality violations as falsification of classical causal explanations, and it sets the stage for viewing such "causal-model collapse" as a metrological resource.

SUPPLEMENTARY NOTE III. CAUSAL FISHER-INFORMATION INEQUALITIES (CFII)

The general framework in Sec. II emphasizes a simple but powerful idea: once a statistical model is required to arise from a classical causal structure, the likelihood is no longer an arbitrary function of the parameter. It must factorize according to the causal graph, and this factorization enforces quantitative restrictions on the Fisher information that can be observed at the endpoints of the experiment. We call such restrictions "causal Fisher-information inequalities (CFIIs)."

Here, we derive the CFIIs that constitute the backbone of our framework. The derivation is intentionally constructive: we isolate two universal principles—a multi-parameter Cramér–Rao geometry and a causal data-processing law—and then show how they fuse into a striking "series law" for Fisher information along a causal path. The conclusion is conceptually sharp: inverse Fisher information behaves as an information-resistance that must accumulate along classical causal bottlenecks. Hence, whenever an experiment exhibits a precision that beats this accumulation rule, the assumed classical causal model is not merely implausible; this is impossible by mathematics alone.

A. Effective Fisher information for causal parameters

CFIIs are most naturally expressed in terms of the causal parameters, which are often not the primitive parameters of the model but rather functions of them. In particular, in causal networks, the parameter of interest is frequently an additive quantity accumulated along edges (e.g., a total delay, a total phase, a total action), while each causal module contributes its own parameter.

Let X be an observation (possibly multivariate) generated by a regular parametric model $p(x|\boldsymbol{\theta})$ with a d-dimensional parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$. We define the score vector and Fisher information matrix as

$$\mathbf{s}_X(x|\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} \ln p(x|\boldsymbol{\theta}), \tag{21}$$

$$\mathbf{F}_X(\boldsymbol{\theta}) := \mathbb{E}_{x|\boldsymbol{\theta}}\Big[\mathbf{s}_X(x|\boldsymbol{\theta})\mathbf{s}_X(x|\boldsymbol{\theta})^T\Big]. \tag{22}$$

We assume standard regularity conditions so that $\mathbb{E}[\mathbf{s}_X] = \mathbf{0}$ and differentiation can be interchanged with integration.

The causal quantity of interest is modeled as a scalar function $g(\boldsymbol{\theta})$. In our applications the crucial case is the linear causal parameter

$$\Theta = g(\boldsymbol{\theta}) = \mathbf{u}^T\boldsymbol{\theta}, \tag{23}$$

for some fixed vector $\mathbf{u} \in \mathbb{R}^d$. The appropriate single-number measure of information about Θ in a multi-parameter model is the effective Fisher information defined by [11, 12]

$$F_X^{(\mathbf{u})}(\boldsymbol{\theta}) := \Big(\mathbf{u}^T\mathbf{F}_X(\boldsymbol{\theta})^{-1}\mathbf{u}\Big)^{-1}, \tag{24}$$

whenever $\mathbf{F}_X(\boldsymbol{\theta})$ is invertible on the support of $\mathbf{u}$. When $d = 1$ and $\mathbf{u} = 1$, this reduces to the standard Fisher information. Now we can state that:

Theorem S3 (Cramér–Rao bound for a linear causal parameter)**.** *Let $X \sim p(x|\boldsymbol{\theta})$ and let $\Theta = \mathbf{u}^T\boldsymbol{\theta}$ be the causal parameter in Eq. (23). For any unbiased estimator $\hat{\Theta}(X)$, one has*

$$\mathrm{Var}(\hat{\Theta}) \geq \mathbf{u}^T \mathbf{F}_X(\boldsymbol{\theta})^{-1}\mathbf{u} = \frac{1}{F_X^{(\mathbf{u})}(\boldsymbol{\theta})}. \tag{25}$$

Proof. —Let $r(X) := \hat{\Theta}(X) - \Theta$. The unbiasedness implies $\mathbb{E}[r] = 0$. Under the regularity assumptions,

$$\frac{\partial}{\partial\theta_i}\mathbb{E}[\hat{\Theta}] = \int dx\, \hat{\Theta}(x)\frac{\partial}{\partial\theta_i}p(x|\boldsymbol{\theta}) = \mathbb{E}\Big[\hat{\Theta}\, s_i(X|\boldsymbol{\theta})\Big], \tag{26}$$

where s_i is the ith component of the score vector in Eq. (21). Since $\mathbb{E}[\hat{\Theta}] = \Theta = \mathbf{u}^T\boldsymbol{\theta}$, we have $\partial_{\theta_i}\mathbb{E}[\hat{\Theta}] = u_i$. Using $\mathbb{E}[s_i] = 0$, Eq. (26) yields

$$\mathbb{E}\Big[r(X)\, s_i(X|\boldsymbol{\theta})\Big] = u_i. \tag{27}$$

Then, by letting $\mathbf{s} = \mathbf{s}_X(X|\boldsymbol{\theta})$ and $\mathbf{u} = (u_1, \ldots, u_d)^T$, consider the nonnegative quantity

$$\mathbb{E}\Big[(r - \mathbf{a}^T\mathbf{s})^2\Big] = \mathbb{E}[r^2] - 2\mathbf{a}^T\mathbb{E}[r\mathbf{s}] + \mathbf{a}^T\mathbb{E}[\mathbf{s}\mathbf{s}^T]\mathbf{a} \geq 0. \tag{28}$$

By Eq. (27) and Eq. (22), this becomes

$$\mathrm{Var}(\hat{\Theta}) - 2\mathbf{a}^T\mathbf{u} + \mathbf{a}^T\mathbf{F}_X(\boldsymbol{\theta})\mathbf{a} \geq 0. \tag{29}$$

Minimization of the right-hand side over $\mathbf{a}$ gives the choice $\mathbf{a} = \mathbf{F}_X(\boldsymbol{\theta})^{-1}\mathbf{u}$ and yields

$$\mathrm{Var}(\hat{\Theta}) \geq \mathbf{u}^T\mathbf{F}_X(\boldsymbol{\theta})^{-1}\mathbf{u}. \tag{30}$$

This is Eq. (25), and the identity with Eq. (24) is immediate. □

B. Causal data processing for Fisher information

A causal model is not only a factorization; it is also a directional constraint: information can only flow forward through the causal arrows, and any omission of variables corresponds to a stochastic post-processing that is independent of the parameter. Such processing cannot create information. We provide the following theorem:

Theorem S4 (Data-processing inequality for the Fisher information matrix)**.** *Let $X \sim p(x|\boldsymbol{\theta})$ and let Y be generated from X through a Markov kernel $p(y|x)$ that does not depend on $\boldsymbol{\theta}$:*

$$p(y|\boldsymbol{\theta}) = \int dx\, p(y|x)p(x|\boldsymbol{\theta}). \tag{31}$$

Then, the Fisher information matrices satisfy

$$\mathbf{F}_Y(\boldsymbol{\theta}) \preceq \mathbf{F}_X(\boldsymbol{\theta}), \tag{32}$$

where $\preceq$ denotes the positive-semidefinite order.

Proof. —Let $\mathbf{s}_X = \nabla_{\boldsymbol{\theta}} \ln p(x|\boldsymbol{\theta})$ and $\mathbf{s}_Y = \nabla_{\boldsymbol{\theta}} \ln p(y|\boldsymbol{\theta})$. Differentiating Eq. (31), we have

$$\nabla_{\boldsymbol{\theta}}\, p(y|\boldsymbol{\theta}) = \int dx\, p(y|x)\nabla_{\boldsymbol{\theta}}p(x|\boldsymbol{\theta}) = \int dx\, p(y|x)p(x|\boldsymbol{\theta})\mathbf{s}_X(x|\boldsymbol{\theta}). \tag{33}$$

Dividing by $p(y|\boldsymbol{\theta})$, we obtain

$$\mathbf{s}_Y(y|\boldsymbol{\theta}) = \frac{1}{p(y|\boldsymbol{\theta})}\int dx\, p(y|x)p(x|\boldsymbol{\theta})\mathbf{s}_X(x|\boldsymbol{\theta}) = \mathbb{E}\big[\mathbf{s}_X(X|\boldsymbol{\theta})\,|\,Y = y\big], \tag{34}$$

where the conditional expectation is taken with respect to the joint distribution $p(x, y|\boldsymbol{\theta}) = p(y|x)p(x|\boldsymbol{\theta})$.

For any fixed vector $\mathbf{v} \in \mathbb{R}^d$,

$$\begin{aligned}\mathbf{v}^T\mathbf{F}_Y(\boldsymbol{\theta})\mathbf{v} &= \mathbb{E}\big[\big(\mathbf{v}^T\mathbf{s}_Y(Y|\boldsymbol{\theta})\big)^2\big] = \mathbb{E}\big[\big(\mathbf{v}^T\mathbb{E}[\mathbf{s}_X|Y]\big)^2\big] \\ &\leq \mathbb{E}\big[\mathbb{E}\big[\big(\mathbf{v}^T\mathbf{s}_X\big)^2\,|\,Y\big]\big] = \mathbb{E}\big[\big(\mathbf{v}^T\mathbf{s}_X(X|\boldsymbol{\theta})\big)^2\big] = \mathbf{v}^T\mathbf{F}_X(\boldsymbol{\theta})\mathbf{v},\end{aligned} \tag{35}$$

where the inequality follows from the fact that conditional expectation is an L^2-contraction. Since Eq. (35) holds for all $\mathbf{v}$, we obtain Eq. (32). □

An immediate consequence is that effective Fisher information for any causal parameter cannot increase under a causal coarse-graining.

Corollary S1 (Monotonicity of effective Fisher information)**.** *Under the assumptions of* **Theorem S4***, for any* $\mathbf{u}$ *one has*

$$F_Y^{(\mathbf{u})}(\boldsymbol{\theta}) \leq F_X^{(\mathbf{u})}(\boldsymbol{\theta}). \tag{36}$$

Proof. —By **Theorem S4**, $\mathbf{F}_Y \preceq \mathbf{F}_X$. For positive definite matrices, the inversion reverses the order, hence $\mathbf{F}_Y^{-1} \succeq \mathbf{F}_X^{-1}$. Therefore, $\mathbf{u}^T\mathbf{F}_Y^{-1}\mathbf{u} \geq \mathbf{u}^T\mathbf{F}_X^{-1}\mathbf{u}$, and Eq. (36) follows from Eq. (24). □

C. The CFII for a causal path: inverse FI accumulates

We now arrive at the core inequality. Consider three nodes $A \to C \to B$ forming a causal path. Operationally, A labels the preparation (or initial condition), C labels an intermediate state (possibly unobserved), and B labels the final measurement record. A classical causal-path hypothesis asserts that the joint conditional distribution factorizes as

$$p(c, b|a, \theta_{ac}, \theta_{cb}) = p(c|a, \theta_{ac})\, p(b|c, \theta_{cb}), \tag{37}$$

which is equivalent to the conditional independence $A \perp B \,|\, C$ together with modular dependence of parameters: θ_{ac} appears only in the $A \to C$ module, and θ_{cb} appears only in the $C \to B$ module.

The causal parameter of interest is the *total* parameter accumulated along the path,

$$\theta_{ab} = \theta_{ac} + \theta_{cb}. \tag{38}$$

This is precisely the form encountered whenever the parameter corresponds to an additive generator under composition (time translations, phase shifts, accumulated action, etc.). We will show that Eq. (37) and Eq. (38) force a stringent bound on the precision available at the endpoint.

To express the bound in a way that is both causal and operational, we introduce the local Fisher informations

$$F_{ac}(\theta_{ac}) := \mathbb{E}_{c|a,\theta_{ac}}\big[\big(\partial_{\theta_{ac}} \ln p(c|a, \theta_{ac})\big)^2\big], \tag{39}$$

$$F_{cb}(\boldsymbol{\theta}) := \mathbb{E}_{c|a,\theta_{ac}}\big[\mathbb{E}_{b|c,\theta_{cb}}\big[\big(\partial_{\theta_{cb}} \ln p(b|c, \theta_{cb})\big)^2\big]\big]. \tag{40}$$

Here, the second definition explicitly acknowledges that the information in the $C \to B$ module may depend on which intermediate value c occurs, and hence must be averaged over the distribution induced by the upstream module (so in general it can also depend on θ_{ac} through $p(c|a, \theta_{ac})$).

We also define the endpoint effective Fisher information for estimating θ_{ab} from B alone, in the two-parameter model in Eq. (37). Let $\boldsymbol{\theta} = (\theta_{ac}, \theta_{cb})^T$ and $\mathbf{u} = (1, 1)^T$, so that $\theta_{ab} = \mathbf{u}^T\boldsymbol{\theta}$. Then, we have

$$F_{ab}^{(B)}(\boldsymbol{\theta}) := F_B^{(\mathbf{u})}(\boldsymbol{\theta}) = \Big(\mathbf{u}^T\mathbf{F}_B(\boldsymbol{\theta})^{-1}\mathbf{u}\Big)^{-1}. \tag{41}$$

When the model is effectively one-parameter (no nuisance degrees of freedom), $F_{ab}^{(B)}$ reduces to the usual Fisher information from $p(b|a, \theta_{ab})$. The advantage of Eq. (41) is that it remains meaningful for the genuinely causal situation where the path is described by modular parameters. We now construct the following theorem:

Theorem S5 (CFII for a classical causal path)**.** *Assume the causal-path factorization Eq. (37) and the additive causal parameter in Eq. (38). Then, the effective Fisher information at the endpoint satisfies*

$$\Big(F_{ab}^{(B)}(\boldsymbol{\theta})\Big)^{-1} \geq \Big(F_{ac}(\theta_{ac})\Big)^{-1} + \Big(F_{cb}(\boldsymbol{\theta})\Big)^{-1}. \tag{42}$$

Proof. —We proceed in two steps: first we evaluate the information about θ_{ab} when the intermediate variable C is accessible, and then we invoke causal data processing when C is discarded.

Step 1: joint record (C, B) yields the "series law" bound.

From Eq. (37), we have

$$\ln p(c, b|a, \theta_{ac}, \theta_{cb}) = \ln p(c|a, \theta_{ac}) + \ln p(b|c, \theta_{cb}). \tag{43}$$

Hence, the score components are

$$\begin{aligned} \partial_{\theta_{ac}} \ln p(c, b|a, \theta_{ac}, \theta_{cb}) &= \partial_{\theta_{ac}} \ln p(c|a, \theta_{ac}), \\ \partial_{\theta_{cb}} \ln p(c, b|a, \theta_{ac}, \theta_{cb}) &= \partial_{\theta_{cb}} \ln p(b|c, \theta_{cb}). \end{aligned} \tag{44}$$

The Fisher information matrix for the joint record (C, B) has the entries

$$\mathbf{F}_{(C,B)}(\boldsymbol{\theta}) = \begin{pmatrix} F_{ac}(\theta_{ac}) & 0 \\ 0 & F_{cb}(\boldsymbol{\theta}) \end{pmatrix} \tag{45}$$

because the cross term vanishes: i.e.,

$$\mathbb{E}\big[\big(\partial_{\theta_{ac}} \ln p\big)\big(\partial_{\theta_{cb}} \ln p\big)\big] = \mathbb{E}_{c|a,\theta_{ac}}\big[\big(\partial_{\theta_{ac}} \ln p(c|a,\theta_{ac})\big)\mathbb{E}_{b|c,\theta_{cb}}\big[\partial_{\theta_{cb}} \ln p(b|c,\theta_{cb})\big]\big] = 0, \tag{46}$$

since the conditional expectation of a score is zero. Now set $\mathbf{u} = (1,1)^T$ so that $\theta_{ab} = \mathbf{u}^T\boldsymbol{\theta}$. By **Theorem S3**, the following holds:

$$\mathrm{Var}(\hat{\theta}_{ab}) \geq \mathbf{u}^T\mathbf{F}_{(C,B)}(\boldsymbol{\theta})^{-1}\mathbf{u} = \frac{1}{F_{ac}(\theta_{ac})} + \frac{1}{F_{cb}(\boldsymbol{\theta})}. \tag{47}$$

Equivalently, the effective Fisher information for θ_{ab} contained in the *joint* record (C, B) is upper bounded by

$$F_{ab}^{(C,B)}(\boldsymbol{\theta}) \;:=\; F_{(C,B)}^{(\mathbf{u})}(\boldsymbol{\theta}) \;=\; \Big(\frac{1}{F_{ac}(\theta_{ac})} + \frac{1}{F_{cb}(\boldsymbol{\theta})}\Big)^{-1}. \tag{48}$$

Step 2: discarding C cannot increase information about θ_{ab}.

The endpoint record B is obtained from (C, B) by the $\boldsymbol{\theta}$-independent coarse-graining that simply forgets C. By **Corollary S1**,

$$F_{ab}^{(B)}(\boldsymbol{\theta}) \leq F_{ab}^{(C,B)}(\boldsymbol{\theta}). \tag{49}$$

By taking the inverse of both sides and using Eq. (48), we obtain

$$\Big(F_{ab}^{(B)}(\boldsymbol{\theta})\Big)^{-1} \geq \frac{1}{F_{ac}(\theta_{ac})} + \frac{1}{F_{cb}(\boldsymbol{\theta})}, \tag{50}$$

which is Eq. (42). The proof is completed. □

The inequality in Eq. (42) is the archetypal CFII: it converts a causal claim—"the process admits a classical intermediate node C with modular parameters"—into a quantitative constraint on observable precision. Here, we remark:

Remark S1 (Inverse Fisher information as information-resistance)**.** *Eq. (42) admits a physical reading that will guide the rest of this study. We here define the information-resistance $R := F^{-1}$. Then, Eq. (42) states that along a classical causal path, the resistances must add:*

$$R_{ab}^{(B)} \geq R_{ac} + R_{cb}. \tag{51}$$

In a classical causal model, the information must traverse the intermediate bottleneck C, and each module imposes an unavoidable resistance. Therefore, if an experiment exhibits an endpoint precision such that $R_{ab}^{(B)} < R_{ac} + R_{cb}$, the conclusion is unambiguous: the assumed classical causal-path model cannot reproduce the statistics. In this sense, a violation of Eq. (42) is not a mere metrological curiosity; it is a certificate of causal-model impossibility.

The causal-path CFII extends immediately to longer chains.

Corollary S2 (CFII for an N-step causal chain)**.** *Consider a chain $X_0 \to X_1 \to \cdots \to X_N$ with modular parameters $\theta_{j-1,j}$ and factorization*

$$p(x_1, \ldots, x_N \,|\, x_0, \boldsymbol{\theta}) = \prod_{j=1}^{N} p(x_j \,|\, x_{j-1}, \theta_{j-1,j}), \tag{52}$$

and define the additive causal parameter $\theta_{0N} = \sum_{j=1}^{N} \theta_{j-1,j}$. Then,

$$\Big(F_{0N}^{(X_N)}(\boldsymbol{\theta})\Big)^{-1} \geq \sum_{j=1}^{N} \Big(F_{j-1,j}(\theta_{j-1,j})\Big)^{-1}, \tag{53}$$

where $F_{0N}^{(X_N)}$ is the endpoint effective Fisher information for θ_{0N} from X_N alone.

Proof. —Apply **Theorem S5** iteratively, or proceed by induction by grouping the last two links into a single effective module. Each application adds one more term of the form $F_{j-1,j}^{-1}$ on the right-hand side. □

D. Series-parallel composition and the network view

The path inequality already reveals the essential mechanism: modularity forces the orthogonality of score contributions, and causal coarse-graining prevents the hidden information from being resurrected at the endpoint. However, the realistic causal models often combine modules not only in series but also in parallel (multiple conditionally independent branches providing information about the same causal parameter). The resulting picture is that Fisher information obeys composition laws reminiscent of electrical networks: Fisher informations add in parallel, while inverse Fisher informations add in series. This analogy is not decorative; it is the algebraic heart of how CFIIs scale to complex causal architectures.

We record the parallel law in a form suited for later use:

Theorem S6 (Parallel composition: additivity of Fisher information)**.** *Let Y_1 and Y_2 be conditionally independent given θ, i.e., $p(y_1, y_2|\theta) = p(y_1|\theta)p(y_2|\theta)$. Then, the Fisher information satisfies*

$$F_{(Y_1,Y_2)}(\theta) = F_{Y_1}(\theta) + F_{Y_2}(\theta). \tag{54}$$

Moreover, any coarse-graining $Z = \Gamma(Y_1, Y_2)$ produced by a θ-independent map obeys

$$F_Z(\theta) \leq F_{Y_1}(\theta) + F_{Y_2}(\theta). \tag{55}$$

Proof. —From $\ln p(y_1, y_2|\theta) = \ln p(y_1|\theta) + \ln p(y_2|\theta)$, the score is the sum of independent score contributions. Thus, by taking the expectation of the square, we obtain Eq. (54). The coarse-grained bound in Eq. (55) is an immediate application of **Theorem S4** in the one-parameter case. □

Taken together, **Theorem S5** (series law) and **Theorem S6** (parallel law) allow one to derive CFIIs for any causal architecture that can be built from series and parallel compositions of independent modules. In particular, one can assign each module an information-resistance $R = F^{-1}$ and reduce the causal graph exactly as one reduces a resistor network: series resistances add, parallel resistances combine harmonically. The endpoint Fisher information is then upper bounded by the effective conductance of this reduced network, and the corresponding endpoint resistance is lower bounded by the effective resistance. This network viewpoint will be exploited in later sections to construct families of testable inequalities tailored to specific causal hypotheses, and to interpret their violation as a collapse of classical causal explanation.

SUPPLEMENTARY NOTE IV. VIOLATION = IMPOSSIBILITY OF THE CLASSICAL CAUSAL MODEL (MODEL FALSIFICATION)

The inequalities derived in Sec. III look, at first glance, like technical bounds on Fisher information. Their true significance is far sharper. A CFII is not merely a limit on precision; it is a logical constraint imposed by a classical causal explanation. Once a causal model class is fixed, the inequality becomes a necessary condition for compatibility. Consequently, a violation is not an invitation to interpret; it is a mathematical verdict. Here, we formalize this verdict and explain why it is physically dramatic: a CFII violation means that the classical causal narrative literally cannot generate the observed statistics.

A. From inequality to falsification: the logical structure of CFIIs

Recall from **Definition S1** that a classical causal model class $\mathcal{M}$ specifies which families of operational distributions $\{p(x|s,\theta)\}$ are admissible, by requiring that they arise from a DAG factorization (possibly with latent variables) together with the associated conditional-independence relations. **Definition S2** then introduced a CFII as an inequality

$$\mathcal{G}\left(\{F_s(\theta)\}_{s\in\mathcal{S}_0}\right) \geq 0 \tag{56}$$

that holds for all θ and for every operational model compatible with $\mathcal{M}$.

To turn this into a falsification principle, we separate two objects. First, we fix a finite set of contexts $\mathcal{S}_0$ and consider the vector of Fisher informations

$$\mathbf{F}(\theta) := (F_s(\theta))_{s\in\mathcal{S}_0}. \tag{57}$$

Second, we define the feasible region of FI-vectors admitted by the model class.

Definition S3 (Feasible FI region)**.** *For a fixed θ, the feasible FI region of a model class $\mathcal{M}$ is the set*

$$\Omega_{\mathcal{M}}(\theta) := \{\mathbf{F}(\theta) : \{p(x|s,\theta)\}_{s\in\mathcal{S}_0} \textit{ is compatible with } \mathcal{M}\}. \tag{58}$$

A CFII is precisely the statement that $\Omega_{\mathcal{M}}(\theta)$ is contained in the half-space defined by $\mathcal{G} \geq 0$.

Theorem S7 (CFII violation implies causal-model impossibility)**.** *Fix a model class $\mathcal{M}$, a context set $\mathcal{S}_0$, and a function $\mathcal{G}$ such that Eq. (56) holds for all operational models compatible with $\mathcal{M}$. Let an experiment produce operational distributions $\{p_{\exp}(x|s,\theta)\}_{s\in\mathcal{S}_0}$ with Fisher informations $\mathbf{F}_{\exp}(\theta)$. If for some θ one has*

$$\mathcal{G}\left(\mathbf{F}_{\exp}(\theta)\right) < 0, \tag{59}$$

then $\{p_{\exp}(x|s,\theta)\}_{s\in\mathcal{S}_0}$ is not compatible with $\mathcal{M}$. Equivalently, $\mathbf{F}_{\exp}(\theta) \notin \Omega_{\mathcal{M}}(\theta)$.

Proof. —Assume, toward contradiction, that $\{p_{\exp}(x|s,\theta)\}_{s\in\mathcal{S}_0}$ is compatible with $\mathcal{M}$. Then, by the definition of CFII, Eq. (56) must hold for this operational model at the same θ. Hence, $\mathcal{G}(\mathbf{F}_{\exp}(\theta)) \geq 0$, contradicting Eq. (59). Therefore, the operational model cannot be compatible with $\mathcal{M}$, and the FI-vector lies outside the feasible region. □

Theorem S7 is logically simple, but its meaning is profound. The experimental task is to estimate FI (or effective FI) from observed statistics; the theoretical task is to derive CFIIs from explicit causal assumptions; the conclusion is model falsification.

B. The causal-path witness: when information seems to appear from nowhere

The abstract falsification principle becomes physically vivid in the archetypal case of a causal path $A \to C \to B$. Here, the classical claim is that the endpoint record B can only depend on the past through the intermediate bottleneck C, and that parameter dependence is modular, as encoded in Eq. (37). **Theorem S5** then states that the endpoint effective Fisher information for the additive causal parameter $\theta_{ab} = \theta_{ac} + \theta_{cb}$ obeys

$$\left(F_{ab}^{(B)}(\boldsymbol{\theta})\right)^{-1} \geq \left(F_{ac}(\theta_{ac})\right)^{-1} + \left(F_{cb}(\boldsymbol{\theta})\right)^{-1}. \tag{60}$$

This inequality is the mathematical form of a narrative: information must traverse C, and every traversal adds an irreducible information-resistance.

For experimental falsification, it is convenient to package Eq. (60) into a single witness.

Definition S4 (Causal-path violation statistic)**.** *For a causal-path hypothesis with parameters $\boldsymbol{\theta} = (\theta_{ac}, \theta_{cb})^T$, let us define*

$$V_{\mathrm{path}}(\boldsymbol{\theta}) := \left(F_{ab}^{(B)}(\boldsymbol{\theta})\right)^{-1} - \left(F_{ac}(\theta_{ac})\right)^{-1} - \left(F_{cb}(\boldsymbol{\theta})\right)^{-1}. \tag{61}$$

Then, we can construct the theorem:

Theorem S8 (Causal-path falsification criterion)**.** *If an operational model is compatible with the classical causal-path assumptions of Sec. III C, then*

$$V_{\mathrm{path}}(\boldsymbol{\theta}) \geq 0 \tag{62}$$

for all admissible $\boldsymbol{\theta}$. Therefore, if an experiment yields $V_{\mathrm{path}}(\boldsymbol{\theta}) < 0$ for some $\boldsymbol{\theta}$, then no classical causal-path model in that class can reproduce the observed statistics.

Proof. —The nonnegativity in Eq. (62) is algebraically equivalent to Eq. (60), which holds by **Theorem S5** under the causal-path assumptions. The contrapositive implication is an instance of **Theorem S7**. □

At this point the physical implication can be stated without qualification. A negative value of V_{path} means that, at the endpoint B, one can estimate the total causal parameter more precisely than would be permitted if the parameter were truly accumulated through two independent causal modules connected in series. In the information-resistance language of **Remark S1**, the experiment behaves as if the series connection had developed an "active element" that cancels the resistance. Classically, such behavior is forbidden because every causal module is a passive channel: it may preserve information, but it cannot conjure additional information about θ that was not already present in its input. A violation therefore confronts us with a dramatic conclusion: either the intermediate bottleneck C does not exist as a classical variable, or the process does not decompose into independent causal modules, or some hidden influence bypasses the assumed causal constraints. In every case, the classical causal story collapses.

C. What exactly is falsified?

A central virtue of the causal interpretation is the precision about what is, and is not, concluded from a violation. A CFII is derived under the explicit assumptions: a DAG factorization, its implied conditional-independence relations, and a modular parameter dependence consistent with the causal decomposition. Hence, a violation falsifies the conjunction of these assumptions. However, it does not claim that the world is "quantum" in the abstract; it claims something sharper:

the observed statistics cannot be generated by any member of the specified classical causal model class.

This distinction is not semantic. The different notions of "classicality" correspond to different model classes $\mathcal{M}$, and thus to different feasible regions $\Omega_{\mathcal{M}}(\theta)$. A violation identifies which classical causal explanations are ruled out, while leaving open the possibility that a larger class of models (with weaker conditional-independence requirements, additional latent memory nodes, or explicit context dependence) may still explain the data. This is exactly why CFIIs are powerful scientific instruments: they do not merely quantify precision, they map which causal explanations remain tenable.

In the causal-path setting, for instance, $V_{\mathrm{path}} < 0$ rules out the existence of any intermediate classical bottleneck C that simultaneously (i) renders A and B conditionally independent given C and (ii) supports a modular decomposition of the parameter into $(\theta_{ac}, \theta_{cb})$ with additive causal parameter θ_{ab}. In dynamical quantum experiments, this naturally resonates with the failure of a trajectory-like description by a fixed series of intermediate states, where the true evolution may involve coherence across segments that cannot be represented by a single classical node C. In such a case, a coherent superposition, for example, a qubit state $|\psi\rangle$ evolving under a Hamiltonian $\hat{H}$, can exhibit statistics that invalidate the classical path decomposition precisely, because the causal bottleneck cannot capture phase-sensitive interference effects. The inequality detects this failure at the level of Fisher information: the putative resistance of a classical bottleneck is overcome by interference, and the only consistent conclusion is that the classical bottleneck model is impossible. We will explicitly show this example later.

D. Statistical certification from finite data

Theorems S7 and **S8** are the statements about true Fisher informations, i.e., about the underlying distributions. In experiment, Fisher information must be estimated from a finite set of samples. Here, we present a statistically principled route to certify the CFII violations with controlled confidence.

We assume that, for each context $s \in \mathcal{S}_0$ and a fixed parameter value θ (e.g., a calibrated or controlled setting), we can evaluate the score

$$\ell_s(x|\theta) := \frac{\partial}{\partial\theta} \ln p(x|s,\theta). \tag{63}$$

This is standard in the metrological models where the encoding $p(x|s,\theta)$ is known up to the value of θ, and it can also be achieved operationally by local calibration. Given i.i.d. outcomes $\{x_{s,i}\}_{i=1}^{N_s}$ sampled from $p(x|s,\theta)$, let us consider the estimator

$$\tilde{F}_s(\theta) := \frac{1}{N_s} \sum_{i=1}^{N_s} \ell_s(x_{s,i}|\theta)^2. \tag{64}$$

By construction, $\mathbb{E}[\tilde{F}_s(\theta)] = F_s(\theta)$ whenever the score has finite second moment.

Theorem S9 (Consistency and asymptotic normality of $\tilde{F}_s(\theta)$)**.** *Assume $\mathbb{E}[\ell_s(X|\theta)^4] < \infty$ for each $s \in \mathcal{S}_0$, where $X \sim p(x|s,\theta)$. Then:*

$$\tilde{F}_s(\theta) \to F_s(\theta) \tag{65}$$

in probability as $N_s \to \infty$, and moreover

$$\sqrt{N_s}\left(\tilde{F}_s(\theta) - F_s(\theta)\right) \Rightarrow \mathcal{N}\left(0, \sigma_s^2(\theta)\right), \tag{66}$$

where

$$\sigma_s^2(\theta) := \mathrm{Var}\left(\ell_s(X|\theta)^2\right). \tag{67}$$

Proof. —Define the i.i.d. random variables $Y_{s,i} := \ell_s(x_{s,i}|\theta)^2$. Then, $\tilde{F}_s(\theta)$ is the sample mean of $Y_{s,i}$:

$$\tilde{F}_s(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} Y_{s,i}. \tag{68}$$

Since $\mathbb{E}[Y_{s,i}] = \mathbb{E}[\ell_s(X|\theta)^2] = F_s(\theta)$ and $\mathbb{E}[Y_{s,i}^2] = \mathbb{E}[\ell_s(X|\theta)^4] < \infty$, the weak law of large numbers implies Eq. (65). The central limit theorem for i.i.d. variables with finite variance yields

$$\sqrt{N_s}\left(\tilde{F}_s(\theta) - F_s(\theta)\right) = \frac{1}{\sqrt{N_s}} \sum_{i=1}^{N_s} \left(Y_{s,i} - \mathbb{E}[Y_{s,i}]\right) \Rightarrow \mathcal{N}\left(0, \mathrm{Var}(Y_{s,1})\right), \tag{69}$$

which is Eq. (66) with Eq. (67). □

Theorem S9 provides an operational route to confidence intervals for each $F_s(\theta)$, and, crucially, for any smooth function $\mathcal{G}$ of the FI-vector. We now translate it to a violation test. Let $\tilde{\mathbf{F}}(\theta) := (\tilde{F}_s(\theta))_{s\in\mathcal{S}_0}$, and define the plug-in estimator

$$\tilde{\mathcal{G}}(\theta) := \mathcal{G}\left(\tilde{\mathbf{F}}(\theta)\right). \tag{70}$$

Assume that outcomes are sampled independently across contexts, so that the estimators $\{\tilde{F}_s(\theta)\}$ are asymptotically independent.

Theorem S10 (Asymptotic certification of CFII violation)**.** *Assume the conditions of* **Theorem S9** *for all $s \in \mathcal{S}_0$, and assume $\mathcal{G}$ is continuously differentiable on an open neighborhood of $\mathbf{F}(\theta)$. Let $N_s \to \infty$ with fixed proportions $q_s := N_s/N$ where $N = \sum_{s\in\mathcal{S}_0} N_s$. Then,*

$$\sqrt{N}\left(\tilde{\mathcal{G}}(\theta) - \mathcal{G}(\mathbf{F}(\theta))\right) \Rightarrow \mathcal{N}\left(0, \tau^2(\theta)\right), \tag{71}$$

where

$$\tau^2(\theta) = \sum_{s\in\mathcal{S}_0} \frac{1}{q_s} \left(\frac{\partial \mathcal{G}}{\partial F_s}\Big|_{\mathbf{F}(\theta)}\right)^2 \sigma_s^2(\theta). \tag{72}$$

Proof. —From **Theorem S9**, the vector

$$\sqrt{N}\left(\tilde{\mathbf{F}}(\theta) - \mathbf{F}(\theta)\right) = \left(\sqrt{N}\left(\tilde{F}_s(\theta) - F_s(\theta)\right)\right)_{s\in\mathcal{S}_0} \tag{73}$$

has asymptotically normal components with variances scaled by q_s:

$$\sqrt{N}\left(\tilde{F}_s(\theta) - F_s(\theta)\right) = \sqrt{\frac{N}{N_s}}\sqrt{N_s}\left(\tilde{F}_s(\theta) - F_s(\theta)\right) \Rightarrow \mathcal{N}\left(0, \frac{\sigma_s^2(\theta)}{q_s}\right). \tag{74}$$

The independence across contexts yields an asymptotically normal vector with diagonal covariance matrix whose sth diagonal entry is $\sigma_s^2(\theta)/q_s$.

Since $\mathcal{G}$ is continuously differentiable, a first-order Taylor expansion gives

$$\tilde{\mathcal{G}}(\theta) - \mathcal{G}(\mathbf{F}(\theta)) = \nabla\mathcal{G}(\mathbf{F}(\theta))^T \left(\tilde{\mathbf{F}}(\theta) - \mathbf{F}(\theta)\right) + r_N, \tag{75}$$

where the remainder r_N satisfies $r_N = o_p(\|\tilde{\mathbf{F}} - \mathbf{F}\|)$. Multiplying by $\sqrt{N}$ and using the multivariate central limit theorem together with Slutsky's theorem, we obtain Eq. (71). The variance is the quadratic form of the gradient with the asymptotic covariance matrix, which reduces to Eq. (72), because the covariance is diagonal. □

Theorem S10 makes the falsification program experimentally concrete. Under the null hypothesis that the data are compatible with $\mathcal{M}$, one must have $\mathcal{G}(\mathbf{F}(\theta)) \geq 0$. If an experiment yields $\tilde{\mathcal{G}}(\theta)$ significantly below zero compared to its standard error $\tau(\theta)/\sqrt{N}$, then the classical causal model class is rejected with controlled significance. In other words, one can attach to a CFII violation the same statistical seriousness as to any hypothesis test in physics.

For the causal-path witness, $\mathcal{G}$ can be chosen as $\mathcal{G} = V_{\mathrm{path}}$ in Eq. (61), and the derivative factors entering Eq. (72) take a simple form. Writing $F_1 = F_{ab}^{(B)}$, $F_2 = F_{ac}$, and $F_3 = F_{cb}$, one has

$$\frac{\partial V_{\mathrm{path}}}{\partial F_1} = -\frac{1}{F_1^2}, \quad \frac{\partial V_{\mathrm{path}}}{\partial F_2} = \frac{1}{F_2^2}, \quad \frac{\partial V_{\mathrm{path}}}{\partial F_3} = \frac{1}{F_3^2}. \tag{76}$$

Substituting Eq. (76) into Eq. (72) yields an explicit asymptotic variance for $\hat{V}_{\mathrm{path}}$. Thus, the inequality, originally a statement about causal structure, becomes an experimentally testable boundary with quantitative confidence.

E. Why this matters

What makes CFII-based falsification distinctive is that the rejected object is not a phenomenological fit but a "causal explanation" of how information about θ may propagate through an experiment. The content of a violation is therefore structural: it tells us that no classical model with the assumed conditional independences can account for the data, even if that model is allowed arbitrary internal parameters and arbitrary hidden variables. When a CFII is violated, the classical causal narrative does not merely require refinement; it ceases to exist. This is precisely the point at which the story becomes metrological. Once a classical causal model collapses, the "extra" information that defeats the classical resistance law cannot be treated as an accident. It is the signature of a nonclassical resource that bypasses the classical bottleneck.

We will show, in later sections, how such violations can be interpreted not only as witnesses but as resources, and how they encompass as special instances contextuality-based protocols in which the failure of a noncontextual causal condition is converted into enhanced precision. The dramatic moral is that, in metrology, impossibility results are not dead ends. They are signposts that point to where Nature stores her most useful nonclassical advantages.

SUPPLEMENTARY NOTE V. FROM WITNESS TO RESOURCE: "CAUSAL-MODEL COLLAPSE" AS A METROLOGICAL RESOURCE

Sections III and IV established the conceptual backbone of this work: a CFII is a necessary condition for the existence of a classical causal explanation, and its violation is a rigorous falsification of that explanation. At this stage, however, we ask a practical question: if a CFII violation certifies that a classical causal story fails, what does that buy us in metrology? Is it merely a diagnostic statement about "nonclassicality," or does it translate into a concrete improvement of estimation performance?

Here, we answer with an engineering-level message. A CFII violation is not only a witness; it is a resource certificate. It certifies that the experiment possesses an information flow that cannot be routed through the assumed classical bottlenecks, and this surplus information is directly convertible into smaller estimation error. The conversion is quantitative, and it is operational: it does not require philosophical interpretation, only a well-defined estimator. The same inequality that collapses the classical causal story simultaneously exposes a metrological advantage that the classical story would have declared impossible.

A. Classical causal precision bounds from CFIIs

A CFII is an inequality among Fisher informations. Since Fisher information is the curvature controlling Cramér–Rao-type bounds, each CFII immediately induces a precision bound that holds whenever the classical causal model class is valid. This is the sense in which CFIIs serve as causal benchmarks for metrology. We formulate the statement in a general form first, and then specialize it to the causal-path inequality.

Theorem S11 (Classical causal precision bound induced by a CFII)**.** *Fix a model class $\mathcal{M}$ and a finite set of contexts $\mathcal{S}_0$. Assume that a CFII of the form*

$$\mathcal{G}\left(\{F_s(\theta)\}_{s\in\mathcal{S}_0}\right) \geq 0 \tag{77}$$

holds for all operational models compatible with $\mathcal{M}$. Let an estimator $\hat{\theta}$ be constructed from N i.i.d. repetitions of an experiment that uses only contexts in $\mathcal{S}_0$. Then, for any operational model compatible with $\mathcal{M}$ and for any unbiased $\hat{\theta}$, the mean-square error satisfies a model-dependent lower bound

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{N}\frac{1}{F_{\mathrm{cl}}(\theta)}, \tag{78}$$

where $F_{\mathrm{cl}}(\theta)$ is any quantity that upper bounds the achievable per-sample Fisher information under the constraint in Eq. (77). In particular, if the CFII can be algebraically rearranged into

$$F_{\mathrm{eff}}(\theta) \leq F_{\mathrm{cl}}(\theta), \tag{79}$$

for a suitable effective Fisher information $F_{\mathrm{eff}}(\theta)$ relevant to the estimation strategy, then Eq. (78) follows with F_{cl}.

Proof. For N i.i.d. repetitions, the Fisher information is additive, hence the total Fisher information equals N times the per-sample Fisher information (**Theorem S2** in Sec. II). For any unbiased estimator, the Cramér–Rao bound gives

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{N}\frac{1}{F_{\mathrm{eff}}(\theta)}, \tag{80}$$

where F_{eff} denotes the per-sample Fisher information appropriate to the statistical model used in the protocol. If the operational model is compatible with $\mathcal{M}$, then its Fisher informations must satisfy the CFII as in Eq. (77). If Eq. (77) implies an upper bound in Eq. (79) on F_{eff}, then by substituting Eq. (79) into Eq. (80), we obtain Eq. (78), completing the proof. □

Theorem S11 is conceptually simple but practically sharp: it tells us that a causal model class defines a *precision frontier*. Any protocol consistent with the model must live above that frontier. The moment an experiment falls below it, the classical model is not merely disfavored; it is logically incompatible with the achieved precision.

The causal-path case makes this concrete. Under the classical causal-path hypothesis of Sec. III C, **Theorem S5** states that

$$\left(F_{ab}^{(B)}(\boldsymbol{\theta})\right)^{-1} \geq (F_{ac}(\theta_{ac}))^{-1} + (F_{cb}(\boldsymbol{\theta}))^{-1}. \tag{81}$$

Here, we define the classical benchmark Fisher information by the harmonic composition

$$F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta}) := \left((F_{ac}(\theta_{ac}))^{-1} + (F_{cb}(\boldsymbol{\theta}))^{-1}\right)^{-1}. \tag{82}$$

Then, Eq. (81) is precisely the upper bound

$$F_{ab}^{(B)}(\boldsymbol{\theta}) \leq F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta}) \tag{83}$$

that **Theorem S11** requires. Thus, if a process genuinely decomposes into two independent causal modules in series, the achievable per-sample Fisher information about the total parameter can never exceed the harmonic mean in Eq. (82). This is the classical "series penalty" expressed in metrological language.

B. Violation-certified gain: witness-to-resource transition

We now convert the previous statement into a resource claim. Recall the causal-path violation statistic

$$V_{\text{path}}(\boldsymbol{\theta}) = \left(F_{ab}^{(B)}(\boldsymbol{\theta})\right)^{-1} - (F_{ac}(\theta_{ac}))^{-1} - (F_{cb}(\boldsymbol{\theta}))^{-1}. \tag{84}$$

The inequality in Eq. (81) is equivalent to $V_{\text{path}}(\boldsymbol{\theta}) \geq 0$. Here, a negative value $V_{\text{path}} < 0$ can be interpreted as the impossibility of the assumed classical causal-path model. Now we show that the same quantity also certifies a quantitative metrological gain.

To state the gain in the most operational form, we consider N i.i.d. repetitions of the endpoint experiment producing outcomes $B_1, \ldots, B_N$ distributed as $p(b|\boldsymbol{\theta})$, and let $\hat{\theta}_{ab}$ be an estimator of $\theta_{ab} = \theta_{ac} + \theta_{cb}$ constructed from $\{B_i\}_{i=1}^N$. Here we provide the following theorem:

Theorem S12 (Violation-certified precision enhancement)**.** *Assume that $F_{ab}^{(B)}(\boldsymbol{\theta})$ is the per-sample effective Fisher information for estimating θ_{ab} from endpoint data B in the model under consideration. Define the classical benchmark $F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta})$ by Eq. (82). If $V_{\text{path}}(\boldsymbol{\theta}) < 0$, then*

$$F_{ab}^{(B)}(\boldsymbol{\theta}) > F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta}), \tag{85}$$

and consequently the Cramér–Rao bound for any unbiased estimator satisfies

$$\frac{1}{N}\frac{1}{F_{ab}^{(B)}(\boldsymbol{\theta})} < \frac{1}{N}\frac{1}{F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta})}. \tag{86}$$

Moreover, defining the resistance quantities

$$R_{\text{ab}}(\boldsymbol{\theta}) = \left(F_{ab}^{(B)}(\boldsymbol{\theta})\right)^{-1}, \quad R_{\text{cl}}(\boldsymbol{\theta}) = \left(F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta})\right)^{-1}, \tag{87}$$

the violation directly yields the certified improvement factor

$$\frac{R_{\text{cl}}(\boldsymbol{\theta})}{R_{\text{ab}}(\boldsymbol{\theta})} = \frac{R_{\text{cl}}(\boldsymbol{\theta})}{R_{\text{cl}}(\boldsymbol{\theta}) + V_{\text{path}}(\boldsymbol{\theta})} > 1. \tag{88}$$

Proof. —If $V_{\text{path}}(\boldsymbol{\theta}) < 0$, then by Eq. (84), we have

$$\left(F_{ab}^{(B)}(\boldsymbol{\theta})\right)^{-1} < \left(F_{\text{cl}}^{(\text{path})}(\boldsymbol{\theta})\right)^{-1}. \tag{89}$$

Taking reciprocals preserves the strict inequality because both sides are positive, proving Eq. (85). By substituting these informations into the Cramér–Rao bound, we obtain Eq. (86). Finally, Eq. (88) follows from the definition $V_{\text{path}} = R_{\text{ab}} - R_{\text{cl}}$. □

Theorem S12 is the operational heart of the "witness-to-resource" transition. It says that the same numerical quantity that falsifies a classical causal explanation also provides a quantitative *performance guarantee*: the endpoint experiment achieves a precision that no member of the classical causal-path model class can match, even in principle.

Remark S2 (Negative $V_{\rm path}$ is a metrology enhancement resource)**.** *The result is dramatic when expressed in the resistance language. In a classical causal narrative, resistances add. The total estimation error behaves as if the parameter had to pass through two passive resistors in series. A negative $V_{\rm path}$ means that the experiment contains an element that cancels part of this series resistance. In classical circuitry that would require an active device. Here the "active device" is not an additional energy source; it is the collapse of a classical causal description. Metrologically, the collapse is a resource because it converts structural impossibility into reduced error.*

C. Achievability: efficient estimators harvest the resource

A resource certificate is meaningful only if the improved bound is not a mirage. **Theorem S12** compares two Cramér–Rao bounds, but does not yet guarantee that the smaller bound can be approached by an explicit estimator. We therefore record a standard but essential fact: under regularity conditions, maximum likelihood estimation asymptotically achieves the Fisher information bound. This is the mechanism by which causal-model collapse becomes practically harvestable.

Theorem S13 (Asymptotic efficiency of maximum likelihood estimation)**.** *Consider $X_1, \ldots, X_N$, the i.i.d. samples drawn from a regular one-parameter family $p(x|\theta)$ with true parameter value θ. Let*

$$\ell_N(\theta) = \sum_{i=1}^{N} \ln p(X_i|\theta) \tag{90}$$

be the log-likelihood and define the maximum likelihood estimator $\hat{\theta}_{\rm ML}$ by the score equation

$$\frac{\partial}{\partial\theta}\ell_N(\hat{\theta}_{\rm ML}) = 0. \tag{91}$$

Assume that $\hat{\theta}_{\rm ML} \to \theta$ in probability and that the following identities hold with finite moments:

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\ln p(X|\theta)\right] = 0, \quad \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln p(X|\theta)\right)^2\right] = F(\theta), \quad \textit{and} \quad \mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\ln p(X|\theta)\right] = F(\theta), \tag{92}$$

where $F(\theta)$ is the per-sample Fisher information. Then,

$$\sqrt{N}\left(\hat{\theta}_{\rm ML} - \theta\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{F(\theta)}\right), \tag{93}$$

and consequently,

$$\mathrm{Var}(\hat{\theta}_{\rm ML}) = \frac{1}{N}\frac{1}{F(\theta)} + o\left(\frac{1}{N}\right). \tag{94}$$

Proof. —Define the score and observed information

$$\begin{aligned} S_N(\theta) &= \frac{\partial}{\partial\theta}\ell_N(\theta) = \sum_{i=1}^{N}\frac{\partial}{\partial\theta}\ln p(X_i|\theta), \\ J_N(\theta) &= -\frac{\partial^2}{\partial\theta^2}\ell_N(\theta) = \sum_{i=1}^{N}\left(-\frac{\partial^2}{\partial\theta^2}\ln p(X_i|\theta)\right). \end{aligned} \tag{95}$$

The defining Eq. (91) reads $S_N(\hat{\theta}_{\rm ML}) = 0$. By Taylor's theorem,

$$0 = S_N(\hat{\theta}_{\rm ML}) = S_N(\theta) - J_N(\tilde{\theta}_N)\left(\hat{\theta}_{\rm ML} - \theta\right), \tag{96}$$

where $\tilde{\theta}_N$ lies between $\hat{\theta}_{\rm ML}$ and θ. By rearranging, we have

$$\sqrt{N}\left(\hat{\theta}_{\rm ML} - \theta\right) = \left(\frac{1}{N}J_N(\tilde{\theta}_N)\right)^{-1}\left(\frac{1}{\sqrt{N}}S_N(\theta)\right). \tag{97}$$

We then analyze the two factors. First, by Eq. (92) and the central limit theorem, we can find

$$\frac{1}{\sqrt{N}} S_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} s(X_i|\theta) \Rightarrow \mathcal{N}\left(0, F(\theta)\right), \tag{98}$$

where $s(X|\theta) = \partial_\theta \ln p(X|\theta)$.

Second, by the law of large numbers and Eq. (92),

$$\frac{1}{N} J_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} j(X_i|\theta) \to F(\theta) \tag{99}$$

in probability, where $j(X|\theta) = -\partial_\theta^2 \ln p(X|\theta)$. Since $\hat{\theta}_{\mathrm{ML}} \to \theta$ in probability by assumption, also $\tilde{\theta}_N \to \theta$ in probability. Under continuity of $j(X|\theta)$ in θ and dominated convergence sufficient to exchange limits and expectations, Eq. (99) extends to

$$\frac{1}{N} J_N(\tilde{\theta}_N) \to F(\theta) \tag{100}$$

in probability. Therefore, by Slutsky's theorem,

$$\left(\frac{1}{N} J_N(\tilde{\theta}_N)\right)^{-1} \to \frac{1}{F(\theta)} \tag{101}$$

in probability.

Finally, by combining Eq. (98), Eq. (101), and Eq. (97) via Slutsky's theorem, we can attain

$$\sqrt{N}\left(\hat{\theta}_{\mathrm{ML}} - \theta\right) \Rightarrow \mathcal{N}\left(0, \frac{1}{F(\theta)}\right), \tag{102}$$

which is Eq. (93). The variance statement Eq. (94) follows. □

Theorem S13 supplies the missing bridge from principle to practice. If an experiment exhibits a violation-certified increase in Fisher information, then an explicit estimator (maximum likelihood, or any asymptotically efficient alternative) can harvest that increase into reduced error. The point is practical: "causal-model collapse" certifies that there exists a *working estimation procedure* whose asymptotic performance beats the classical causal benchmark.

D. Mechanism: Fisher-information synergy from correlated score contributions

We established that the CFII violation implies a gain and that the gain is achievable. We now expose the structural mechanism behind the gain in a way that is independent of any particular physical platform. The mechanism is the emergence of "synergy" between the causal modules, visible as off-diagonal terms of the Fisher information matrix. These off-diagonal terms are precisely what classical modularity forbids, as made explicit by the vanishing cross term in Eq. (46).

Consider a two-parameter description $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$ of a process and an observation Y distributed as $p(y|\theta_1, \theta_2)$. Let the Fisher information matrix be

$$\mathbf{F}_Y(\boldsymbol{\theta}) = \begin{pmatrix} F_1(\boldsymbol{\theta}) & J(\boldsymbol{\theta}) \\ J(\boldsymbol{\theta}) & F_2(\boldsymbol{\theta}) \end{pmatrix}, \tag{103}$$

where

$$\begin{aligned} F_1(\boldsymbol{\theta}) &= \mathbb{E}\left[\left(\partial_{\theta_1} \ln p(Y|\theta_1, \theta_2)\right)^2\right], \\ F_2(\boldsymbol{\theta}) &= \mathbb{E}\left[\left(\partial_{\theta_2} \ln p(Y|\theta_1, \theta_2)\right)^2\right], \\ J(\boldsymbol{\theta}) &= \mathbb{E}\left[\left(\partial_{\theta_1} \ln p(Y|\theta_1, \theta_2)\right)\left(\partial_{\theta_2} \ln p(Y|\theta_1, \theta_2)\right)\right]. \end{aligned} \tag{104}$$

The quantity $J(\boldsymbol{\theta})$ is the covariance between the score contributions of the two parameters. In a classical causal-path model with modular parameter dependence, this covariance vanishes by construction (as shown explicitly in Eq. (46)). When it does not vanish, the two modules are not independent in the sense required by the classical causal decomposition.

We now compute the effective Fisher information for the additive parameter

$$\Theta = \theta_1 + \theta_2. \tag{105}$$

Let $\mathbf{u} = (1,1)^T$. By the definition of Eq. (24),

$$F_Y^{(\mathbf{u})}(\boldsymbol{\theta}) = \left(\mathbf{u}^T \mathbf{F}_Y(\boldsymbol{\theta})^{-1} \mathbf{u}\right)^{-1}. \tag{106}$$

We then provide the theorem:

Theorem S14 (Synergy formula and the condition for beating the series law)**.** *Assume $\mathbf{F}_Y(\boldsymbol{\theta})$ in Eq. (103) is positive definite, i.e., $F_1F_2 - J^2 > 0$. Then the effective Fisher information for $\Theta = \theta_1 + \theta_2$ is*

$$F_Y^{(\mathbf{u})}(\boldsymbol{\theta}) = \frac{F_1F_2 - J^2}{F_1 + F_2 - 2J}. \tag{107}$$

Moreover, comparing to the classical series benchmark

$$F_{\text{series}}(\boldsymbol{\theta}) = \left(\frac{1}{F_1} + \frac{1}{F_2}\right)^{-1} = \frac{F_1F_2}{F_1 + F_2}, \tag{108}$$

one has

$$F_Y^{(\mathbf{u})}(\boldsymbol{\theta}) > F_{\text{series}}(\boldsymbol{\theta}) \tag{109}$$

if and only if

$$0 < J(\boldsymbol{\theta}) < \frac{2F_1(\boldsymbol{\theta})F_2(\boldsymbol{\theta})}{F_1(\boldsymbol{\theta}) + F_2(\boldsymbol{\theta})}. \tag{110}$$

Finally, for fixed F_1 and F_2, the supremum of $F_Y^{(\mathbf{u})}$ over J under $F_1F_2 - J^2 > 0$ is

$$\sup_J F_Y^{(\mathbf{u})}(\boldsymbol{\theta}) = \min\left\{F_1(\boldsymbol{\theta}), F_2(\boldsymbol{\theta})\right\}. \tag{111}$$

Proof. —We first compute $\mathbf{F}_Y^{-1}$. For a 2×2 matrix in Eq. (103) with the determinant $F_1F_2 - J^2$, one has

$$\mathbf{F}_Y(\boldsymbol{\theta})^{-1} = \frac{1}{F_1F_2 - J^2}\begin{pmatrix} F_2 & -J \\ -J & F_1 \end{pmatrix}. \tag{112}$$

Therefore,

$$\mathbf{u}^T \mathbf{F}_Y^{-1} \mathbf{u} = \frac{1}{F_1F_2 - J^2}\left(F_1 + F_2 - 2J\right), \tag{113}$$

and Eq. (107) follows by inversion.

To compare with the series benchmark in Eq. (108), we define the resistance quantities

$$\begin{aligned} R_{\text{eff}}(\boldsymbol{\theta}) &= \left(F_Y^{(\mathbf{u})}(\boldsymbol{\theta})\right)^{-1} = \frac{F_1 + F_2 - 2J}{F_1F_2 - J^2}, \\ R_{\text{series}}(\boldsymbol{\theta}) &= \left(F_{\text{series}}(\boldsymbol{\theta})\right)^{-1} = \frac{F_1 + F_2}{F_1F_2}. \end{aligned} \tag{114}$$

The inequality $F_Y^{(\mathbf{u})} > F_{\text{series}}$ is equivalent to $R_{\text{eff}} < R_{\text{series}}$. A direct algebraic manipulation yields

$$R_{\text{series}} - R_{\text{eff}} = \frac{J\left(J(F_1 + F_2) - 2F_1F_2\right)}{F_1F_2\left(J^2 - F_1F_2\right)}. \tag{115}$$

Since $F_1F_2 - J^2 > 0$, the denominator in Eq. (115) is negative. Therefore, $R_{\text{series}} - R_{\text{eff}} > 0$ holds if and only if

$$J\left(J(F_1 + F_2) - 2F_1F_2\right) < 0, \tag{116}$$

which is equivalent to the strict range in Eq. (110). This proves the condition for beating the series law.

For the optimization over J, observe from Eq. (114) that maximizing $F_Y^{(\mathbf{u})}$ is equivalent to minimizing R_{eff} over J with $J^2 < F_1F_2$. By differentiating R_{eff} with respect to J, we have

$$\frac{\partial}{\partial J} R_{\text{eff}} = \frac{-2(J - F_1)(J - F_2)}{\left(J^2 - F_1F_2\right)^2}. \tag{117}$$

If $F_1 \neq F_2$, the unique stationary point in the allowed interval $\left|J\right| < \sqrt{F_1F_2}$ is $J = \min\{F_1, F_2\}$, and substituting this value into Eq. (107) yields $F_Y^{(\mathbf{u})} = \min\{F_1, F_2\}$. In the symmetric case $F_1 = F_2 =: F$, Eq. (107) simplifies to $F_Y^{(\mathbf{u})} = (F + J)/2$, whose supremum over $\left|J\right| < F$ is F, approached as $J \to F^-$. Hence, in all cases, the supremum equals Eq. (111). □

Theorem S14 reveals a unifying mechanism. In a classical modular causal decomposition, score contributions are orthogonal and $J = 0$, forcing the harmonic series penalty. Causal-model collapse manifests as a nonzero J, and when J is positive in the range in Eq. (110), the effective resistance for the total parameter drops below the classical series value. The gain is therefore not mysterious: it is the information-theoretic shadow of an interference-like correlation between modules that classical causality forbids.

Remark S3 (The synergy in the Fisher information geometry)**.** *This interpretation is precisely what makes the framework practical. The designer does not need to guess which "quantum resource" is present in a vague sense. The design target is concrete: engineer the experiment so that the score contributions for the would-be causal modules become positively correlated. In quantum dynamics, such correlations naturally arise when an attempted segmentation into independent modules is inconsistent with coherent evolution. In contextual schemes, they arise when the statistics cannot be embedded into a context-free causal model. In both cases, the resource is the same mathematical object: synergy in the Fisher information geometry.*

E. Scaling intuition: cancelling the series penalty in long chains

The causal-path CFII already carries a message that is easy to miss if one reads it only as a bound on precision: the familiar "series penalty" is not a law of nature, but a law of a model class. In a strictly classical causal chain, the information about an additive parameter must traverse a sequence of causal bottlenecks, and the CFII forces the inverse Fisher information to accumulate. Once the classical causal modularization collapses (as certified by a CFII violation), the bottleneck accounting rule stops being applicable, and the very same experiment can exhibit an endpoint sensitivity that would be mathematically impossible under the classical causal narrative.

Why the $1/N$ dilution appears classically

To isolate the scaling, consider an N-parameter description $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)^T$ and an additive causal quantity

$$\Theta = \sum_{j=1}^{N} \theta_j = \mathbf{u}^T \boldsymbol{\theta}, \quad \mathbf{u} = (1, \ldots, 1)^T. \tag{118}$$

The appropriate single-number information about Θ in a multi-parameter model is the effective Fisher information

$$F^{(\mathbf{u})}(\boldsymbol{\theta}) = \left(\mathbf{u}^T \mathbf{F}(\boldsymbol{\theta})^{-1} \mathbf{u}\right)^{-1}, \tag{119}$$

where $\mathbf{F}(\boldsymbol{\theta})$ is the Fisher information matrix (FIM). This definition is not cosmetic: it is precisely the Cramér–Rao geometry for estimating the sum direction in parameter space.

In a strictly modular classical chain, the segments behave as independent "causal modules" in the precise sense used to derive the causal-path CFII: score contributions associated with different segments are orthogonal, hence the FIM is diagonal (or block-diagonal in more general modular networks). For the simplest equal-strength case, one may write

$$\mathbf{F}_{\mathrm{cl}}(\boldsymbol{\theta}) = F \mathbb{1}_N, \tag{120}$$

so that $\mathbf{F}_{\mathrm{cl}}^{-1} = (1/F)\mathbb{1}_N$ (Here, $\hat{\mathbb{1}}_N$ is the $N \times N$ identity matrix) and therefore

$$\mathbf{u}^T \mathbf{F}_{\mathrm{cl}}^{-1} \mathbf{u} = \frac{N}{F}, \quad F_{\mathrm{cl}}^{(\mathbf{u})} = \frac{F}{N}. \tag{121}$$

This is exactly the classical "series penalty": when one tries to estimate the *total* parameter accumulated over N modules whose information contents are comparable, the effective information about the total parameter scales as $1/N$. Equivalently, the information-resistance $R := F^{-1}$ grows linearly with N. The physical picture is the one we emphasized in the causal-path CFII: a classical chain is a sequence of passive mediators, and passive resistances in series must add.

What changes under causal-model collapse: off-diagonal Fisher synergy

The CFII derivation identifies the microscopic reason behind the series penalty: modularity forces score orthogonality and therefore removes off-diagonal entries of $\mathbf{F}$. Under causal-model collapse, those orthogonality relations need not hold. The score contributions from different segments become correlated, and the FIM develops off-diagonal structure. This is the multi-segment analogue of the two-segment "synergy" mechanism discussed in Sec. V D: the gain is the information-geometric shadow of correlations between score contributions that the classical causal model forbids.

To see how such correlations can cancel the series penalty in a transparent and analytically solvable way, it is useful to introduce a minimal toy model that cleanly separates diagonal "local information" from off-diagonal "collective synergy." Thus, let us consider an "equal-information, equal-correlation" FIM of the form

$$\mathbf{F}(\boldsymbol{\theta}) = F\left((1-\varepsilon)\mathbb{1}_N + \varepsilon \mathbf{J}_N\right), \tag{122}$$

where $\mathbf{J}_N$ is the all-ones matrix. This corresponds to a score vector $\mathbf{s} = (s_1, \ldots, s_N)^T$ whose covariance is $\mathbf{F} = \mathbb{E}[\mathbf{s}\mathbf{s}^T]$ with

$$\mathbb{E}[s_j^2] = F, \quad \mathbb{E}[s_i s_j] = F\varepsilon \quad (i \neq j), \tag{123}$$

so that ε directly quantifies the uniform pairwise correlation of segment-wise score contributions. For $0 \leq \varepsilon < 1$, $\mathbf{F}$ is positive definite (and hence invertible), while $\varepsilon \to 1$ corresponds to an almost perfectly correlated score structure.

There are two equally instructive ways to compute $F^{(\mathbf{u})}$: an eigenmode decomposition and a direct inversion.

Eigenmode viewpoint: The vector $\mathbf{u}$ is an eigenvector of $\mathbf{J}_N$ with eigenvalue N, and any vector orthogonal to $\mathbf{u}$ is an eigenvector of $\mathbf{J}_N$ with eigenvalue 0. Hence, $\mathbf{u}$ is an eigenvector of $\mathbf{F}$ with eigenvalue $F(1-\varepsilon+\varepsilon N)$, while the $(N-1)$ orthogonal directions have eigenvalue $F(1-\varepsilon)$. Since the total parameter Θ lives exactly in the $\mathbf{u}$ direction, the relevant information for estimating Θ is controlled by the inverse eigenvalue along $\mathbf{u}$, yielding

$$\mathbf{u}^T \mathbf{F}(\boldsymbol{\theta})^{-1} \mathbf{u} = \frac{N}{F(1-\varepsilon+\varepsilon N)}. \tag{124}$$

Therefore, the effective Fisher information becomes

$$F^{(\mathbf{u})}(\boldsymbol{\theta}) = \left(\mathbf{u}^T \mathbf{F}(\boldsymbol{\theta})^{-1} \mathbf{u}\right)^{-1} = \frac{F(1-\varepsilon+\varepsilon N)}{N} = F\left(\varepsilon + \frac{1-\varepsilon}{N}\right). \tag{125}$$

Direct inversion viewpoint: One may also verify Eq. (124) by explicitly inverting Eq. (122). By using $\mathbf{J}_N^2 = N\mathbf{J}_N$ and a short algebra, we have

$$\mathbf{F}(\boldsymbol{\theta})^{-1} = \frac{1}{F(1-\varepsilon)}\left(\mathbb{1}_N - \frac{\varepsilon}{1-\varepsilon+\varepsilon N}\mathbf{J}_N\right), \tag{126}$$

and contracting with $\mathbf{u}$ yields $\mathbf{u}^T\mathbf{F}^{-1}\mathbf{u}$ in Eq. (124).

Eq. (125) makes the scaling mechanism completely explicit. When $\varepsilon = 0$ (no synergy), one recovers the classical harmonic scaling $F^{(\mathbf{u})} = F/N$. When $\varepsilon > 0$ is fixed and N becomes large, the second term $(1-\varepsilon)/N$ dies out and one finds the saturation

$$\lim_{N\to\infty} F^{(\mathbf{u})}(\boldsymbol{\theta}) = F\varepsilon, \tag{127}$$

namely, an $O(1)$ effective Fisher information for the total parameter Θ even for arbitrarily long chains. In the extreme synergy limit $\varepsilon \to 1$ (maximal correlation compatible with invertibility), Eq. (125) yields $F^{(\mathbf{u})} \to F$, i.e., the chain behaves as if there were no series penalty at all.

It is worth pausing on what this means physically. The equicorrelated matrix has a simple interpretation: it decomposes the information geometry into one "collective mode" along $\mathbf{u}$ and $(N-1)$ "relative modes" orthogonal to it. As $\varepsilon \to 1$, the collective mode becomes highly informative, while the relative modes become poorly conditioned (the eigenvalue $F(1-\varepsilon)$ tends to 0). This captures an important metrological moral: the same correlations that make the sum parameter Θ easy to estimate can make the individual segment parameters hard to disentangle. In other words, the synergy is not magic that creates information about everything, but it is a geometric reallocation of information into the direction that the estimation task actually cares about.

From scaling intuition to operational design

The relevance to our causal perspective is immediate. A long chain is precisely the regime where classical causal reasoning would insist that precision must become hopelessly diluted, because each causal module adds an irreducible information-resistance. A large CFII violation is precisely the regime where Nature declares that this dilution narrative is not merely pessimistic but structurally wrong for the observed statistics. In such a regime, the causal-model collapse is not only a philosophical diagnosis; it is a design instruction: engineer the experiment so that the would-be segment-wise score contributions become positively correlated, thereby generating off-diagonal Fisher synergy and converting a classically series-limited architecture into an effectively collective sensor.

F. CFII-guided metrological design

The results above suggest a practical methodology for metrological design. A CFII is derived from an explicit classical causal hypothesis. If the hypothesis were true, it would impose a precision frontier (Sec. V A). If the experiment violates the CFII, the violation simultaneously (i) falsifies the hypothesis (Sec. IV) and (ii) certifies a precision gain that can be harvested by an efficient estimator (Sec. V B and Sec. V C). Thus, a CFII is more than a test; it is a metrology design compass.

The most important practical implication is that the "resource" is *self-certified*. To claim a metrological advantage, one typically must compare to a theoretical bound (such as a QFI benchmark) and argue that the implementation truly adheres to the assumptions behind that bound. In contrast, a CFII-based certification proceeds internally: it compares the achieved endpoint information to a causal benchmark derived from an explicit classical model. When the inequality is violated, the classical benchmark is not merely beaten; it is proven inapplicable. The very data that achieve the enhanced precision also certify that the enhancement arises from causal-model collapse. This is precisely the kind of internal certification that is valuable in practical sensing, where device imperfections often blur the meaning of idealized theoretical limits.

Equally important is that the mechanism identified in Sec. V D translates into a concrete control objective. A classical causal decomposition forces the score covariance J to vanish. A metrological protocol that aims to exploit causal-model collapse should therefore be designed to generate positive score correlations between the would-be causal modules, while maintaining estimator stability (regularity of the likelihood and positivity of the effective statistical model used by the inference procedure). When these conditions are met, the improvement factor in Eq. (88) becomes the experimentally measurable figure of merit: by tuning preparation, controls, and measurement contexts to push V_{path} deeper below zero, one directly tunes the achievable estimation error.

SUPPLEMENTARY NOTE VI. EXAMPLES: COHERENT DYNAMICS BEYOND THE CLASSICAL CAUSAL-PATH FRONTIER

In the preceding sections, we established that a Fisher-information inequality becomes a *causal* inequality once one commits to a classical causal model class. In particular, under the causal-path hypothesis $A \to C \to B$ with a modular (segment-wise) parameterization, Fisher information must obey a strict series law. When the series law is violated, the conclusion is that the entire classical causal-path explanation is falsified.

Here we illustrate these messages in an elementary coherent-dynamics setting. A single qubit undergoing a coherent rotation violates the causal-path CFII even with a fixed, simple projective readout. We then show that the resulting causal-path collapse is visible in Fisher-information landscapes, estimator-level performance, and adversarial classical benchmarks that optimize the intermediate split time.

A. Unified benchmark: CFII gap and its metrological meaning

We start with an additive parameter decomposition

$$\theta_{ab} = \theta_{ac} + \theta_{cb}, \quad (\theta_{ac}, \theta_{cb} > 0), \tag{128}$$

and let $F(\theta)$ denote the (classical) Fisher information of the chosen readout statistics for estimating θ after an evolution of duration θ.

Under the classical causal-path model class formalized in Sec. III, the causal-path CFII reads

$$F(\theta_{ab})^{-1} \geq F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}, \tag{129}$$

and the violation statistic is defined by

$$V(\theta_{ac}, \theta_{cb}) = F(\theta_{ab})^{-1} - F(\theta_{ac})^{-1} - F(\theta_{cb})^{-1}. \tag{130}$$

The causal-path hypothesis requires $V(\theta_{ac}, \theta_{cb}) \geq 0$ for all admissible splits. Thus, $V < 0$ is a model-falsification certificate in the precise sense of Sec. IV.

To connect the witness to the operational precision, we define the classical causal-path benchmark Fisher information for a given split,

$$F_{\mathrm{cl}}^{(\mathrm{path})}(\theta_{ac}, \theta_{cb}) = \left(F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}\right)^{-1}. \tag{131}$$

Then, Eq. (129) is equivalent to $F(\theta_{ab}) \leq F_{\rm cl}^{(\rm path)}$. Since the Cramér–Rao bound gives $\Delta\theta \geq 1/\sqrt{NF(\theta)}$ for N i.i.d. samples [2, 3], the causal-path model predicts an error floor

$$\Delta\theta_{\rm cl} \geq \frac{1}{\sqrt{NF_{\rm cl}^{(\rm path)}(\theta_{ac},\theta_{cb})}}. \tag{132}$$

In contrast, the actual experiment yields

$$\Delta\theta \geq \frac{1}{\sqrt{NF(\theta_{ab})}}. \tag{133}$$

Here we package the advantage by the error-ratio indicator

$$G(\theta_{ac},\theta_{cb}) := \log\left(\frac{\Delta\theta}{\Delta\theta_{\rm cl}}\right) = \frac{1}{2}\log\left(\frac{F_{\rm cl}^{(\rm path)}(\theta_{ac},\theta_{cb})}{F(\theta_{ab})}\right). \tag{134}$$

Note that $G < 0$ is exactly the metrological-gain regime. Thus, we can summarize as:

Theorem S15 (Witness–resource equivalence for the causal-path CFII)**.** *For $F(\theta_{ab}), F(\theta_{ac}), F(\theta_{cb}) > 0$, the following statements are equivalent:*

$$V(\theta_{ac},\theta_{cb}) < 0, \quad F(\theta_{ab}) > F_{\rm cl}^{(\rm path)}(\theta_{ac},\theta_{cb}), \quad G(\theta_{ac},\theta_{cb}) < 0. \tag{135}$$

Proof. By definition,

$$\begin{aligned} V(\theta_{ac},\theta_{cb}) < 0 &\Longleftrightarrow F(\theta_{ab})^{-1} < F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1} \\ &\Longleftrightarrow F(\theta_{ab}) > \left(F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}\right)^{-1} = F_{\rm cl}^{(\rm path)}(\theta_{ac},\theta_{cb}). \end{aligned} \tag{136}$$

By substituting this inequality into Eq. (134), we obtain

$$F(\theta_{ab}) > F_{\rm cl}^{(\rm path)}(\theta_{ac},\theta_{cb}) \Longleftrightarrow \frac{F_{\rm cl}^{(\rm path)}(\theta_{ac},\theta_{cb})}{F(\theta_{ab})} < 1 \Longleftrightarrow G(\theta_{ac},\theta_{cb}) < 0. \tag{137}$$

The proof is completed. □

Theorem S15 is the operational core and summary of Sec. V. Thus, a CFII violation is simultaneously a falsification of the classical causal-path model class and a guaranteed precision advantage relative to that class.

B. Single-qubit coherent rotation

We begin with a single qubit evolving under a coherent unitary rotation. We cast a simple model:

$$\hat{H} = \frac{1}{2}\hat{\sigma}_x, \quad \hat{U}(\theta) = e^{-i\theta\hat{H}} = e^{-i\theta\hat{\sigma}_x/2}. \tag{138}$$

We prepare a (pure) probe state

$$|\psi(\vartheta,\varphi)\rangle = \cos\frac{\vartheta}{2}|0\rangle + e^{i\varphi}\sin\frac{\vartheta}{2}|1\rangle, \tag{139}$$

and perform a fixed projective measurement in the $\hat{\sigma}_z$ basis,

$$\hat{M}_0 = |0\rangle\langle 0|, \quad \hat{M}_1 = |1\rangle\langle 1|. \tag{140}$$

Writing the binary probabilities as $p_0(\theta) = \frac{1+z(\theta)}{2}$ and $p_1(\theta) = \frac{1-z(\theta)}{2}$, one finds

$$z(\theta) = \cos\vartheta\cos\theta + \sin\vartheta\sin\varphi\sin\theta. \tag{141}$$

For binary statistics, the Fisher information is given by

$$F(\theta) = \sum_{m=0,1} p_m(\theta)\left(\frac{\partial}{\partial\theta}\ln p_m(\theta)\right)^2 = \frac{\left(\frac{\partial}{\partial\theta}z(\theta)\right)^2}{1-z(\theta)^2}, \tag{142}$$

where the removable singularities at $z(\theta)^2 = 1$ are treated by continuity.

Then, we provide the result that shows that the causal-path CFII can be violated in a maximally clean way: no entanglement, no post-selection, no measurement optimization. A single fixed projective measurement suffices, yet the classical causal-path model collapses deterministically.

Result S1 (Deterministic single-qubit collapse of the causal-path model)**.** *Fix $\varphi = \pi/2$ in Eq. (139). Then, the Fisher information of the readout in Eq. (140) is constant,*

$$F(\theta) = 1 \quad (\forall\, \theta), \tag{143}$$

and the causal-path CFII gap in Eq. (130) satisfies

$$V(\theta_{ac}, \theta_{cb}) = -1 \quad (\forall\, \theta_{ac}, \theta_{cb} > 0). \tag{144}$$

Equivalently, $F_{\mathrm{cl}}^{(\mathrm{path})} = \frac{1}{2}$ and the achieved Fisher information exceeds the classical causal-path benchmark such that:

$$\frac{F(\theta_{ab})}{F_{\mathrm{cl}}^{(\mathrm{path})}(\theta_{ac}, \theta_{cb})} = 2. \tag{145}$$

The deterministic point makes the causal content unmistakable. A classical causal-path narrative demands that information about θ_{ab} must traverse an intermediate classical mediator C and therefore pay a series penalty. The coherent qubit rotation refuses: it yields twice the Fisher information permitted by any such bottleneck.

This advantage is not merely a bound-level statement. At the deterministic point, the endpoint model is a simple binary fringe $p_0(\theta) = \cos^2((\theta - \vartheta)/2)$. With N i.i.d. samples, a maximum-likelihood estimate is obtained by matching the frequency $\widehat{p_0} = n_0/N$ and inverting

$$\widehat{\theta} = \vartheta + 2 \arccos \sqrt{\widehat{p_0}}, \tag{146}$$

on an interval where the mapping is one-to-one. Asymptotically, the estimator achieves $\Delta\theta \simeq 1/\sqrt{N}$, whereas the classical causal-path frontier predicts $\Delta\theta_{\mathrm{cl}} \geq \sqrt{2}/\sqrt{N}$.

C. Generic regimes, numerical landscapes, and adversarial benchmarks

The deterministic result above might appear as a special symmetric point. Its true role is conceptual: at that point the Fisher information is constant, and the classical series law fails uniformly for every split. The natural question is what happens away from such a symmetry. The answer is that the uniform collapse deforms into structured regions of collapse rather than disappearing. Thus, here we show: (i) a representative landscape where $V < 0$ and the metrological advantage $G < 0$ persist broadly, (ii) estimator-level achievability of the advantage, and (iii) robustness against an "adversarial" classical benchmark that optimizes the split time to help the classical model as much as possible.

(i) *Representative landscape: $V < 0$ and $G < 0$ in a generic single-qubit setting*

The CFII gap V and the gain indicator G for a representative nontrivial single-qubit setting are visualized in Fig. 1: a generic probe $(\vartheta, \varphi) = (0.7\pi, 0.3\pi)$ with $\hat{\sigma}_z$ readout.

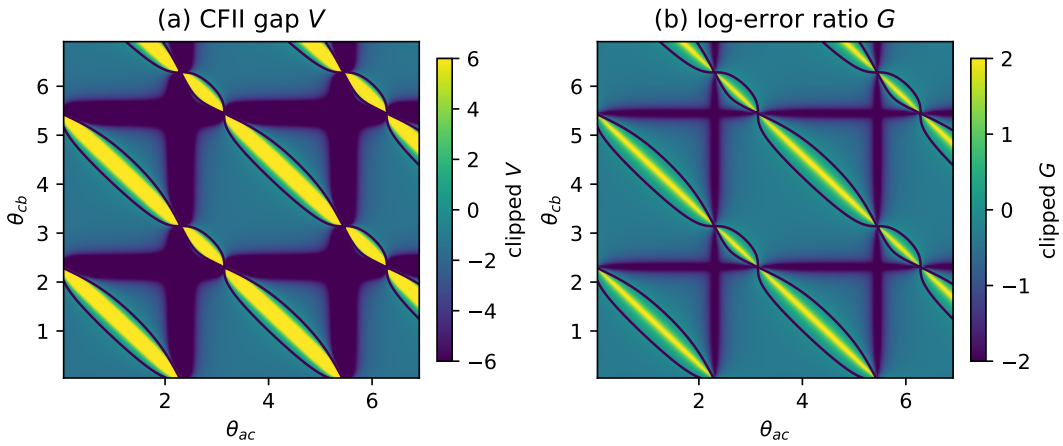
By Theorem S15, the negative region in the V landscape is exactly the negative region in the G landscape. Hence, the figure simultaneously shows where the classical causal-path explanation fails and where the metrological advantage over that explanation is guaranteed.

(ii) *Achievability: the gain is not only a bound, but an estimator-level reality*

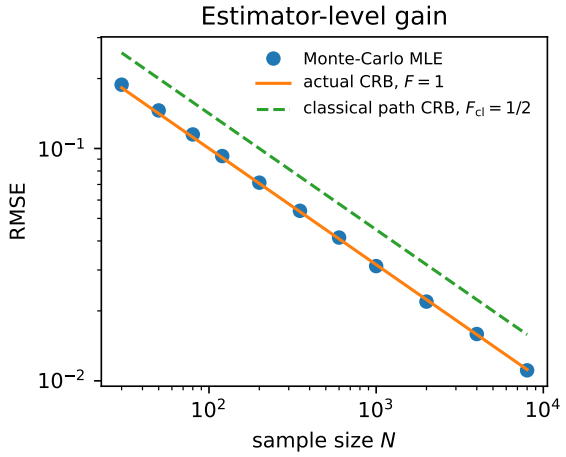
Fig. 2 demonstrates that the gain is operationally attainable. For the deterministic point of **Result S1**, the endpoint likelihood is a simple binary fringe, and a maximum-likelihood estimator saturates the Cramér–Rao bound asymptotically. The Monte-Carlo RMSE scaling shows the RMSE approaching $1/\sqrt{NF(\theta_{ab})}$ while staying strictly below the classical causal-path bound $1/\sqrt{NF_{\mathrm{cl}}^{(\mathrm{path})}}$ by the predicted factor $\sqrt{2}$ in standard deviation.

(iii) *Adversarial classical benchmark: optimizing the split time*

A skeptical classicalist might attempt the following escape: "Perhaps the CFII violation occurs only because we chose an unfortunate split time t_c. If we pick a different intermediate time, maybe the inequality recovers." Within the causal-path hypothesis, this intuition is already inconsistent: a trajectory-like classical mediator should exist at any intermediate time, so



Supplementary Fig. 1. Representative single-qubit landscapes for the generic setting $(\vartheta, \varphi) = (0.7\pi, 0.3\pi)$ with $\hat{\sigma}_z$ readout. (a) $V(\theta_{ac}, \theta_{cb})$ in Eq. (130) (clipped). (b) $G(\theta_{ac}, \theta_{cb})$ in Eq. (134) (clipped), where $G < 0$ certifies an operational precision advantage over the classical causal-path benchmark.



Supplementary Fig. 2. Achievability of the advantage for the deterministic single-qubit point. Monte-Carlo RMSE of a simple MLE (markers) versus sample size N (log–log), compared to the quantum CRB $1/\sqrt{NF(\theta_{ab})}$ (solid) and the classical causal-path bound $1/\sqrt{NF_{\mathrm{cl}}^{(\mathrm{path})}}$ (dashed). Here $F = 1$ and $F_{\mathrm{cl}} = 1/2$.

the CFII must hold for every split. Nevertheless, to make the comparison maximally conservative, we now give the classical causal-path model the power to choose the split that benefits it the most.

For a fixed total interval θ_{ab}, define the split-optimized classical benchmark

$$F_{\mathrm{cl}}^{(\mathrm{opt})}(\theta_{ab}) := \max_{0<\theta_{ac}<\theta_{ab}} \left(F(\theta_{ac})^{-1} + F(\theta_{ab}-\theta_{ac})^{-1}\right)^{-1}. \tag{147}$$

We also define the optimized gain ratio

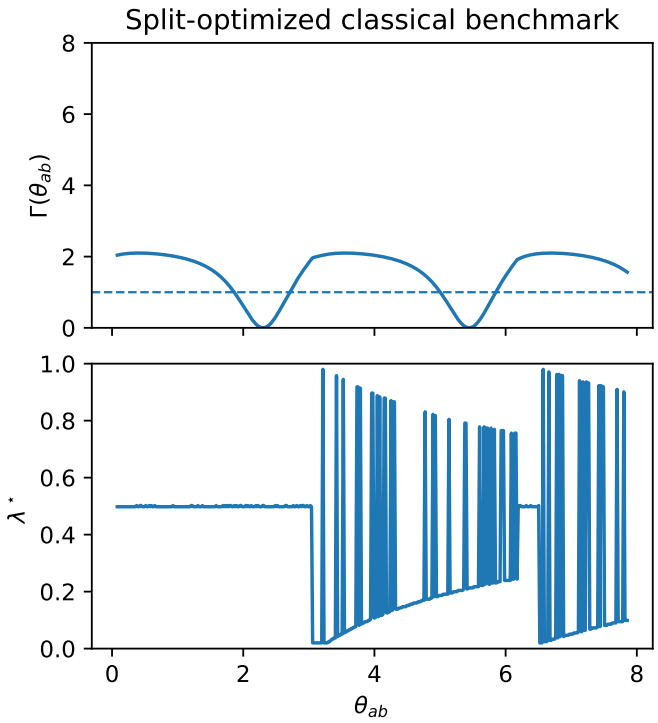
$$\Gamma(\theta_{ab}) := \frac{F(\theta_{ab})}{F_{\mathrm{cl}}^{(\mathrm{opt})}(\theta_{ab})}, \tag{148}$$

and the corresponding optimal split fraction

$$\lambda^\star(\theta_{ab}) := \frac{\theta_{ac}^\star(\theta_{ab})}{\theta_{ab}}, \tag{149}$$

where

$$\theta_{ac}^\star(\theta_{ab}) \in \arg\max_{0<\theta_{ac}<\theta_{ab}} \left(F(\theta_{ac})^{-1} + F(\theta_{ab}-\theta_{ac})^{-1}\right)^{-1}. \tag{150}$$



Supplementary Fig. 3. Adversarial (split-optimized) classical benchmark for the representative single-qubit setting used in Fig. 1. Top: the ratio $\Gamma(\theta_{ab}) = F(\theta_{ab})/F_{\mathrm{cl}}^{(\mathrm{opt})}(\theta_{ab})$ from Eq. (148). The dashed line $\Gamma = 1$ is the most generous classical causal-path frontier once the split is optimized; $\Gamma > 1$ certifies the split-robust causal-model collapse (**Result S2**). Bottom: the optimal split fraction $\lambda^\star(\theta_{ab}) = \theta^\star_{ac}/\theta_{ab}$ from Eq. (149).

The line $\Gamma(\theta_{ab}) = 1$ is the most generous classical causal-path frontier once the split is optimized. Here, we can prove the following result:

Result S2 (Split-robust collapse under the optimized classical benchmark)**.** *Fix θ_{ab} and define $F_{\mathrm{cl}}^{(\mathrm{opt})}(\theta_{ab})$ by Eq. (147). If the following holds*

$$\Gamma(\theta_{ab}) > 1, \tag{151}$$

then the causal-path CFII is violated for every admissible split $\theta_{ab} = \theta_{ac} + \theta_{cb}$:

$$V(\theta_{ac}, \theta_{cb}) < 0 \quad \forall\, \theta_{ac} \in (0, \theta_{ab}),\ \theta_{cb} = \theta_{ab} - \theta_{ac}. \tag{152}$$

In particular, no choice of intermediate time t_c can restore a classical causal-path explanation for that θ_{ab}.

Proof. —For each admissible split, let us define the split-dependent benchmark

$$F_{\mathrm{cl}}^{(\mathrm{path})}(\theta_{ac}, \theta_{cb}) = \left(F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}\right)^{-1}. \tag{153}$$

By definition of the maximum in Eq. (147),

$$F_{\mathrm{cl}}^{(\mathrm{path})}(\theta_{ac}, \theta_{cb}) \le F_{\mathrm{cl}}^{(\mathrm{opt})}(\theta_{ab}) \quad \text{for all admissible splits.} \tag{154}$$

If $\Gamma(\theta_{ab}) > 1$, then $F(\theta_{ab}) > F_{\mathrm{cl}}^{(\mathrm{opt})}(\theta_{ab})$. By combining with Eq. (154), we obtain

$$F(\theta_{ab}) > F_{\mathrm{cl}}^{(\mathrm{path})}(\theta_{ac}, \theta_{cb}) \quad \text{for all admissible splits.} \tag{155}$$

This is equivalent to $V(\theta_{ac}, \theta_{cb}) < 0$ by the definition of Eq. (130). Thus, the proof is completed. □

Fig. 3 shows $\Gamma(\theta_{ab})$ and $\lambda^\star(\theta_{ab})$ for the generic single-qubit setting used in Fig. 1. The top panel shows that $\Gamma(\theta_{ab})$ exceeds unity over wide ranges of θ_{ab} even after the split is optimized. This is precisely the "even an adversarial classical benchmark cannot catch up" regime: the causal-path explanation collapses regardless of how one tries to place the intermediate classical mediator.

D. Chain-amplified gain in long causal-path decompositions

The classical causal-path narrative is, in fact, far more committal than the existence of one intermediate node. If the evolution is truly describable by a classical trajectory-like mediation, then every refinement of the path into more intermediate bottlenecks must also admit a consistent classical causal decomposition. This observation leads to a stronger and more dramatic test (as already discussed in Sec. V E): instead of a single split $\theta_{ab} = \theta_{ac} + \theta_{cb}$, we impose an entire discrete chain of K segments, and ask whether the classical "series law" can survive such a refined causal claim.

From a single split to a K-step causal chain

Fix a total interval $\theta_{ab} > 0$ and choose intermediate times $t_0 = t_a < t_1 < \cdots < t_{K-1} < t_K = t_b$, so that the total parameter decomposes additively as

$$\theta_{ab} = \sum_{j=1}^{K} \theta_{j-1,j}, \quad \theta_{j-1,j} := t_j - t_{j-1} > 0. \tag{156}$$

For the same measurement used in Sec. VI B, let $F(\theta)$ denote the (classical) Fisher information of the endpoint measurement statistics after evolution duration θ. The K-step causal-path hypothesis asserts that the information about θ_{ab} must traverse $K-1$ intermediate classical bottlenecks in series, with modular parameter dependence on each link. By the N-step chain CFII (**Corollary S2**), this implies the multi-split series law

$$F(\theta_{ab})^{-1} \geq \sum_{j=1}^{K} F(\theta_{j-1,j})^{-1}. \tag{157}$$

It is convenient to package this into a K-step violation statistic,

$$V_K(\theta_{0,1}, \ldots, \theta_{K-1,K}) := F(\theta_{ab})^{-1} - \sum_{j=1}^{K} F(\theta_{j-1,j})^{-1}, \tag{158}$$

so that the classical causal-path model class requires $V_K \geq 0$ for all admissible partitions of θ_{ab}. The corresponding K-step classical benchmark Fisher information is the harmonic-series composition

$$F_{\mathrm{cl}}^{(K)}(\theta_{0,1}, \ldots, \theta_{K-1,K}) := \left(\sum_{j=1}^{K} F(\theta_{j-1,j})^{-1} \right)^{-1}. \tag{159}$$

Hence, $V_K < 0$ is again a witness-resource certificate: it simultaneously falsifies the K-step classical causal-path class and certifies a precision gap relative to that class.

Here we summarize our result:

Result S3 (Chain-amplified collapse and gain for the single-qubit coherent example)**.** *Assume that, for the fixed readout employed in Sec. VI B, the Fisher information is constant,*

$$F(\theta) = F_0 > 0 \quad (\forall\, \theta > 0). \tag{160}$$

Then, for every integer $K \geq 2$ and for every admissible partition $\theta_{ab} = \sum_{j=1}^{K} \theta_{j-1,j}$ with $\theta_{j-1,j} > 0$, the K-step causal-path CFII is violated uniformly:

$$V_K(\theta_{0,1}, \ldots, \theta_{K-1,K}) = -\frac{K-1}{F_0} < 0. \tag{161}$$

Equivalently, the K-step classical benchmark equals

$$F_{\mathrm{cl}}^{(K)}(\theta_{0,1}, \ldots, \theta_{K-1,K}) = \frac{F_0}{K}, \tag{162}$$

and the certified gain ratio is exactly linear in the chain depth,

$$\Gamma_K(\theta_{ab}) = \frac{F(\theta_{ab})}{F_{\mathrm{cl}}^{(K)}(\theta_{0,1}, \ldots, \theta_{K-1,K})} = K. \tag{163}$$

In particular, for N_{s} i.i.d. repetitions of the endpoint experiment, the classical causal-path model predicts an error floor

$$\Delta\theta_{\mathrm{cl}}^{(K)} \geq \frac{1}{\sqrt{N_{\mathrm{s}} F_{\mathrm{cl}}^{(K)}}} = \sqrt{\frac{K}{N_{\mathrm{s}} F_0}}, \tag{164}$$

whereas the actual coherent-dynamics readout achieves asymptotically

$$\Delta\theta \simeq \frac{1}{\sqrt{N_{\mathrm{s}} F(\theta_{ab})}} = \frac{1}{\sqrt{N_{\mathrm{s}} F_0}}, \tag{165}$$

thus beating the K-step classical causal-path frontier by a factor $\sqrt{K}$ in standard deviation.

Proof. —Under the assumption $F(\theta) = F_0$, one has $F(\theta_{ab}) = F_0$ and $F(\theta_{j-1,j}) = F_0$ for all segments. By substituting into the definition of V_K, we find

$$V_K = F_0^{-1} - \sum_{j=1}^{K} F_0^{-1} = \frac{1-K}{F_0} = -\frac{K-1}{F_0} < 0, \tag{166}$$

proving the uniform violation. Similarly,

$$F_{\mathrm{cl}}^{(K)} = \left(\sum_{j=1}^{K} F_0^{-1} \right)^{-1} = \left(\frac{K}{F_0} \right)^{-1} = \frac{F_0}{K}, \tag{167}$$

and therefore $\Gamma_K = F_0/(F_0/K) = K$. The error statements follow from the Cramér–Rao scaling $\Delta\theta \simeq 1/\sqrt{N_{\mathrm{s}} F}$ applied to the endpoint Fisher information and to the classical benchmark Fisher information. □

Result S3 applies immediately to the deterministic coherent point established earlier. For the single-qubit setting of Sec. VI B at $\varphi = \pi/2$, **Result S1** yields $F(\theta) = 1$ for all θ, hence $F_0 = 1$ and the chain-amplified violation reads

$$V_K = -(K-1), \quad F_{\mathrm{cl}}^{(K)} = \frac{1}{K}, \quad \Gamma_K = K. \tag{168}$$

The metrological meaning is clear: the endpoint statistics behave as if the total parameter were estimated without suffering the classical series dilution, whereas any classical K-step causal-path explanation must pay a $1/K$ penalty in Fisher information.

In the causal language of this work, increasing K does not correspond to increasing experimental resources. Rather, it corresponds to tightening the classical causal claim. A K-step causal-path model asserts the existence of $K-1$ intermediate classical mediators such that the parameter dependence modularizes across the chain. Each additional mediator is a new bottleneck and therefore contributes an additional information-resistance term $F(\theta_{j-1,j})^{-1}$ that must add in series. The coherent dynamics violates the modularity that enforces this orthogonality of score contributions, and the consequence is as dramatic as it is quantitative: the more finely one insists on discretizing a classical trajectory, the more severely the classical series law is contradicted.

Adversarially optimized multi-split benchmarks

One may extend the "adversarial benchmark" logic of Sec. VI C(iii) from a single split to an entire partition. Let us define the K-split-optimized classical benchmark by

$$F_{\mathrm{cl}}^{(K,\mathrm{opt})}(\theta_{ab}) := \max_{\substack{\theta_{0,1},\ldots,\theta_{K-1,K}>0 \\ \sum_{j=1}^{K} \theta_{j-1,j} = \theta_{ab}}} \left(\sum_{j=1}^{K} F(\theta_{j-1,j})^{-1} \right)^{-1}, \tag{169}$$

and the corresponding optimized gain ratio

$$\Gamma_K^{(\mathrm{opt})}(\theta_{ab}) := \frac{F(\theta_{ab})}{F_{\mathrm{cl}}^{(K,\mathrm{opt})}(\theta_{ab})}. \tag{170}$$

In the constant Fisher information coherent point of **Result S1**, the optimization is powerless: since $F(\theta) = F_0$ for all segment lengths, every partition yields the same harmonic value F_0/K, and therefore

$$F_{\mathrm{cl}}^{(K,\mathrm{opt})}(\theta_{ab}) = \frac{F_0}{K}, \quad \Gamma_K^{(\mathrm{opt})}(\theta_{ab}) = K. \tag{171}$$

Thus, the chain-amplified collapse is not only split-robust; it is partition-robust.

More generally, whenever the chosen readout yields a bounded Fisher information $F(\theta) \le F_{\max}$ over the relevant domain of segment lengths, the classicalist cannot evade the series penalty even by optimizing the entire partition:

$$F_{\rm cl}^{(K,\rm opt)}(\theta_{ab}) = \max \left(\sum_{j=1}^{K} F(\theta_{j-1,j})^{-1} \right)^{-1} \le \left(\sum_{j=1}^{K} F_{\max}^{-1} \right)^{-1} = \frac{F_{\max}}{K}. \tag{172}$$

Hence, a $1/K$ dilution of the classical benchmark is an unavoidable consequence of insisting on K passive classical bottlenecks. Whenever the coherent dynamics maintains an $O(1)$ endpoint Fisher information for the total parameter (as it does at the deterministic point, and approximately does in broad regions of the generic landscape), the gain relative to the K-step causal-path frontier is naturally amplified with chain depth.

SUPPLEMENTARY NOTE VII. AI-ASSISTED ADVERSARIAL FINITE-DATA STRESS TEST

The examples in Sec. VI establish the CFII violation analytically and show that the resulting gain is estimator-achievable. In this section, we add a numerical stress test designed to address a more practical question: can the negative witness be certified from noisy finite data even when a flexible classical causal adversary is allowed to optimize its hidden mediator and stochastic kernels? The answer is affirmative. The purpose of the simulation is not to replace the theorems, but to make their operational content harder to dismiss: we deliberately give the classical causal-path model a strong, differentiable, AI-optimized adversary and verify that it can saturate, but not cross, the CFII frontier.

A. Noisy coherent-fringe data generator

We use the same binary readout structure as in the coherent single-qubit example, but include two elementary imperfections: visibility loss and symmetric readout error. The simulated endpoint probabilities are

$$p_0^{(\gamma)}(\theta) = \frac{1 + z_\gamma(\theta)}{2}, \quad p_1^{(\gamma)}(\theta) = \frac{1 - z_\gamma(\theta)}{2}, \tag{173}$$

where

$$z_\gamma(\theta) = \eta_{\rm r} e^{-\gamma\theta} \cos(\theta - \vartheta_0), \quad \eta_{\rm r} = 1 - 2\epsilon_{\rm r}. \tag{174}$$

Here γ is a dimensionless dephasing-rate parameter and $\epsilon_{\rm r}$ is a symmetric readout-error probability. The derivative is

$$\dot{z}_\gamma(\theta) = \frac{\partial z_\gamma(\theta)}{\partial\theta} = -\eta_{\rm r} e^{-\gamma\theta} \left[\gamma \cos(\theta - \vartheta_0) + \sin(\theta - \vartheta_0) \right], \tag{175}$$

and the binary Fisher information is

$$F_\gamma(\theta) = \frac{\dot{z}_\gamma(\theta)^2}{1 - z_\gamma(\theta)^2}. \tag{176}$$

The numerical results below use

$$T = \frac{\pi}{2}, \quad K = 4, \quad \vartheta_0 = 0, \quad \epsilon_{\rm r} = 0.02, \tag{177}$$

and compare the endpoint experiment at T with an equal K-segment classical causal-chain benchmark. The equal-segment witness and gain ratio are

$$V_K(\gamma) = F_\gamma(T)^{-1} - K F_\gamma(T/K)^{-1}, \tag{178}$$

$$\Gamma_K(\gamma) = \frac{F_\gamma(T)}{F_\gamma(T/K)/K}. \tag{179}$$

For example, at $\gamma = 0.25$ one obtains

$$F_\gamma(T) = 0.4202, \quad F_\gamma(T/K) = 0.8065, \quad V_K = -2.5799, \quad \Gamma_K = 2.0841. \tag{180}$$

The optimized coherent advantage disappears only when Γ_K reaches unity; for the parameters above this crossing occurs near $\gamma \simeq 0.444$.

B. Finite-shot FI estimation and delta-method certification

For the binary model in Eq. (173), the score values are explicit:

$$s_0(\theta) = \frac{\dot{z}_\gamma(\theta)}{1 + z_\gamma(\theta)}, \quad s_1(\theta) = -\frac{\dot{z}_\gamma(\theta)}{1 - z_\gamma(\theta)}. \tag{181}$$

Given N_s independent samples in a context θ_s, the plug-in FI estimator is

$$\widehat{F}_s = \frac{1}{N_s} \sum_{i=1}^{N_s} s_{x_i}(\theta_s)^2. \tag{182}$$

This is the concrete version of the score estimator discussed in Sec. IV D. Let

$$\mu_{4,s} = \sum_{x=0,1} p_x^{(\gamma)}(\theta_s) s_x(\theta_s)^4. \tag{183}$$

Then

$$\mathrm{Var}(\widehat{F}_s) = \frac{\mu_{4,s} - F_s^2}{N_s}. \tag{184}$$

Applying the delta method to $R_s = F_s^{-1}$ gives

$$\mathrm{Var}(\widehat{R}_s) \simeq \frac{\mu_{4,s} - F_s^2}{N_s F_s^4}. \tag{185}$$

For the equal-segment K-chain test with independent data in each context, the standard error of

$$\widehat{V}_K = \widehat{F}(T)^{-1} - \sum_{j=1}^{K} \widehat{F}(T/K)^{-1} \tag{186}$$

is therefore

$$\mathrm{SE}(\widehat{V}_K)^2 \simeq \frac{\mu_{4,T} - F_T^2}{N_s F_T^4} + K \frac{\mu_{4,T/K} - F_{T/K}^2}{N_s F_{T/K}^4}. \tag{187}$$

We report the certification significance as

$$Z = -\frac{V_K}{\mathrm{SE}(\widehat{V}_K)}. \tag{188}$$

For $\gamma = 0.25$ and $N_s = 10^3$ shots per context, Eq. (187) gives $\mathrm{SE} = 0.2121$ and hence $Z = 12.17$. A direct finite-shot Monte-Carlo simulation gives a 95% interval approximately $[-3.06, -2.22]$ for $\widehat{V}_K$, well separated from the classical boundary at zero.

C. Classifier score estimation

The previous subsection used the analytic score only to make the certification transparent. In an actual complex experiment, the likelihood and score may not be available in closed form. We therefore also implement a classifier likelihood-ratio estimator, a standard simulation-based inference idea [15]. Suppose that one can generate or collect samples at two nearby parameters $\theta + \delta$ and $\theta - \delta$. A binary classifier trained with equal class priors to distinguish these two sample sets has an optimal logit

$$g^\star(x) = \log \frac{p(x|\theta + \delta)}{p(x|\theta - \delta)}. \tag{189}$$

For small δ,

$$\frac{g^\star(x)}{2\delta} = \frac{\partial}{\partial \theta} \log p(x|\theta) + O(\delta^2), \tag{190}$$

so the FI can be estimated by averaging the squared classifier score. For the binary readout considered here, a saturated neural classifier is equivalent to a two-logit model over $x = 0, 1$. The closed-form finite-sample estimate used in the code is

$$\widehat{s}_{\mathrm{cls}}(x) = \frac{1}{2\delta} \log \frac{(n_{+,x} + \alpha)/(N_+ + 2\alpha)}{(n_{-,x} + \alpha)/(N_- + 2\alpha)}, \tag{191}$$

where $n_{\pm,x}$ are the counts of outcome x in the $\theta \pm \delta$ training samples and α is a smoothing parameter. In the reported figure we used $\delta = 0.10$ and $\alpha = 5$. The same formula is replaced by a neural network logit when x is high-dimensional.

D. Differentiable modular classical adversary

We now describe the AI-optimized classical causal adversary. The adversary is not allowed to abandon the causal assumptions that define the CFII. It must remain a modular causal path,

$$p_\phi(c, b|\theta_1, \theta_2) = \alpha_\phi(c|\theta_1)\, \beta_\phi(b|c, \theta_2), \tag{192}$$

with $\theta_1 = \theta_{ac}$ and $\theta_2 = \theta_{cb}$. However, within this constraint it is made deliberately flexible: c is a latent mediator with L possible values, b has M possible endpoint outcomes, and the local stochastic kernels are optimized by gradient descent.

Because a CFII is a local Fisher-information statement, it is enough to parameterize the local values and local derivatives of the kernels. We use softmax tangent models,

$$\alpha_\phi(c|\theta_1) = \mathrm{softmax}_c \left[a_c + \dot{a}_c(\theta_1 - \theta_1^0)\right], \tag{193}$$

$$\beta_\phi(b|c, \theta_2) = \mathrm{softmax}_b \left[d_{cb} + \dot{d}_{cb}(\theta_2 - \theta_2^0)\right]. \tag{194}$$

All logits and logit-derivatives are trainable variables. At the expansion point, let $\alpha_c = \alpha(c|\theta_1^0)$, $\dot{\alpha}_c = \partial_{\theta_1}\alpha(c|\theta_1)|_{\theta_1^0}$, $\beta_{bc} = \beta(b|c, \theta_2^0)$, and $\dot{\beta}_{bc} = \partial_{\theta_2}\beta(b|c, \theta_2)|_{\theta_2^0}$. The local module FIs are

$$F_{ac} = \sum_c \frac{\dot{\alpha}_c^2}{\alpha_c}, \quad F_{cb} = \sum_c \alpha_c \sum_b \frac{\dot{\beta}_{bc}^2}{\beta_{bc}}. \tag{195}$$

The endpoint marginal and its derivatives are

$$p_b = \sum_c \alpha_c \beta_{bc}, \quad \partial_1 p_b = \sum_c \dot{\alpha}_c \beta_{bc}, \quad \partial_2 p_b = \sum_c \alpha_c \dot{\beta}_{bc}. \tag{196}$$

Thus the endpoint Fisher matrix is

$$[F_B]_{ij} = \sum_b \frac{(\partial_i p_b)(\partial_j p_b)}{p_b}, \quad i, j \in \{1, 2\}. \tag{197}$$

The effective endpoint FI for the sum direction $u = (1, 1)^T$ is

$$F_B^{(u)} = \left(u^T F_B^{-1} u\right)^{-1}, \tag{198}$$

with the Moore–Penrose inverse used in singular cases. The adversary attempts to maximize

$$\Gamma_{\mathrm{adv}} = \frac{F_B^{(u)}}{\left(F_{ac}^{-1} + F_{cb}^{-1}\right)^{-1}}. \tag{199}$$

The causal-path CFII proves that every modular adversary must satisfy $\Gamma_{\mathrm{adv}} \leq 1$. Numerically, we set $L = M = 5$ and ran 36 independent Adam optimizations [16]. The largest value obtained was

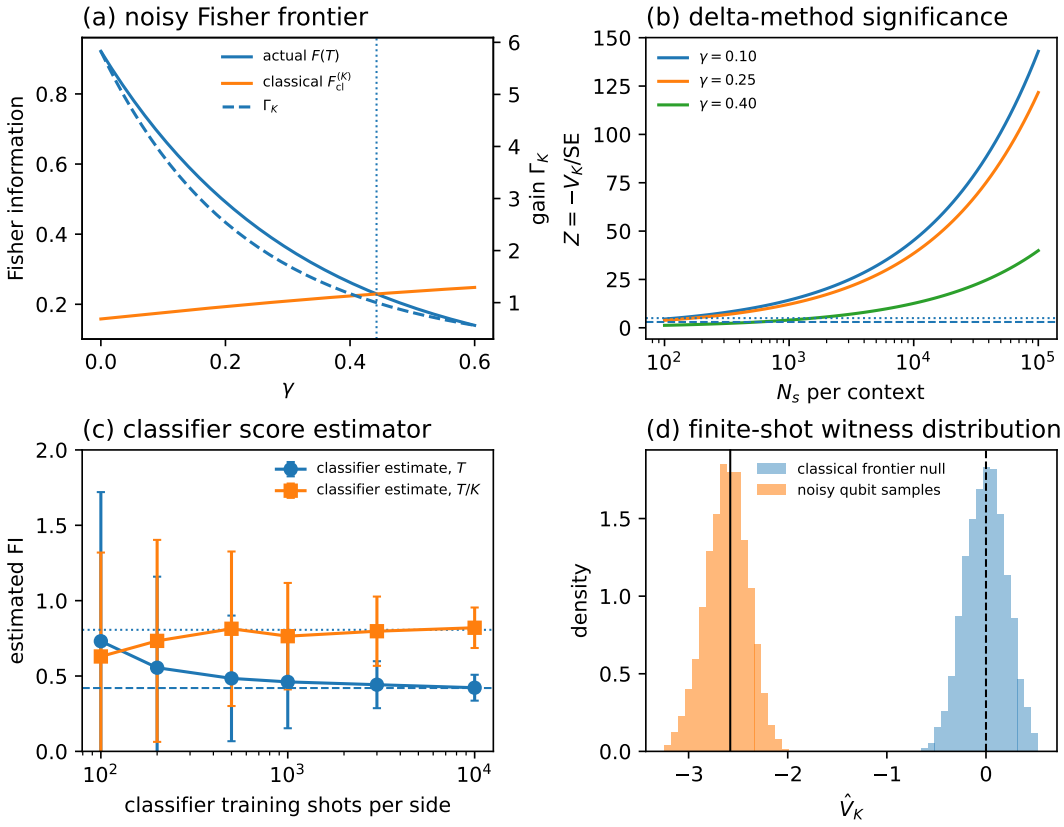
$$\max_{\text{restarts}} \Gamma_{\mathrm{adv}} = 0.9999999998, \tag{200}$$

with mean 0.99988 and minimum 0.99869. Thus the adversary learns to saturate the causal frontier, but it does not enter the forbidden region where the noisy coherent data lie.

E. Numerical results and interpretation

Fig. 4 gives the detailed numerical checks supporting the main-text stress-test figure. Panel (a) shows that increasing γ reduces the endpoint FI faster than the segment FI, eventually closing the advantage. Nevertheless, a finite noise window remains where $F(T) > F(T/K)/K$. Panel (b) shows that this region can be certified with modest sample sizes. At $\gamma = 0.40$, the violation is weaker but becomes a multi-sigma effect once N_s is increased to the 10^3–10^4 range. Panel (c) verifies that the classifier score estimator converges to the analytic FI values as its training sample size grows. Panel (d) displays the finite-shot separation between the noisy coherent witness and the classical frontier null.

The stress test should be read in the following way. The noisy coherent data are not idealized: visibility loss, readout error, finite samples, and likelihood-free score estimation are included. The classical comparator is not weak: it is a high-capacity differentiable modular path with latent mediators and optimized local kernels. Yet the two objects occupy different sides of the same CFII frontier. This is precisely the desired robustness statement for a PRL-scale claim: the negative witness is not an artifact of an analytic toy model, nor is it rescued by a flexible hidden-variable path as long as the defining modular causal assumptions remain in force.



Supplementary Fig. 4. Details of the AI-assisted finite-data stress test. (a) Actual noisy endpoint FI $F(T)$, equal-segment classical benchmark $F_{\rm cl}^{(K)} = F(T/K)/K$, and gain ratio Γ_K as functions of dephasing rate γ. The vertical dotted line marks the $\Gamma_K = 1$ crossing. (b) Delta-method significance $Z = -V_K/\mathrm{SE}$ versus shots per context for representative dephasing rates; horizontal lines mark $Z = 3$ and $Z = 5$. (c) Classifier score-estimator calibration at $\gamma = 0.25$. Dashed and dotted horizontal lines indicate the analytic FIs for T and T/K, respectively. (d) Finite-shot distribution of $\widehat{V}_K$ for the noisy coherent data at $\gamma = 0.25$ and $N_s = 10^3$, compared with a classical frontier null centered at zero.

SUPPLEMENTARY NOTE VIII. DISCUSSION, OUTLOOK, AND CONCLUSION

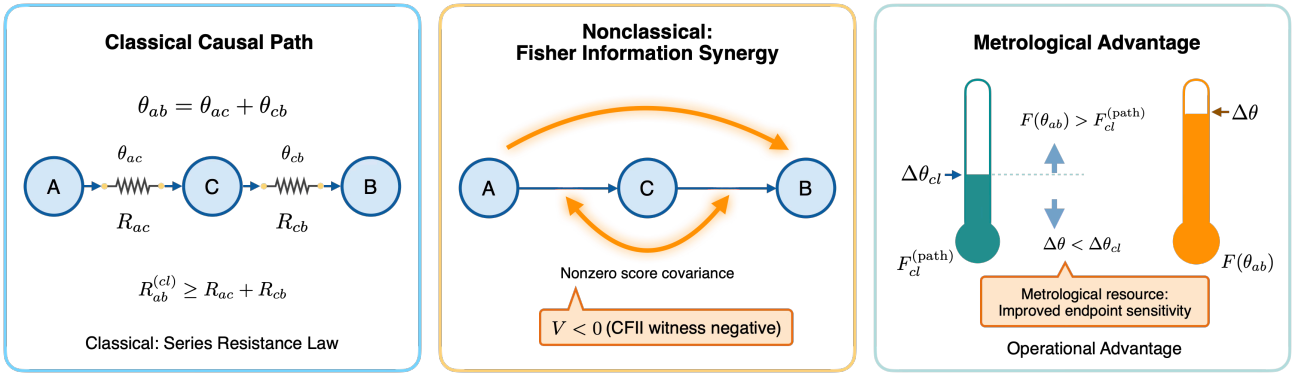
This work began from a deceptively simple observation: Fisher information is not merely a figure of merit for parameter estimation. It is also a geometric object encoding how probability models deform under infinitesimal parameter changes. Once a probability model is constrained by a classical causal structure, its admissible deformations are constrained as well. The resulting constraints are Fisher-information inequalities of a fundamentally different kind: they are causal Fisher-information inequalities (CFIIs). Their violation is therefore not a numerical surprise, but a logical collapse in the underlying classical causal narrative.

Figure 5 is intended as a compressed map of the entire witness-to-resource program developed in this SI: a classical causal assumption fixes a precision frontier; a CFII violation certifies that the assumed modular causal mediation is impossible; and the same negative witness directly quantifies a metrological gain. The schematic also makes explicit why the inverse Fisher information is the natural quantity to read as an information resistance, because it is the object that adds along classical causal paths and whose failure to add signals the breakdown of classical mediation.

A. Discussions and remarks

From Fisher-information inequalities to causal-model criteria

The central conceptual step of this work is to elevate Fisher-information inequalities from a metrological statement to a causal statement. In Sec. II–III, we showed how a classical causal model class (specified by a Bayesian network, conditional independences, and modular parameterizations) implies the nontrivial inequalities among Fisher informations associated with



Supplementary Fig. 5. Schematic summary of the logical flow of the paper. Left: under a classical causal-path description, the total parameter $\theta_{ab} = \theta_{ac} + \theta_{cb}$ must be mediated through an intermediate classical node C, so inverse Fisher information plays the role of an information resistance $R := F^{-1}$ and obeys the series law $R_{ab}^{(\mathrm{cl})} \geq R_{ac} + R_{cb}$. Center: coherent or otherwise nonclassical dynamics generate Fisher-information synergy (nonzero score covariance between the would-be modules), so the modular classical decomposition fails and the CFII witness becomes negative, $V < 0$. Right: the same violation is operationally a metrological resource, because it implies $F(\theta_{ab}) > F_{\mathrm{cl}}^{(\mathrm{path})}$ and therefore $\Delta\theta < \Delta\theta_{\mathrm{cl}}$. For longer decompositions, the classical frontier suffers the $1/K$ series dilution, whereas engineered synergy can maintain an $O(1)$ endpoint sensitivity.

the different parameter segments or experimental conditions. In Sec. IV–V, we established a sharpened interpretation: A CFII is a necessary condition for the existence of the assumed classical causal model. Therefore, if the CFII is violated, the appropriate conclusion is "the assumed classical causal model class is impossible for the observed statistics." This is what we call causal-model collapse.

At the same time, the Fisher-information form of these constraints makes the collapse immediately operational. A Fisher information inequality is an inequality about precision frontiers. Thus, when a classical causal model collapses, it collapses in a way that can be quantified as a precision gap: there exists a benchmark error floor enforced by the classical model class, and the experiment achieves a strictly smaller error. This witness-resource equivalence is formalized in **Theorem S15** and operationalized in the adversarial-benchmark example in Sec. VI C.

The physical meaning of CFII violation: breaking the series law of classical mediation

The causal-path CFII, expressed in the "series" form, i.e., as highlighted in Fig. 5,

$$F(\theta_{ab})^{-1} \geq F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}, \quad (\theta_{ab} = \theta_{ac} + \theta_{cb}), \tag{201}$$

is the prototypical example of the general framework. Its physical meaning is vivid. A classical causal-path explanation asserts that information about the total parameter θ_{ab} must traverse an intermediate classical mediator C. That assertion induces a modular decomposition into two independent estimation segments. In such a modularization, F^{-1} behaves as an information resistance, and the resistances in series must add. The inequality Eq. (201) is therefore the unavoidable accounting rule of classical mediation.

The example in Sec. VI shows why coherent quantum dynamics violates this rule. In the single-qubit coherent setting, the parameter is imprinted in a way that is not compatible with a classical intermediate mediator that would render the two segments independent in the sense required by the model. When the experiment violates Eq. (201), it is not that the data misbehave; it is that the classical causal-path story cannot be sustained without contradicting the observed local information geometry. In this sense, the CFII violation can be a diagnostic of causal nonclassicality: the incompatibility is not merely with a classical state space, but with a classical causal modularization.

Relation to earlier trajectory-testing Fisher-information inequalities

Historically, Fisher-information inequalities of the form in Eq. (201) were motivated as the tests of discrete evolution trajectories and intermediate-state hypotheses. That viewpoint already contained the core intuition: inserting the intermediate states amounts to imposing a conditional-independence-like structure on the statistical description of dynamics. Our framework clarifies and generalizes this intuition in two decisive ways.

First, it disentangles the logical content of the inequality from the interpretational language of "trajectory." The inequality is not intrinsically about whether the quantum state "really passes through" intermediate states. It is about whether the observed family of probability distributions is compatible with a classical causal model class whose defining feature is a modular, conditionally independent causal mediation. The trajectories constitute one special narrative within that class, but the causal logic is more general.

Second, the framework is systematic. Rather than proposing a single inequality for a single hypothesized structure, we provide a general route from a chosen causal model class (DAG plus independence/modularity assumptions) to a family of testable Fisher-information inequalities. This shifts the task from ad hoc witness construction to principled causal-model engineering.

Practical implications: designing metrology by causal-model engineering

A distinctive practical implication of the present framework is that it suggests a design principle for quantum sensing that is not limited to "use entanglement" or "optimize measurements." One may instead proceed as follows:

- First, identify a classical causal model class that captures the desired operational notion of classicality in the target platform (for instance, a trajectory-like mediation, a context-free constraint, or a Markovian modular decomposition).
- Second, derive the corresponding CFII constraints.
- Third, design probe states, dynamics, and readout schemes that maximize the violation of those constraints while respecting experimental limitations.
- Finally, turn the violation into a resource by explicitly constructing estimators and benchmarks, as done in Sec. VI via the V–G equivalence and the adversarial split-optimised benchmark.

The crucial point is that this design principle is model-relative and therefore operationally meaningful: it does not claim an absolute advantage over all conceivable strategies, but a certified advantage over an entire classical causal narrative.

Limitations and methodological remarks

Several caveats are also worth emphasizing.

- First, Fisher information is a local quantity. Our inequalities are therefore statements about local estimation geometry around a parameter point, and their operational meaning is clearest in asymptotic, regular regimes where maximum-likelihood estimators saturate Cramér–Rao bounds.
- Second, a failure to observe a violation does not prove the classical causal model. It only establishes consistency. This is a fundamental asymmetry of model falsification.
- Third, finite-sample implementations require careful statistical treatment. Estimating Fisher information from data introduces bias and variance, and one must propagate uncertainty to obtain confidence intervals for CFII violation. The Monte-Carlo achievability demonstrations in Sec. VI are a simple operational check, and developing a full hypothesis-testing toolkit for CFII violation is an important direction.
- Finally, the causal assumptions are explicit and therefore falsifiable, but also context-dependent. The physical meaning of a causal-model collapse depends on the chosen model class. This is a feature rather than a bug: it forces clarity about what is meant by "classical explanation" in each platform.

B. Outlook

The present work opens several concrete extensions.

A first direction is algorithmic. Given a DAG with latent variables and a specified parameter modularization, can one systematically enumerate all independent CFIIs, analogously to how entropic inequalities can be algorithmically generated for graphical models. Such a program would turn CFII derivation into a push-button tool for causal metrology.

A second direction is multiparameter estimation. The Fisher information matrix carries richer geometric content than the scalar Fisher information. Generalizing CFIIs to matrix inequalities may reveal new causal constraints and new resource regimes, especially in scenarios where the tradeoffs among parameters are central.

A third direction concerns noise and open dynamics. Realistic sensors operate under decoherence, and the relevant "classical" causal narratives may involve additional latent environment nodes or non-Markovian memory. Embedding such effects into causal graphs and deriving corresponding CFIIs could provide a new route to diagnosing non-Markovianity and to harvesting memory effects as metrological resources.

A fourth direction is to interface with broader causal notions in quantum theory. The classical causal models used here are intentionally conservative; one may ask how CFII-like statements extend to quantum causal models, process matrices, or generalized intervention frameworks [13, 14, 17]. This would sharpen the boundary between classical and quantum causality in information-geometric terms.

C. Conclusion

We have presented a framework in which Fisher-information inequalities become causal-model criteria. For a chosen classical causal model class, the framework yields testable causal Fisher-information inequalities (CFIIs). Their violation has a dual meaning: it falsifies the classical causal model class and simultaneously certifies a metrological resource, in the concrete sense of a precision gap relative to the best performance compatible with that class.

The examples show that this causal reading is not abstract. Even minimal coherent dynamics can cross the classical causal-path frontier, and the crossing can be made robust against adversarial classical benchmarks. The coQM scheme, discussed in the appendix, fits naturally into this narrative: NSIT is a classical causal constraint, its violation is a causal-model collapse, and the collapsed regime can be harvested as a Fisher-information resource.

As summarized schematically in Fig. 5, the framework yields a single causal-to-operational pipeline:

$$\text{classical causal hypothesis} \implies \text{CFII frontier} \implies \text{violation / falsification} \implies \text{certified metrological gain.} \tag{202}$$

In short, CFIIs unify three themes that are often treated separately: causal inference, nonclassicality, and precision measurement. They provide a principled route to reinterpreting the nonclassical advantage as the operational signature of classical causal-model impossibility, and to designing metrological protocols that deliberately exploit that impossibility.

SUPPLEMENTARY APPENDIX A. CONTEXTUAL QUANTUM METROLOGY AS AN NSIT-BASED CAUSAL-MODEL COLLAPSE AND ITS SCOPE

The main text develops a general principle: Fisher-information (FI) constraints become causal constraints once one commits to a classical causal model class, and the violation of such constraints simultaneously (i) falsifies that model class and (ii) certifies a metrological resource. The contextual quantum metrology (coQM) scheme, recently studied in Ref. [18], provides a complementary success story in which the "classicality" assumption is not the existence of an intermediate causal bottleneck, but a context-free causal description encoded by the condition of no-signaling in time (NSIT). This appendix makes the connection, in a form compatible with the causal language of our work.

The discussion proceeds in the following steps. At first, we introduce the NSIT as a conditional-independence (CI) constraint induced by a classical causal model. We then show that coQM reduces to conventional metrology when NSIT holds, hence no enhancement is possible in the context-free regime. Finally, we prove that an NSIT-based framework, by itself, cannot subsume the general CFII program developed in our work.

1. Preliminaries: two contexts and the NSIT constraint

We consider a parameter estimation problem with a single real parameter θ. A probe state $\hat{\rho}(\theta)$ is prepared, and the experimenter can choose between two "contexts" as below:

$$\begin{aligned} S = 0 &: \text{perform measurement } \hat{B} \text{ only,} \\ S = 1 &: \text{perform measurement } \hat{A} \text{ and then } \hat{B}. \end{aligned} \tag{A1}$$

Let $b \in \mathcal{B}$ denote the outcome of $\hat{B}$, and let $a \in \mathcal{A}$ denote the outcome of $\hat{A}$. The observable data are:

$$p(b|S=0,\theta) \equiv p(b|B,\theta), \tag{A2}$$

and

$$p(a,b|S=1,\theta) \equiv p(a,b|A,B,\theta), \quad p(b|S=1,\theta) = \sum_a p(a,b|S=1,\theta). \tag{A3}$$

The NSIT condition [19] is the requirement that the marginal statistics of B are invariant under the context of whether A is performed beforehand:

$$p(b|S=0,\theta) = p(b|S=1,\theta) \quad (\forall\, b, \theta.) \tag{A4}$$

Note that Eq. (A4) is a constraint directly testable from the measured frequencies. Conceptually, it expresses the absence of a causal influence of the "selection of A" on the marginal prediction of B.

2. NSIT as a classical causal-model constraint

We now formulate NSIT as a CI constraint, and thereby, as a causal-model class in the sense of our work.

Proposition S1 (NSIT is a conditional-independence constraint)**.** *Fix θ and consider the joint distribution of (S, B) defined by the operational conditional distributions $p(b|S,\theta)$ and an exogenous distribution $p(s)$ of the context choice. Then, the NSIT condition in Eq. (A4) holds for all b if and only if*

$$B \perp S \,|\, \theta. \tag{A5}$$

Proof. —Assume NSIT holds. Then $p(b|S=0,\theta) = p(b|S=1,\theta)$ for all b, hence $p(b|s,\theta)$ is independent of s. Therefore

$$p(b,s|\theta) = p(s)p(b|s,\theta) = p(s)p(b|\theta), \tag{A6}$$

which is exactly $B \perp S|\theta$. Conversely, if $B \perp S|\theta$, then $p(b|s,\theta) = p(b|\theta)$ for all s, in particular for $s=0$ and $s=1$. This is NSIT. □

Proposition S1 shows that NSIT is not a specifically quantum notion; it is the operational face of a CI constraint. In the causal language of Sec. II–IV, it corresponds to a model class $\mathcal{M}_{\mathrm{NSIT}}$ whose defining feature is that the exogenous context variable S has no causal arrow into B.

A standard latent-variable embedding makes this explicit. Let Λ denote a latent variable (a "classical state" prior to measurement), distributed as $p(\lambda|\theta)$. A minimal noncontextual (context-free) causal model class $\mathcal{M}_{\mathrm{NSIT}}$ can be defined by the DAG constraints

$$S \to A, \quad \Lambda \to A, \quad \Lambda \to B, \quad S \not\to B, \tag{A7}$$

together with the factorization

$$p(a,b,\lambda|s,\theta) = p(\lambda|\theta)\, p(a|s,\lambda,\theta)\, p(b|\lambda,\theta). \tag{A8}$$

By marginalizing over λ, we yield

$$p(b|s,\theta) = \sum_{\lambda} p(\lambda|\theta)p(b|\lambda,\theta), \tag{A9}$$

which is manifestly independent of s and hence implies NSIT. Thus, NSIT is a necessary consequence of the classical causal story in Eq. (A7) and Eq. (A8). Within the viewpoint of our work, one may take Eq. (A4) itself as the defining operational signature of $\mathcal{M}_{\mathrm{NSIT}}$.

We can now state that:

Theorem S16 (NSIT violation implies impossibility of the context-free causal model)**.** *If there exist θ and b such that $p(b|S=0,\theta) \neq p(b|S=1,\theta)$, then no model in the class $\mathcal{M}_{\mathrm{NSIT}}$ defined by Eq. (A8) can reproduce the observed operational statistics.*

Proof. —Any model in $\mathcal{M}_{\mathrm{NSIT}}$ implies Eq. (A9), hence $p(b|s,\theta)$ must be independent of s for all (b,θ). This is exactly NSIT. Therefore, if NSIT is violated by the observed statistics, those statistics cannot arise from any member of $\mathcal{M}_{\mathrm{NSIT}}$. For more details, see Ref. [18]. □

Theorem S16 is the precise sense in which NSIT violation is a causal-model collapse: it falsifies the entire class of context-free classical causal explanations that forbid the arrow $S \to B$.

3. Operational quasiprobability and the reduction under NSIT

The central operational move in coQM is to integrate the two contexts into a single statistical model, the operational quasiprobability (OQ) [18, 20, 21]. The OQ is constructed as

$$w(a,b|\theta) = p(a,b|S=1,\theta) + \frac{1}{2}\big[p(b|S=0,\theta) - p(b|S=1,\theta)\big]. \tag{A10}$$

Note that OQ is constructed purely from observable probabilities. When NSIT holds, the correction term vanishes and OQ collapses to an ordinary joint probability.

We then have the following proposition:

Proposition S2 (Reduction of OQ under NSIT)**.** *If NSIT in Eq. (A4) holds, then*

$$w(a,b|\theta) = p(a,b|S=1,\theta) \quad \forall\, a,b,\theta. \tag{A11}$$

Conversely, if Eq. (A11) holds for all a, b, θ, then NSIT holds.

Proof. If NSIT holds, then $p(b|S=0,\theta) - p(b|S=1,\theta) = 0$ for all b, θ, and Eq. (A10) reduces to Eq. (A11). Conversely, summing Eq. (A11) over a yields

$$\sum_a w(a,b|\theta) = p(b|S=1,\theta). \tag{A12}$$

But summing Eq. (A10) over a gives

$$\sum_a w(a,b|\theta) = p(b|S=1,\theta) + \frac{1}{2}\big[p(b|S=0,\theta) - p(b|S=1,\theta)\big]. \tag{A13}$$

The two expressions imply $p(b|S=0,\theta) = p(b|S=1,\theta)$ for all b, θ, i.e., NSIT. □

4. From witness to resource: Fisher-information consequences

Proposition S2 already exposes the metrological logic. If NSIT holds, OQ does not introduce a new statistical structure; it merely reproduces the probability of the consecutive measurement. Thus, in the context-free causal regime, coQM cannot create an advantage. To make this statement quantitative, we now relate the OQ to Fisher information.

In Ref. [18], coQM defines the contextual Fisher information (coFI) by treating $w(a,b|\theta)$ as a statistical model in the parameter regime where $w(a,b|\theta) > 0$:

$$F_{\rm co}(\theta) = \sum_{a,b} w(a,b|\theta)\left(\frac{\partial}{\partial\theta}\ln w(a,b|\theta)\right)^2. \tag{A14}$$

For comparison, let $F_q(\theta)$ denote the quantum Fisher information (QFI) of the probe family $\hat{\rho}(\theta)$, defined as the supremum of the classical FI over all POVMs on the probe.

Then we have the following theorem:

Theorem S17 (No advantage in the NSIT (context-free) regime)**.** *Assume NSIT holds and $w(a,b|\theta) > 0$. Then,*

$$F_{\rm co}(\theta) = F_{AB}(\theta) \le F_q(\theta), \tag{A15}$$

where $F_{AB}(\theta)$ is the classical FI of the consecutive measurement statistics $p(a,b|S=1,\theta)$. Consequently, when coQM is compared to a conventional strategy using the same total number of samples, no precision enhancement beyond the conventional QFI benchmark is possible in the NSIT regime.

Proof. —By **Proposition S2**, NSIT implies $w(a,b|\theta) = p(a,b|S=1,\theta)$. By substituting this into Eq. (A14), we obtain $F_{\rm co}(\theta) = F_{AB}(\theta)$. The inequality $F_{AB}(\theta) \le F_q(\theta)$ follows from the definition of QFI as the maximum FI achievable by any measurement on $\hat{\rho}(\theta)$. The consecutive measurement defines a valid measurement scheme on the probe and is therefore included in the maximization defining $F_q(\theta)$. Thus, Eq. (A15) holds. Finally, if a total of $2N_s$ probe uses are available, any conventional strategy satisfies the Cramér–Rao benchmark $\Delta\theta \ge 1/\sqrt{2N_s F_q(\theta)}$. In the NSIT regime, $F_{\rm co} \le F_q$ and the effective model underlying coQM reduces to the consecutive-measurement model; hence coQM cannot produce an error scaling that beats the conventional benchmark. □

Theorem S17 is the rigorous content behind the informal statement "if the causal model does not collapse, there is no resource to harvest." In the NSIT regime, the OQ carries no additional information beyond the ordinary probability model, and the Fisher information cannot exceed the QFI benchmark.

The decisive point is therefore the violation of NSIT. When $p(b|S=0,\theta) \neq p(b|S=1,\theta)$, the correction term in Eq. (A10) becomes nonzero. Then, w becomes a new effective statistical model that integrates two contexts. This is precisely where coQM can exhibit a precision enhancement.

5. NSIT does not subsume the CFII program

The NSIT condition is an extremely sharp and operationally meaningful constraint, but it constrains only one specific kind of conditional independence, namely the absence of a causal arrow from a context choice into a marginal prediction. In the coQM/NSIT setting, one introduces a context variable S (e.g., whether a preliminary measurement A is selected before measuring B) and imposes the constraint

$$p(b|S=0,\theta) = p(b|S=1,\theta) \quad (\forall\, b,\theta), \tag{A16}$$

which is nothing but the CI relation $B \perp S|\theta$ in a particular DAG. In other words, NSIT decides whether the marginal statistics of B are invariant under the context choice S.

Our CFII program addresses a different question. Given an arbitrary Bayesian network (DAG, possibly with latent variables), one extracts its CI relations and then derives Fisher-information inequalities implied by the entire causal structure. The causal-path CFII in our work is not a constraint about context dependence at fixed θ, but a constraint about modularity and additivity along a causal chain. In particular, the causal-path CFII

$$F(\theta_{ab})^{-1} \geq F(\theta_{ac})^{-1} + F(\theta_{cb})^{-1}, \quad (\theta_{ab} = \theta_{ac} + \theta_{cb}), \tag{A17}$$

encodes the series law of information that follows when a causal-path model enforces (i) a modular decomposition into two independent estimation modules and (ii) an additive parameter split. NSIT, by itself, does not impose either (i) or (ii), and therefore cannot be expected to reproduce Eq. (A17).

The separation can be made completely explicit. We now exhibit an operational model that is perfectly NSIT (hence fully compatible with the NSIT-based causal class $\mathcal{M}_{\mathrm{NSIT}}$), yet violates the causal-path CFII everywhere once one attempts to apply the causal-path modularity logic.

Proposition S3 (NSIT-compatible models can be maximally CFII-violating)**.** *There exist operational models for which the NSIT condition in Eq. (A16) holds identically for all θ (hence the model is compatible with $\mathcal{M}_{\mathrm{NSIT}}$), but for which the causal-path CFII in Eq. (A17) is violated for* every *nontrivial split $\theta_{ab} = \theta_{ac} + \theta_{cb}$ with $\theta_{ac} > 0$ and $\theta_{cb} > 0$. Therefore, an NSIT-only framework cannot subsume the causal-path CFII framework.*

Proof. —Consider the binary one-parameter family

$$p(0|\theta) = \cos^2\left(\tfrac{\theta}{2}\right), \quad p(1|\theta) = \sin^2\left(\tfrac{\theta}{2}\right). \tag{A18}$$

This family is regular and has constant (classical) Fisher information. Since $\partial_\theta p(0|\theta) = -\frac{1}{2}\sin\theta$ and $p(0|\theta)\big(1-p(0|\theta)\big) = \frac{1}{4}\sin^2\theta$, the Fisher information

$$F(\theta) = \sum_{b=0,1} p(b|\theta)\left(\frac{\partial}{\partial\theta}\log p(b|\theta)\right)^2 = \frac{\big(\partial_\theta p(0|\theta)\big)^2}{p(0|\theta)\big(1-p(0|\theta)\big)} = 1 \quad (\forall\,\theta), \tag{A19}$$

where the apparent $0/0$ points at $\theta = 0, \pi, 2\pi, \ldots$ are removable and the continuous extension yields the same constant.

Step 1: enforce NSIT. Introduce a context label $S \in \{0,1\}$ and define the marginal model of B to be context-independent:

$$p(b|S=0,\theta) = p(b|S=1,\theta) = p(b|\theta) \quad (\forall\, b,\theta), \tag{A20}$$

with $p(b|\theta)$ given by Eq. (A18). Then, Eq. (A16) holds identically, i.e., the model is perfectly NSIT, and therefore, compatible with $\mathcal{M}_{\mathrm{NSIT}}$. Operationally, one may realize such an NSIT situation, for example, by letting the "optional" measurement A commute with (or even coincide with) B and ensuring that no parameter encoding occurs between them; then selecting A does not disturb the marginal prediction of B.

Step 2: apply the causal-path CFII to an additive split. Now consider an additive decomposition $\theta_{ab} = \theta_{ac} + \theta_{cb}$ with $\theta_{ac}, \theta_{cb} > 0$ and evaluate the CFII gap

$$V(\theta_{ac},\theta_{cb}) := F(\theta_{ab})^{-1} - F(\theta_{ac})^{-1} - F(\theta_{cb})^{-1}. \tag{A21}$$

Using Eq. (A19), we obtain

$$V(\theta_{ac}, \theta_{cb}) = 1 - 1 - 1 = -1 < 0, \tag{A22}$$

for every nontrivial split. Hence, the causal-path CFII in Eq. (A17) is violated everywhere. This proves the claim. □

Proposition S3 sharpens the scope statement in a causal-model language. NSIT constrains a single arrow–prohibition ($S \not\to B$) and is silent about whether a process admits a modular causal mediation $A \to C \to B$ with an additive parameter split. The causal-path CFII, by contrast, is precisely the quantitative footprint of that modular mediation: it is derived from the path factorization and from the causal data-processing law (Sec. III). Therefore, an NSIT-based viewpoint cannot be a universal "generalization principle" for the CFII program.

This also clarifies the narrative hierarchy. coQM (and NSIT) fits naturally as one special instance of our broader causal perspective: NSIT is one CI constraint, its violation is one type of causal-model collapse, and coQM turns that collapse into a Fisher-information gain, exactly in the sense of the witness-to-resource transition of Sec. V. Our CFII framework, however, is strictly broader: it provides a systematic route from arbitrary classical causal model classes (DAGs + modularity) to testable Fisher-information inequalities and to resource interpretations of their violations.

[1] B. R. Frieden, *Science from Fisher Information: A Unification* (Cambridge University Press, 2004).
[2] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, 1946).
[3] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[4] S. L. Braunstein and C. M. Caves, Statistical distance and the geometry of quantum states, Phys. Rev. Lett. **72**, 3439 (1994).
[5] M. G. A. Paris, Quantum estimation for quantum technology, International Journal of Quantum Information **7**, 125 (2009).
[6] R. Zamir, A proof of the fisher information inequality via a data processing argument, IEEE Transactions on Information Theory **44**, 1246 (1998).
[7] K. C. Tan, V. Narasimhachar, and B. Regula, Fisher information from quantum resources, Phys. Rev. Lett. **127**, 200402 (2021).
[8] D. R. M. Arvidsson-Shukur, N. Yunger Halpern, and S. Lloyd, The value of temporal correlations in quantum metrology, Nature Communications **11**, 3775 (2020).
[9] N. B. Lupu-Gladstein, B. Y. Yilmaz, D. R. M. Arvidsson-Shukur, A. Brodutch, A. O. T. Pang, A. M. Steinberg, and N. Yunger Halpern, Negative quasiprobabilities enhance phase estimation in quantum metrology, Phys. Rev. Lett. **128**, 220504 (2022).
[10] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, Cambridge, 2009).
[11] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, Multi-parameter estimation beyond quantum fisher information, Journal of Physics A: Mathematical and Theoretical **53**, 363001 (2020).
[12] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, Quantum fisher information matrix and multiparameter estimation, Journal of Physics A: Mathematical and Theoretical **53**, 023001 (2020).
[13] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Quantum common causes and quantum causal models, Phys. Rev. X **7**, 031021 (2017).
[14] J. Barrett, R. Lorenz, and O. Oreshkov, Cyclic quantum causal models, Nature Communications **12**, 885 (2021).
[15] K. Cranmer, J. Brehmer, and G. Louppe, The frontier of simulation-based inference, Proceedings of the National Academy of Sciences **117**, 30055 (2020).
[16] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *International Conference on Learning Representations* (2015).
[17] L. A. Rozema, T. Strömberg, and P. Walther, Experimental aspects of indefinite causal order in quantum mechanics, Nature Reviews Physics **6**, 483 (2024).
[18] J. Jae, J. Lee, M. S. Kim, K.-G. Lee, and J. Lee, Contextual quantum metrology with operational quasiprobabilities, npj Quantum Information **10**, 68 (2024).
[19] J. J. Halliwell, The interpretation of no-signalling in time, Phys. Rev. A **96**, 012121 (2017).
[20] J. Ryu, J. Lim, S. Hong, and J. Lee, Operational quasiprobabilities for qudits, Phys. Rev. A **88**, 052123 (2013).
[21] J. Jae, J. Ryu, and J. Lee, Operational quasiprobabilities for continuous variables, Phys. Rev. A **96**, 042121 (2017).