

Interaction as Interference: A Quantum-Inspired Aggregation Approach^{*}

Pilsung Kang

Department of Software Science, Dankook University, Yongin, 16890, South Korea

Abstract

Classical approaches often treat interaction as engineered product terms or as emergent patterns in flexible models, offering little control over how synergy or antagonism arises. We take a quantum-inspired view: following the Born rule (probability as squared amplitude), *coherent* aggregation sums complex amplitudes before squaring, creating an interference cross-term, whereas an *incoherent* proxy sums squared magnitudes and removes it. In a minimal linear-amplitude model, this cross-term equals the standard potential-outcome interaction contrast Δ_{INT} in a 2×2 factorial design, giving relative phase a direct, mechanism-level control over synergy versus antagonism.

We instantiate this idea in a lightweight *Interference Kernel Classifier* (IKC) and introduce two diagnostics: *Coherent Gain* (log-likelihood gain of coherent over the incoherent proxy) and *Interference Information* (the induced Kullback–Leibler gap). A controlled phase sweep recovers the identity. On a high-interaction synthetic task (XOR), IKC outperforms strong baselines under paired, budget-matched comparisons; on real tabular data (*Adult* and *Bank Marketing*) it is competitive overall but typically trails the most capacity-rich baseline in paired differences. Holding learned parameters fixed, toggling aggregation from incoherent to coherent consistently improves negative log-likelihood, Brier score, and expected calibration error, with positive Coherent Gain on both datasets.

Keywords: Causal interaction, Quantum-inspired machine learning, Coherent

^{*}Preprint submitted to *Pattern Recognition*, November 2025.

Email address: pilsungk@dankook.ac.kr (Pilsung Kang)

1. Introduction

Interaction effects are of central importance across science and machine learning. They arise when the combined influence of multiple factors departs from simple additivity—producing synergy, where effects amplify each other, or antagonism, where they suppress one another. For example, combining two drugs can lead to markedly beneficial or adverse outcomes depending on their interaction. Yet frameworks for understanding interaction remain incomplete. In causal inference, interactions are commonly introduced as ad-hoc product terms in a regression, describing the phenomenon without explaining its generative mechanism [1, 2]. Likewise, classical information measures such as mutual information quantify statistical dependence but do not reveal how multiple inputs compose—e.g., they do not distinguish synergy from redundancy nor encode the sign of interaction. In contrast, a foundational principle of quantum theory, the Born rule, posits that observable probabilities are squared amplitudes [3]; this coherent (phase-sensitive) aggregation naturally produces interference cross-terms that can amplify or attenuate outcome probabilities. The gap between “interaction as a statistical add-on” and “interaction as a property of an aggregation rule” motivates our quantum-inspired perspective on causal interaction.

In this work, we propose that the origin of interaction lies not in an extra product term but in the rule by which causal influences are aggregated. We recast interaction as a consequence of coherent aggregation in the sense of the Born rule. To build intuition, think of two ripples on water: alignment produces amplification (synergy) and misalignment produces cancellation (antagonism). In our framework, each influence is represented by a complex amplitude ψ (a pre-probability score); observable probability is the squared magnitude $|\psi|^2$ per the Born rule. With two binary causes $A, B \in \{0, 1\}$, we write the cell probability $p_{ab} \equiv P(Y=1 \mid A=a, B=b)$ and define the interaction contrast

$$\Delta_{\text{INT}} \equiv p_{11} - p_{10} - p_{01} + p_{00}.$$

When these causes are randomized or well controlled, this contrast equals the potential-outcome interaction contrast; we use the p -based definition throughout for clarity.

To formalize aggregation, we distinguish two rules. The incoherent proxy adds energies (squared magnitudes), $p_{\text{inc}} = \sum_k |\psi_k|^2$. By contrast, coherent aggregation adds amplitudes before squaring, $p_{\text{coh}} = \left| \sum_k \psi_k \right|^2$, which introduces an interference cross-term. In a minimal amplitude-linear model for the 2×2 design,

$$\psi(a, b) = u_0 + a u_A + b u_B, \quad u_0, u_A, u_B \in \mathbb{C}, \quad p_{ab} = |\psi(a, b)|^2,$$

the interaction contrast reduces exactly to the cross-term

$$\Delta_{\text{INT}} = 2 \operatorname{Re}(u_A u_B^*),$$

making the relative phase between u_A and u_B a direct geometric control of synergy versus antagonism. (A short proof appears in Methods 3.) This identity establishes Δ_{INT} as a direct observable signature of the aggregation rule itself.

Our contributions are thus conceptual, methodological, and empirical. We formalize interaction as a property of a phase-sensitive aggregation rule, instantiate it in a practical classifier (the Interference Kernel Classifier, or IKC), and introduce two diagnostic metrics to quantify its effect: *Coherent Gain* (G_{coh}), the per-instance log-likelihood gain over an interaction-free incoherent proxy, and *Interference Information* (\mathcal{J}_{int}), the Kullback-Leibler (KL) divergence between the coherent and incoherent conditional distributions. Through a series of experiments—from controlled phase sweeps that validate the core identity to applications on high-interaction tasks (XOR) and real-world data (*Adult* and *Bank Marketing*) that demonstrate improved calibration and robustness—we validate our framework. Our approach is classical in implementation and makes no quantum-advantage claim; it provides a practical toolkit and a unifying lens for dissecting and engineering interaction in causal learning.

The remainder of this paper is organized as follows. Section 2 situates our work relative to prior research on interaction modeling. Section 3 develops our theoretical framework, formalizes the interference-interaction identity, and details the proposed model. Section 4 describes the protocols for our five empirical studies, and Section 5 presents their findings. Finally, we discuss the implications and limitations of our work

in Section 6 and conclude in Section 7.

2. Related Work

This section situates our contribution within three strands: (i) classical interaction modeling in statistics and machine learning (ML), (ii) causal definitions of interaction in the potential-outcome and structural frameworks, and (iii) quantum-inspired aggregation alongside specialized ML architectures that capture interactions. Prior approaches typically either enumerate interaction terms or recover them implicitly, offering limited mechanism-level control over synergy versus antagonism. In contrast, we frame interaction as an aggregation principle: a learnable phase governs synergy versus antagonism; the concrete mechanism and its connection to the potential-outcome interaction contrast Δ_{INT} are developed later in the paper.

2.1. Classical Interaction Modeling

Generalized Additive Models (GAMs) estimate nonlinear main effects additively, and in their basic form exclude interactions [4, 5]. Interactions can be added via tensor-product smooths or explicit products, but this requires pre-selecting orders and strong regularization, so apparent interaction strength can be an artifact of modeling choices. Post-hoc diagnostics such as the H-statistic quantify interaction strength in tree ensembles [6], yet they assess rather than control interactions. By contrast, our approach treats interaction as a property of the aggregation rule: coherent (phase-sensitive) addition yields an interference cross-term whose sign and magnitude are governed by a learned phase, providing mechanism-level control.

Tree ensembles, especially gradient boosting [7], capture high-order interactions via recursive partitioning and additive expansions. Yet these interactions are emergent, not tunable, components—offering limited mechanism-level control or interpretation compared to our phase parameter that directly sets the sign and strength of the interference term.

Hierarchical interaction selection imposes heredity constraints so interactions appear only with main effects (e.g., hierarchical lasso [8]); it still requires prespecifying

interaction order and primarily induces sparsity rather than a mechanism that distinguishes synergy from antagonism. Our approach instead encodes interaction in the aggregation rule itself, tying the interference cross-term directly to the potential-outcome contrast Δ_{INT} without enumerating pairs.

Rule ensembles make interactions explicit as sparse conjunctions extracted from trees, improving interpretability over black-box boosting, yet these terms are discovered indirectly from recursive splits and remain difficult to tune mechanistically [6]. In our framework, the coherent–incoherent toggle manipulates only the aggregation step (holding parameters fixed), providing mechanism-level control and diagnostics $\{G_{\text{coh}}, \mathcal{J}_{\text{int}}\}$ rather than post-hoc identification of interaction structure.

2.2. Causal Interaction in Potential-Outcome and SCM Frameworks

In the potential-outcome framework with two binary causes $A, B \in \{0, 1\}$ and potential outcomes $Y(a, b)$, the additive-scale interaction contrast Δ_{INT} quantifies synergy and antagonism as a contrast on cell probabilities (formal definition in Sec. 3.2). Under randomized or well-controlled assignment (no unmeasured confounding), this contrast is identifiable from observed conditional probabilities. Prior works [9, 10, 11] clarify interpretability on additive vs. multiplicative scales and provide conditions under which Δ_{INT} is identifiable.

From the structural causal model (SCM) viewpoint, interaction corresponds to non-additive composition in the structural equation for Y [1] and relates to sufficient-cause (synergy or antagonism) notions and to effect decompositions on additive or multiplicative scales. Our use of Δ_{INT} follows these definitions and aligns notation with the rest of the paper. Building on these definitions, we later show (Sec. 3.2) that in a minimal coherent aggregation model the interference cross-term equals Δ_{INT} exactly in a 2×2 design, making the relative phase a tunable mechanism for the sign and magnitude of interaction (instantiated in Sec. 3.3).

2.3. Quantum-Inspired Aggregation and Specialized Interaction Models

Quantum cognition has long used interference—adding amplitudes before squaring—to explain context and order effects in human judgment [12, 13]. These models typically target explanatory adequacy on behavioral tasks and do not connect to causal

interaction formalizations. By contrast, we cast interaction as an aggregation principle: coherent (phase-sensitive) addition yields an interference cross-term whose sign and magnitude are governed by a learned phase, and we link this term exactly to the potential-outcome contrast Δ_{INT} , delivering a practical classifier (IKC) with paired diagnostics $\{G_{\text{coh}}, \mathcal{J}_{\text{int}}\}$.

Specialized ML models also capture interactions but with different goals and mechanisms. Factorization Machines efficiently parameterize pairwise feature interactions in sparse settings [14]; Wide&Deep combines memorization and generalization for recommendation [15]; Neural FM and xDeepFM learn higher-order interactions via neural parameterizations [16, 17]; attention mechanisms learn data-dependent interaction weights across tokens [18]. These approaches excel at prediction yet offer limited mechanism-level control over whether interactions are synergistic or antagonistic: the effect emerges from fitted parameters rather than a tunable aggregation rule. In IKC, a phase parameter directly controls the interference sign, and a coherent–incoherent toggle (holding parameters fixed) isolates the contribution of phase-coherent aggregation from representation capacity. For evaluation, we also employ standard post-hoc calibration via temperature scaling (TS) [19], which complements but is orthogonal to the aggregation mechanism.

3. Methods

3.1. A Quantum-Inspired View of Causal Interaction

The central proposal of this work is to view interaction not as a special feature term, but as an emergent property of the aggregation rule. We contrast a classical, incoherent rule with a quantum-inspired, coherent rule. In ML terms, we consider two outcome *classes* $y \in \{0, 1\}$ (we use “channel” and “class” interchangeably) and define class-specific complex amplitudes whose squared magnitudes yield probabilities.

Notation. We use uppercase P for probabilities and E for unnormalized energies (squared amplitude magnitudes); Let $x \in \mathbb{R}^d$ denote the input feature vector. For each class $y \in \{0, 1\}$, let $\{a_{y,k}(x)\}_{k=1}^K \subset \mathbb{C}$ denote K internal amplitude components (class- y

"branches"), and define the class amplitude

$$\psi_y(x) = \sum_{k=1}^K a_{y,k}(x).$$

Write $\text{Re}(\cdot)$ for the real part and $(\cdot)^*$ for complex conjugation.

Incoherent vs. coherent class energies. The incoherent (square-then-sum) class energy removes all cross-terms:

$$E_y^{\text{inc}}(x) = \sum_{k=1}^K |a_{y,k}(x)|^2.$$

The coherent (add-then-square) class energy includes interference between components:

$$E_y^{\text{coh}}(x) = \left| \sum_{k=1}^K a_{y,k}(x) \right|^2 = \sum_{k=1}^K |a_{y,k}(x)|^2 + \sum_{1 \leq i < j \leq K} 2 \text{Re}(a_{y,i}(x) a_{y,j}(x)^*).$$

From energies to probabilities. Binary class probabilities are obtained by normalizing class energies:

$$P_{\text{inc}}(Y=1 | x) = \frac{E_1^{\text{inc}}(x)}{E_0^{\text{inc}}(x) + E_1^{\text{inc}}(x)}, \quad P_{\text{coh}}(Y=1 | x) = \frac{E_1^{\text{coh}}(x)}{E_0^{\text{coh}}(x) + E_1^{\text{coh}}(x)}.$$

Thus the only difference between the two readouts is whether within-class cross-terms are present (E^{coh}) or removed (E^{inc}).

A minimal amplitude-linear parameterization. For a practical instantiation we use a class-specific affine complex map

$$\psi_y(x) = b_y + \sum_{j=1}^d x_j w_{y,j}, \quad w_{y,j}, b_y \in \mathbb{C}.$$

This corresponds to $K=1$ (a single amplitude per class). Then the coherent class energy expands as

$$E_y^{\text{coh}}(x) = |\psi_y(x)|^2 = |b_y|^2 + 2 \text{Re}\left(b_y \sum_j x_j w_{y,j}^*\right) + \sum_j |w_{y,j}|^2 x_j^2 + 2 \text{Re}\left(\sum_{i < j} w_{y,i} w_{y,j}^* x_i x_j\right),$$

whereas the incoherent proxy drops all phase-sensitive cross-terms:

$$E_y^{\text{inc}}(x) \equiv |b_y|^2 + \sum_j |w_{y,j}|^2 x_j^2.$$

(For one-hot features, $x_j^2 = x_j$, so the removal of cross-terms is explicit.)

The two $2 \operatorname{Re}(\cdot)$ cross-terms absent from E^{inc} are the interference terms that give rise to interaction effects in coherent aggregation. In the general d -dimensional case, these encode all pairwise feature interactions (bias-weight and weight-weight). The next subsection establishes the precise connection to causal interaction for a minimal 2×2 design.

3.2. The Interference–Interaction Identity

We connect the classical potential-outcome interaction contrast to the interference cross-term induced by coherent aggregation.

Definition (Interaction contrast). Let $(a, b) \in \{0, 1\}^2$ denote the two-factor design and $p_{ab} \in [0, 1]$ the cell probability $p_{ab} = P(Y=1 \mid a, b)$. Define

$$\Delta_{\text{INT}} \equiv p_{11} - p_{10} - p_{01} + p_{00}.$$

Setup. We consider a minimal amplitude–linear model with a single complex amplitude whose squared magnitude yields the Bernoulli success probability (one channel):

$$\psi(a, b) = u_0 + a u_A + b u_B, \quad u_0, u_A, u_B \in \mathbb{C}.$$

Under the Born rule, the corresponding cell probability is $p_{ab} = |\psi(a, b)|^2$. We choose magnitudes so that $0 \leq p_{ab} \leq 1$ ensures a valid probability; the identity below, which links the interaction contrast to the interference cross-term, is algebraic and does not depend on normalization across outcomes. This one–channel formulation suffices to establish the identity; the practical binary classifier in Sec. 3.3 uses two outcome channels $y \in \{0, 1\}$.

Proposition (Interference–interaction identity). With the above definitions,

$$\Delta_{\text{INT}} = 2 \operatorname{Re}(u_A u_B^*),$$

which depends only on the pair (u_A, u_B) .

Proof. Expand each cell:

$$\begin{aligned}
p_{11} &= |u_0 + u_A + u_B|^2 = |u_0|^2 + |u_A|^2 + |u_B|^2 + 2 \operatorname{Re}(u_0 u_A^* + u_0 u_B^* + u_A u_B^*), \\
p_{10} &= |u_0 + u_A|^2 = |u_0|^2 + |u_A|^2 + 2 \operatorname{Re}(u_0 u_A^*), \\
p_{01} &= |u_0 + u_B|^2 = |u_0|^2 + |u_B|^2 + 2 \operatorname{Re}(u_0 u_B^*), \\
p_{00} &= |u_0|^2.
\end{aligned}$$

Taking the alternating sum $p_{11} - p_{10} - p_{01} + p_{00}$ cancels all terms except the cross-term $2 \operatorname{Re}(u_A u_B^*)$. \square

Consequences and interpretation. Writing $u_k = r_k e^{i\phi_k}$ (with $r_k \geq 0$, $\phi_k \in \mathbb{R}$) yields

$$\Delta_{\text{INT}} = 2 r_A r_B \cos(\phi_B - \phi_A),$$

so (i) the sign of interaction is a geometric property of the relative phase $\phi_B - \phi_A$ (synergy > 0 , antagonism < 0); (ii) the magnitude is bounded by $2 r_A r_B$; and (iii) the result is invariant to global phase shifts $u_k \mapsto e^{i\theta} u_k$ and to the background amplitude u_0 . The identity holds algebraically for each cell ($A = a, B = b$).

3.3. The Interference Kernel Classifier (IKC) Model

Instantiation. We implement phase-coherent aggregation as a practical binary classifier (IKC). In the notation of Sec. 3.1, where $\psi_y = \sum_k a_{y,k}$, we instantiate the class (output channel) amplitude as an affine complex map

$$\psi_y(x) = w_y^\top x + b_y, \quad w_y \in \mathbb{C}^d, \quad b_y \in \mathbb{C}, \quad y \in \{0, 1\},$$

with $x \in \mathbb{R}^d$ the preprocessed feature vector. In implementation, we represent $\psi_y(x) = \alpha_y(x) + i\beta_y(x)$ using two real-valued affine maps per class; the relative phases among coefficients determine the sign and magnitude of the within-class interference terms.

Born-aggregated probability. The IKC predictive distribution is given by the coherent probability $P_{\text{IKC}}(Y=1 \mid x) = P_{\text{coh}}(Y=1 \mid x)$ defined above, where $E_y^{\text{coh}}(x) = |\psi_y(x)|^2 = \alpha_y(x)^2 + \beta_y(x)^2$. This map is invariant to a common nonzero complex rescaling, i.e., $\psi_y \mapsto c \psi_y$ for any $c \in \mathbb{C} \setminus \{0\}$ leaves P_{IKC} unchanged.

Training objective and regularization. Parameters $\{w_y, b_y\}$ are learned by minimizing the negative log-likelihood (NLL) on the training split using the Born probabilities.

For numerical stability we clip probabilities to $[10^{-7}, 1 - 10^{-7}]$. We apply ℓ_2 weight decay (i.e., an L_2 regularization term on parameter magnitudes) to the real-valued parameterization (α, β) , which helps stabilize the global scale non-identifiability noted above. Optimization uses first-order methods (e.g., Adam [20]); the optimizer and its hyperparameters are selected by budget-matched randomized search.

Temperature scaling. We apply post-hoc TS as specified in the common protocol (Sec. 4): a single temperature is fit on the calibration split and used with the safety-switch policy described there.

Incoherent toggle for ablation (Sec. 4.4). Holding the learned parameters fixed, we isolate the aggregation rule by switching the readout to the incoherent proxy $P_{\text{inc}}(Y=1 \mid x)$ defined in Sec. 3.1, which removes all phase-sensitive cross-terms. This toggle changes only the aggregation rule (add-then-square \rightarrow square-then-sum) and leaves architecture, parameters, and splits unchanged.

Diagnostics. Alongside standard metrics (NLL, Brier, and ECE), we report two coherence diagnostics. Let $P_{\text{coh}}(\cdot \mid x)$ and $P_{\text{inc}}(\cdot \mid x)$ denote the binary class probabilities under coherent and incoherent aggregation, respectively, using identical learned parameters (the incoherent proxy removes within-class cross-terms only).

Coherent Gain quantifies the average per-instance log-likelihood improvement:

$$G_{\text{coh}} = \frac{1}{n} \sum_{i=1}^n [\log P_{\text{coh}}(y_i \mid x_i) - \log P_{\text{inc}}(y_i \mid x_i)].$$

Interference Information measures the distributional shift induced by coherence via the average KL divergence:

$$\mathcal{J}_{\text{int}} = \frac{1}{n} \sum_{i=1}^n \text{KL}(P_{\text{coh}}(\cdot \mid x_i) \parallel P_{\text{inc}}(\cdot \mid x_i)), \quad \text{with } \text{KL}(p \parallel q) = \sum_{y \in \{0,1\}} p_y \log \frac{p_y}{q_y}.$$

Both diagnostics are computed from raw (pre-TS) probabilities and quantify the *local* (within-representation) benefit of coherent aggregation: holding learned parameters fixed, $G_{\text{coh}} > 0$ indicates that phase-sensitive composition improves per-instance likelihood relative to an incoherent proxy, and $\mathcal{J}_{\text{int}} > 0$ measures the distributional shift induced by the interference cross-term. These diagnostics isolate the aggregation rule’s contribution from representational capacity, complementing global performance met-

rics (NLL, Brier, and ECE) that conflate both. Note that $\mathcal{J}_{\text{int}} \geq 0$ by the non-negativity of KL divergence.

4. Experimental Setup

We evaluate the IKC in five experiments that progress from theory to practice: **(E1)** identity validation on a minimal 2×2 synthetic design; **(E2)** a high-interaction XOR task compared against strong baselines; **(E3)** robustness to training-only label noise with a TS safety switch; **(E4)** an ablation that toggles coherent vs. incoherent aggregation while keeping learned parameters fixed; and **(E5)** real-world benchmarks on *Adult* and *Bank Marketing* with calibrated evaluation. Detailed protocols for each experiment follow.

Common protocol. Unless otherwise noted, all experiments follow this protocol:

- **Splits:** For each dataset and seed, we use fixed training, validation, calibration, and test splits (61.25%, 15%, 8.75%, and 15%).
- **Hyperparameter optimization:** We run budget-matched randomized HPO for each model and select the configuration with the lowest validation NLL.
- **Retraining:** The selected configuration is retrained on the union of the training and validation splits before final evaluation.
- **Temperature scaling:** A single temperature is fit on the calibration split and applied to test predictions only if it lowers calibration NLL (safety switch); otherwise, raw probabilities are retained.
- **Metrics:** We report NLL, Brier score, ECE (15 equal-width bins), and accuracy on the test split.
- **Statistical analysis:** Across seeds, we aggregate with means and two-sided bootstrap 95% CIs (5,000 replicates), and compute exact paired sign-flip p -values.

Deviations from common protocol. (E1) Follows its own simulation protocol (not using dataset splits nor TS). Table 1 summarizes the procedural differences across experiments.

Table 1: Summary of selection and evaluation differences across experiments. E2 uses $n_{\text{seeds}} = 10$ (smaller sample for rapid iteration on synthetic data); E3, E4, and E5 use $n_{\text{seeds}} = 20$.

Experiment	Baseline selection criterion	Retraining on training and validation	TS policy
E2 (XOR)	Best classical baseline chosen by test NLL on the fixed test split (within seed)	No (comparison uses validation-tuned models without retraining on training and validation)	Always applied (no safety switch; TS fit on calibration and applied to test)
E3 (Robustness)	Best classical baseline chosen by validation NLL (within seed and noise level)	Yes (selected configuration re-trained on training and validation before test)	Safety switch (apply TS to test only if it lowers calibration NLL)
E4 (Ablation)	N/A (coherent vs. incoherent toggle; no baseline comparison)	Yes (default protocol)	Safety switch independently for each mode
E5 (Real-world)	Anchor (best) baseline chosen by test NLL on the fixed test split (within seed)	Yes (default protocol: retraining on training and validation for final evaluation)	Safety switch (apply TS to test only if it lowers calibration NLL)

4.1. Protocol for Experiment 1 (E1): Phase Sweep Validation

Objective. Provide a direct, quantitative validation of the identity between the potential-outcome interaction contrast and the interference cross-term.

Theoretical model. We use the minimal amplitude-linear model (Sec. 3.2), where $\Delta_{\text{INT}} = 2r_A r_B \cos(\Delta\phi)$ with $\Delta\phi \equiv \phi_B - \phi_A$ the relative phase.

Simulation procedure. Fix $(r_0, r_A, r_B) = (0.20, 0.35, 0.35)$ and sweep the relative phase $\Delta\phi$ over 121 points in $[-\pi, \pi]$. For each $\Delta\phi$, simulate a balanced 2×2 design with $n = 2000$ Bernoulli trials per cell, estimate \widehat{p}_{ab} , and compute $\widehat{\Delta}_{\text{INT}}$. Without loss of generality set $\phi_A = 0$ and realize the sweep by $\phi_B = \Delta\phi$. (The magnitudes keep p_{ab} away from 0/1 and match the theoretical range $2r_A r_B = 0.245$.)

Statistical analysis. Two-sided 95% bootstrap CIs are formed with 5,000 replicates (seed 108).

4.2. Protocol for Experiment 2 (E2): Comparison with Strong Baselines

Objective. Compare the coherent (Born-aggregating) model against strong classical baselines on a canonical high-interaction task (XOR) to assess performance when the

signal resides purely in interaction.

Data-generating process. Inputs are binary and independent, $X_1, X_2 \sim \text{Bernoulli}(1/2)$, with the label $Y = X_1 \oplus X_2$. We generate $N = 20,000$ samples. A single, fixed test split of 3,000 samples (15%) is created using the first random seed and is reused for all models and seeds to enable paired comparisons. The remaining 17,000 samples are, for each seed, partitioned into training (61.25%), validation (15%), and calibration (8.75%).

Models and baselines. We compare IKC with the following baselines:

- **Logistic Regression with an Interaction Term.** Logistic regression on degree-2 polynomial features including the product $X_1 X_2$.
- **Quadratic Logistic Regression.** Logistic regression using the features X_1, X_2 , and their squares X_1^2, X_2^2 . For binary inputs (where $X_i^2 = X_i$), this reduces to a main-effects-only model without interaction, serving as a deliberate misspecification baseline.
- **XGBoost.** A strong tree-based baseline.

Training and hyperparameter search. For each of $n_{\text{seeds}} = 10$ seeds, we run randomized HPO with a fixed budget of twenty trials per model under the same search space, early-stopping policy, and seeding policy. The configuration with the lowest *validation* NLL is selected for each model (intra-model selection). For paired comparisons, we then identify the classical baseline with the lowest *test* NLL on the fixed test split (inter-model anchor selection). This design provides a conservative comparison against the empirically strongest baseline on the test set, effectively tilting the comparison against IKC, which may introduce upward bias in baseline performance due to post-hoc anchor selection. No model is retrained or further tuned using test data.

4.3. Protocol for Experiment 3 (E3): Robustness Evaluation

Objective. Evaluate robustness under controlled label corruption, focusing on how performance degrades as noise increases and whether IKC’s advantages, if any, persist.

Datasets and preprocessing. We use *Adult* (Census Income; UCI repository) to predict whether annual income exceeds \$50K and *Bank Marketing* (UCI) to predict

Table 2: Datasets and preprocessing used in E3 and E5.

Dataset	Target	Preprocessing
Adult	income (> 50K)	<ul style="list-style-type: none"> • Categorical: impute (mode) \rightarrow OneHotEncoder (handle_unknown=ignore) • Numeric: impute (median) \rightarrow standardize
Bank	y (= yes)	<ul style="list-style-type: none"> • Same as Adult • Drop duration • Treat “?” and “unknown” as missing

whether a client subscribes to a term deposit [21, 22]. Categorical features undergo mode imputation and one-hot encoding; numerical features undergo median imputation and standardization. For *Bank*, we exclude `duration` because it is observed only after the call and leaks target information, and we treat both “?” and “unknown” as missing [22]. Table 2 summarizes datasets and preprocessing. For each seed, we create fixed splits; within a seed, the same test split is reused across all noise levels to enable paired comparisons.

Noise injection. Corruption is applied only to the training split; validation, calibration, and test remain clean to avoid leakage. We inject symmetric label noise by flipping the training label independently with probability $p \in \{0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$.

Models. Same as E2.

Deviations from common protocol. The best baseline is selected by validation NLL within each (noise level, seed) pair. We use $n_{\text{seeds}} = 20$.

4.4. Protocol for Experiment 4 (E4): Ablation on Coherent vs. Incoherent Aggregation

Objective. Isolate the contribution of the aggregation rule by evaluating a single trained IKC model in two modes that differ only at the final aggregation step: (i) coherent (add amplitudes then square; cross terms present) and (ii) an incoherent proxy (square then sum; all cross terms removed). Holding all learned parameters fixed, we test whether the interference cross term improves likelihood and calibration.

Datasets and preprocessing. We use *Adult* and *Bank* with the E3 preprocessing

pipeline.

Ablation design. For each trained IKC model (parameterized by $\{w_y, b_y\}_{y \in \{0,1\}}$ with $\psi_y(x) = w_y^\top x + b_y$), we form two probability distributions using identical parameters:

Coherent mode: Uses $P_{\text{coh}}(Y=1 \mid x)$ with $E_y^{\text{coh}}(x) = |\psi_y(x)|^2$ (Born aggregation retains phase-sensitive cross-terms).

Incoherent mode: Uses $P_{\text{inc}}(Y=1 \mid x)$ with $E_y^{\text{inc}}(x) = |b_y|^2 + \sum_j |w_{y,j}|^2 x_j^2$ (cross-terms removed, including bias-feature and feature-feature interactions; for one-hot features, $x_j^2 = x_j$).

Both modes use the same Born normalization and parameters (see Sec. 3.3 for derivations).

Deviations from common protocol. TS is applied independently for each mode using the safety switch. Diagnostics $\{G_{\text{coh}}, \mathcal{J}_{\text{int}}\}$ are computed from raw (pre-TS) probabilities. Paired differences $\Delta_{\text{coh-inc}} \equiv m_{\text{coh}} - m_{\text{inc}}$ (where $m \in \{\text{NLL}, \text{Brier}, \text{ECE}\}$) are analyzed. We use $n_{\text{seeds}} = 20$.

4.5. Protocol for Experiment 5 (E5): Real-World Application

Objective. Assess whether the coherent (Born-aggregating) model provides practical benefit on real tabular problems, using two standard benchmarks (*Adult* and *Bank*). We emphasize fair training budgets, calibrated evaluation, and paired statistical comparisons against strong baselines.

Datasets and preprocessing. We use *Adult* and *Bank* with the preprocessing pipeline from E3 (see Table 2 for details).

Models and hyperparameter search. We use the same models as in E2 (Sec. 4.2): IKC, logistic regression with pairwise interactions, quadratic logistic regression, and XGBoost. Within each seed and dataset, every model receives the same randomized hyperparameter search budget (25 trials). The configuration with the lowest validation NLL is selected, and the model is retrained on the union of the training and validation splits before evaluation.

Metrics. Per seed, we anchor to the classical baseline with the lowest test NLL—a conservative choice against IKC—and compute paired differences (IKC–Best-Baseline)

for NLL, Brier, and ECE. For interpretability, we additionally compute $\{G_{\text{coh}}, \mathcal{J}_{\text{int}}\}$ from the IKC model’s raw (pre-TS) probabilities.

5. Experimental Results

All experiments were run in Python 3 with PyTorch [23], NumPy [24], SciPy [25], and Matplotlib [26] for numerical routines and visualization. Classical baselines used `scikit-learn` [27] and XGBoost [28]. Our IKC implementation follows the methods in Sec. 3.3; TS and sign-flip tests use the same stack.

5.1. E1: Phase Sweep Validation of the Interference–Interaction Identity

We validate the identity from Sec. 3.2 by sweeping the relative phase $\Delta\phi$ over 121 points (see Sec. 4.1 for parameters). The theoretical prediction is $\Delta_{\text{INT}} = 2 r_A r_B \cos(\Delta\phi) = 0.245 \cos(\Delta\phi)$. Figure 1 shows that empirical estimates $\widehat{\Delta}_{\text{INT}}$ closely track the theoretical curve. Representative checkpoints:

- $\Delta\phi \approx -\pi$: theory -0.245 , observed -0.229 $[-0.253, -0.205]$.
- $\Delta\phi \approx -\pi/2$: theory 0.000 , observed $+0.002$ $[-0.033, +0.037]$.
- $\Delta\phi \approx 0$: theory $+0.245$, observed $+0.233$ $[+0.198, +0.267]$.
- $\Delta\phi \approx +\pi/2$: theory 0.000 , observed $+0.029$ $[-0.004, +0.063]$.
- $\Delta\phi \approx +\pi$: theory -0.245 , observed -0.264 $[-0.288, -0.239]$.

At neutral points ($\pm\pi/2$), confidence intervals include zero; at extremes ($0, \pm\pi$), they exclude zero with the correct sign.

5.2. E2: Comparison with Strong Baselines

Main results. Paired, budget-matched comparisons (IKC – Best-Baseline) show that IKC significantly outperforms on XOR (Table 3; $n_{\text{seeds}}=10$): $\Delta\text{NLL} = -0.159 \pm 0.134$ (exact $p = 0.016$), $\Delta\text{Brier} = -0.062 \pm 0.057$, and $\Delta\text{ECE} = -0.034 \pm 0.023$ (exact $p \approx 0.121$). IKC yields clear likelihood and Brier improvements; the ECE reduction is consistent in magnitude but not statistically significant under the sign-flip test. Because

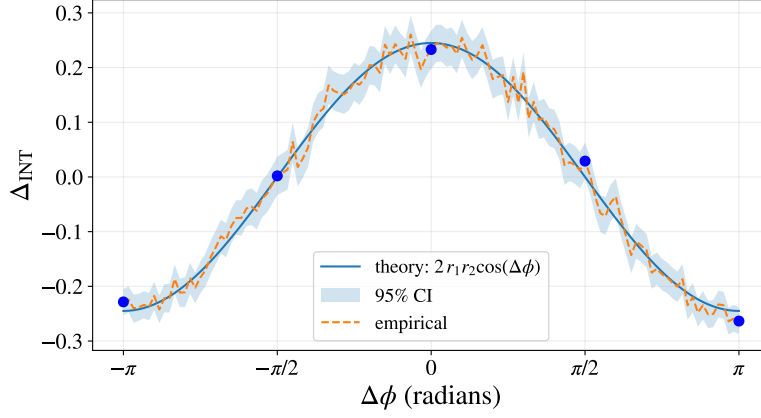


Figure 1: **E1: Phase sweep validates the interference–interaction identity.** Interaction contrast Δ_{INT} as a function of the relative phase $\Delta\phi$. The plot overlays the theoretical law $2r_Ar_B\cos(\Delta\phi)$ from the amplitude–linear model with $(r_0, r_A, r_B) = (0.20, 0.35, 0.35)$ and empirical plug-in estimates $\widehat{\Delta}_{\text{INT}}$ with two-sided 95% bootstrap confidence bands (balanced 2×2 design; $n=2000$ per cell; 121 phase points; 5,000 bootstrap replicates; a fixed random seed).

the anchor baseline is chosen post-hoc from $K=3$ candidates, the reported p -value may be optimistic. Even under a conservative Bonferroni correction (threshold: $0.05/3 \approx 0.017$), the result remains significant ($p=0.016$), though this adjustment may be overly stringent given the paired design on a fixed test split.

Interpretation. XOR is a pure-interaction task: main effects are zero, and only the interaction term is informative. IKC’s coherent aggregation captures this via the phase-sensitive cross-term in the amplitude-linear model, whereas baselines that omit the explicit interaction (e.g., *Quadratic*) are mis-specified; even when the interaction feature is present (logistic regression with an interaction term), IKC’s Born aggregation confers a likelihood advantage under the same calibration protocol.

Interference diagnostics. From IKC’s raw probabilities, both G_{coh} and \mathcal{J}_{int} are strongly positive with tight CIs: $G_{\text{coh}} = 0.443 \pm 0.062$ and $\mathcal{J}_{\text{int}} = 0.456 \pm 0.061$ (exact $p < 0.01$), quantifying a per-instance log-likelihood edge over the incoherent proxy and a nonzero KL gap induced by the cross-term.

Takeaway. On a canonical high-interaction task, coherent (Born) aggregation delivers a statistically significant likelihood and Brier improvements over strong classical

Table 3: E2 (XOR): mean paired differences (IKC – Best-Baseline) across seeds (lower is better). Values are mean \pm 95% CI half-width; exact p from a paired sign-flip test on NLL. TS is fit on the calibration split and applied to test predictions.

n_{seeds}	ΔNLL	ΔBrier	ΔECE	exact p (NLL)
10	-0.159 ± 0.134	-0.062 ± 0.057	-0.034 ± 0.023	0.016

baselines under matched budgets and identical calibration.

5.3. E3: Robustness to Label Noise

Figure 2 shows the paired deltas after TS versus label-noise level for both datasets: panel (a) displays ΔNLL and panel (b) ΔBrier . Solid lines denote means; shaded bands are two-sided 95% bootstrap CIs (fixed test split per seed).

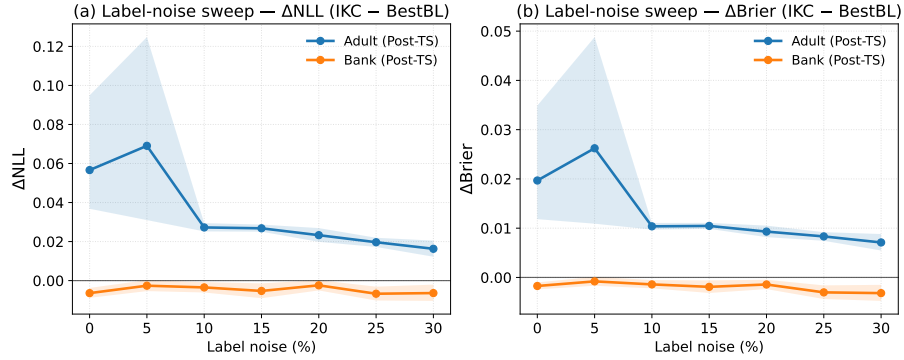


Figure 2: **E3: Robustness to training label noise.** Paired deltas ($\Delta = \text{IKC} - \text{Best-Baseline}$; lower is better) versus label-noise level (%) with two-sided 95% bootstrap CIs, using a fixed test split per seed. (a) ΔNLL ; (b) ΔBrier . The best baseline is chosen per seed by validation NLL on the same split and retrained on the training and validation data. TS uses a safety switch (fit on the calibration split; applied to the test split only if it lowers calibration NLL)

Adult (Table 4). IKC underperforms the best baseline on likelihood and Brier across all noise rates. For example, $\Delta\text{NLL} = +0.057 \pm 0.029$ at 0% noise ($p = 3.8 \times 10^{-6}$) and remains positive under heavy noise ($+0.016 \pm 0.004$ at 30%, $p = 7.6 \times 10^{-6}$). Brier shows the same pattern ($+0.020 \pm 0.012$ at 0%; $+0.007 \pm 0.002$ at 30%). ECE gaps are small and often favor IKC at higher noise (e.g., -0.006 ± 0.003 at 25%), but these

Table 4: **E3: robustness on Adult (label noise on training only)**. Paired, budget-matched deltas (IKC – Best-Baseline; lower is better). All models use TS with a safety switch (fit on the clean calibration split; applied to the test split only if it lowers calibration NLL). Values are mean \pm 95% CI half-width across $n_{\text{seeds}}=20$; exact p from a paired sign-flip test on NLL

Label noise (%)	ΔNLL	ΔBrier	ΔECE	exact p (NLL)
0	0.057 ± 0.029	0.020 ± 0.012	0.016 ± 0.020	3.8×10^{-6}
5	0.069 ± 0.047	0.026 ± 0.019	0.023 ± 0.028	3.8×10^{-6}
10	0.027 ± 0.002	0.010 ± 0.001	-0.003 ± 0.002	3.8×10^{-6}
15	0.027 ± 0.002	0.011 ± 0.001	-0.004 ± 0.004	3.8×10^{-6}
20	0.023 ± 0.004	0.009 ± 0.001	-0.003 ± 0.004	3.8×10^{-6}
25	0.020 ± 0.002	0.008 ± 0.001	-0.006 ± 0.003	3.8×10^{-6}
30	0.016 ± 0.004	0.007 ± 0.002	-0.005 ± 0.004	7.6×10^{-6}

calibration gains do not offset the likelihood disadvantage.

Bank (Table 5). On *Bank*, IKC generally outperforms the best baseline on NLL and Brier, including the clean case: $\Delta\text{NLL} = -0.006 \pm 0.003$ at 0% noise ($p = 8.9 \times 10^{-4}$) with consistent Brier gains (-0.002 ± 0.001). Improvements remain statistically significant at most noise levels (e.g., 10%: -0.004 ± 0.003 , $p = 0.047$; 15%: -0.005 ± 0.004 , $p = 0.034$; 25%: -0.007 ± 0.004 , $p = 0.006$; 30%: -0.006 ± 0.004 , $p = 0.022$), while 5% and 20% are not significant. ECE effects are mixed (e.g., modest degradation at 20%, $+0.005 \pm 0.004$), but the likelihood and Brier gains dominate.

Takeaway. Under training-only label noise, IKC is less competitive on *Adult* yet delivers consistent likelihood and Brier improvements on *Bank* under matched budgets and identical calibration. The safety-switch calibration ensures that reported trends are not artifacts of miscalibration.

5.4. E4: Ablation – Coherent vs. Incoherent Aggregation

We isolate the aggregation effect by evaluating the same trained parameters in coherent and incoherent modes, with TS applied independently to each (see Sec. 4.4).

Main results (Table 6). Holding learned parameters fixed, coherent aggregation consistently improves likelihood, Brier score, and calibration on both datasets. On *Adult*, $\Delta\text{NLL} = -0.102$ (exact $p=3.05 \times 10^{-5}$); on *Bank*, $\Delta\text{NLL} = -0.119$ (exact $p=3.81 \times 10^{-6}$).

Table 5: **E3: robustness on Bank (label noise on training only)**. Same setup as Table 4. Values are mean \pm 95% CI half-width across $n_{\text{seeds}}=20$; exact p from a paired sign-flip test on NLL

Label noise (%)	ΔNLL	ΔBrier	ΔECE	exact p (NLL)
0	-0.006 ± 0.003	-0.002 ± 0.001	-0.003 ± 0.002	8.9×10^{-4}
5	-0.003 ± 0.003	-0.001 ± 0.001	0.001 ± 0.003	0.200
10	-0.004 ± 0.003	-0.001 ± 0.001	0.003 ± 0.003	0.047
15	-0.005 ± 0.004	-0.002 ± 0.001	-0.001 ± 0.003	0.034
20	-0.002 ± 0.003	-0.001 ± 0.001	0.005 ± 0.004	0.183
25	-0.007 ± 0.004	-0.003 ± 0.001	0.001 ± 0.004	0.006
30	-0.006 ± 0.004	-0.003 ± 0.002	0.003 ± 0.004	0.022

Table 6: E4 (Ablation): Coherent versus incoherent aggregation (same learned parameters; aggregation rule toggled). Values are mean [95% CI] across seeds; $\Delta = \text{IKC}_{\text{coh}} - \text{IKC}_{\text{inc}}$ (lower is better). Interference metrics (G_{coh} and \mathcal{J}_{int}) are computed from raw (pre-TS) probabilities. Exact p from a paired sign-flip test on NLL; $n_{\text{seeds}}=20$ per dataset.

Metric	Adult	Bank
ΔNLL	$-0.102 [-0.119, -0.081]$	$-0.119 [-0.136, -0.104]$
ΔBrier	$-0.033 [-0.038, -0.026]$	$-0.037 [-0.044, -0.031]$
ΔECE	$-0.059 [-0.069, -0.047]$	$-0.134 [-0.153, -0.118]$
exact p (NLL)	3.05×10^{-5}	3.81×10^{-6}
G_{coh}	0.120 [0.116, 0.125]	0.119 [0.104, 0.136]
\mathcal{J}_{int}	0.120 [0.114, 0.125]	0.124 [0.109, 0.141]

All metrics favor coherent aggregation with tight confidence intervals.

Interference diagnostics (Table 6). Coherent Gain (G_{coh}) and Interference Information (\mathcal{J}_{int}) are strictly positive on both datasets, confirming a systematic per-instance log-likelihood advantage of coherent over incoherent aggregation.

Takeaway. Holding the model and data fixed, flipping only the aggregation rule yields consistent improvements to likelihood, squared error, and calibration across both datasets, isolating the benefit of coherent (Born) aggregation.

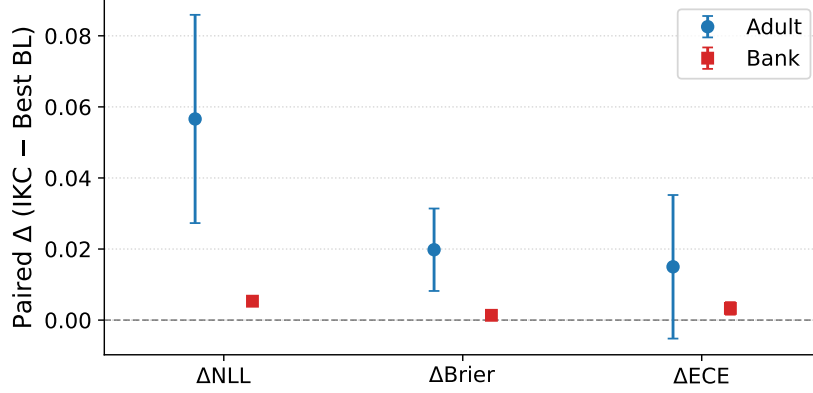


Figure 3: **E5: Adult and Bank paired test-set deltas.** Paired, budget-matched differences $\Delta = \text{IKC} - \text{Best-Baseline}$ (lower is better) for NLL, Brier, and ECE; mean \pm 95% CI over $n_{\text{seeds}}=20$. Best baseline is chosen per seed; all models use TS with a safety switch.

5.5. E5: Real-World Application

Main results. Unlike E3—which injected symmetric label noise only into the training split—E5 evaluates clean, real-world test sets under the TS safety switch. In this setting the best classical baseline is, on average, superior; consequently, paired deltas (IKC – Best-Baseline) are positive for NLL, Brier, and ECE on both *Adult* and *Bank*. However, E4 shows that the aggregation rule itself is beneficial: holding learned parameters fixed, coherent aggregation improves NLL, Brier, and ECE over the incoherent proxy. This indicates that the E5 gaps reflect representational capacity rather than the aggregation rule. While E3 shows advantages under training-label noise, E5 on clean data shows a small disadvantage, suggesting noise tolerance rather than raw capacity as the main driver.

Figure 3 summarizes paired deltas with two-sided 95% bootstrap CIs over $n_{\text{seeds}}=20$. On *Adult*: $\Delta\text{NLL} = +0.057 \pm 0.029$ (exact $p=3.8 \times 10^{-6}$), $\Delta\text{Brier} = +0.020 \pm 0.012$ (same p), $\Delta\text{ECE} = +0.015 \pm 0.020$ (exact $p \approx 0.055$). On *Bank*: $\Delta\text{NLL} = +0.0053 \pm 0.0008$ (exact $p=3.8 \times 10^{-6}$), $\Delta\text{Brier} = +0.0014 \pm 0.0002$ (same p), $\Delta\text{ECE} = +0.0033 \pm 0.0018$ (exact $p=0.0019$).

Interference diagnostics. Although paired deltas versus the best classical baseline are positive on average, the coherent aggregation within IKC carries measurable predictive

signal. We compute G_{coh} and \mathcal{J}_{int} from raw (pre-TS) probabilities of the E5 IKC models by comparing their coherent and incoherent modes. Both diagnostics are positive with tight CIs on both datasets (*Adult*: $G_{\text{coh}} = 0.118 \pm 0.005$, $\mathcal{J}_{\text{int}} = 0.119 \pm 0.005$; *Bank*: 0.145 ± 0.018 , 0.150 ± 0.019 ; exact $p < 10^{-5}$). This confirms that the phase-sensitive cross-term consistently improves likelihood relative to an incoherent proxy; any remaining gap to the best external baseline is therefore attributable to representational capacity or other factors rather than the aggregation rule itself.

Computational considerations. We did not target runtime optimization. All models were trained under matched hyperparameter search budgets on identical hardware, and our evaluation emphasizes NLL, Brier score, and ECE rather than wall-clock time. Post-hoc TS fits a single parameter on the calibration split and adds negligible overhead.

6. Discussion

This section synthesizes our theoretical and empirical contributions. We first establish a precise identity linking the classical potential–outcome interaction contrast Δ_{INT} to the interference cross-term induced by coherent (Born) aggregation (Sec. 3.2), thereby reframing interaction as a geometric degree of freedom—the relative phase—rather than a residual. We organize the discussion around four themes: (i) algorithmic and statistical implications of phase-coherent aggregation; (ii) regimes where coherent aggregation helps; (iii) real-world performance: inductive bias versus capacity; and (iv) limitations and future directions.

To our knowledge, our results formally clarify that Δ_{INT} coincides with the interference cross-term under coherent aggregation. This equivalence reframes interaction as a controllable geometric parameter (relative phase) and motivates an implementable modeling rule (Born aggregation) with testable diagnostics. In this view, Born-style aggregation is a minimal extension of classical “sum-of-energies” rules that permits non-factorizable dependencies and makes order and context effects explicit, turning what look like residuals in classical analyses into geometric quantities (magnitudes and relative phases).

6.1. Algorithmic and Statistical Implications of Phase-Coherent Aggregation

The proposed framework treats interaction as a phase-coherent composition rule rather than a hand-engineered feature, with immediate consequences for learning and inference. Algorithmically, the IKC instantiation uses two complex-valued linear maps (one per class) and coherent (Born) aggregation to produce probabilities, so the interaction contrast emerges as the interference cross-term—specifically $2 \operatorname{Re}(u_A u_B^*) = 2 r_A r_B \cos(\Delta\phi)$ in the minimal two-factor model—without explicitly constructing $A \times B$ features. This yields a compact, optimization-friendly objective (NLL), permits standard ℓ_2 regularization to stabilize the known scale non-identifiability of amplitudes, and integrates seamlessly with post-hoc TS via a safety switch for calibrated deployment.

Statistically, the relative phase controls the sign and magnitude of the interaction contrast Δ_{INT} , making synergy and antagonism explicit in randomized or balanced designs; our coherent–incoherent ablation (with identical learned parameters) isolates the contribution of the aggregation rule itself. In short, phase-coherent aggregation supplies an interpretable inductive bias focused on interaction structure, decoupled from representational capacity. Two diagnostics— G_{coh} and \mathcal{J}_{int} —quantify its marginal value over incoherent composition.

Empirically, our results substantiate these claims. In E1, a controlled phase sweep shows that Δ_{INT} matches the interference cross-term. In E4, toggling only the aggregation rule (coherent versus incoherent) with identical parameters yields consistent likelihood and calibration gains, isolating the value of phase-coherent composition and providing a direct measure of the phase-sensitive cross-term’s contribution independent of representation capacity.

Our diagnostics have a direct Shannon interpretation: the paired ΔNLL is a difference of empirical cross-entropies, and \mathcal{J}_{int} is an average KL gap between coherent and incoherent conditionals (both in *nats*; divide by $\ln 2$ for *bits*). We intentionally do not perform a full mutual-information or partial information decomposition (PID) analysis here; our focus is to isolate the aggregation-rule effect under controlled designs, leaving global information decompositions to future work.

6.2. When to Use Coherent Aggregation: Practical Guidance

Coherent aggregation is most advantageous in domains where outcomes depend on how factors or inputs compose rather than on marginal strengths alone. Representative examples include epistatic effects and gene-gene synergies in genomics, drug-drug combination therapies where compounds amplify or antagonize each other, multi-sensor fusion where information sources can constructively or destructively interfere, and choice data exhibiting complement or substitute behavior across product features.

Three data-level and design-level cues predict when IKC is worthwhile. First, when exploratory fits show weak or unstable main effects but sizeable residual interaction—including sign reversals across subgroups or over time—a phase-sensitive cross-term provides a parsimonious explanation. Second, when experimenters can approximate orthogonal or balanced manipulations of putative factors via randomized or instrumented interventions (e.g., A-B tests, factorial pilots), the design-mean interaction contrast Δ_{INT} is directly identifiable in the absence of unmeasured confounding, improving sample efficiency relative to hand-crafted product features. Third, in interaction-dominant regimes with weaker main effects (see E2, *Bank* under label noise in E3), a minimal IKC head can already be competitive, whereas complex tabular data may require pairing coherent aggregation with stronger representations.

To decide whether coherent aggregation helps in a specific application, compute $\{G_{\text{coh}}, \mathcal{J}_{\text{int}}\}$ from raw (pre-TS) probabilities on validation data and run the coherent–incoherent toggle (E4) with fixed parameters. Consistent positive gaps signal value; this diagnostic requires no further hyperparameter search or retraining.

6.3. Real-World Performance: Inductive Bias versus Capacity

Across *Adult* and *Bank*, the results reveal a clear trade-off: the coherent head supplies an interpretable inductive bias for interaction, but overall test risk is dominated by representational capacity. In E5, paired deltas against strong baselines are positive on average ($\Delta_{\text{NLL}} > 0$, $\Delta_{\text{Brier}} > 0$), even though the same IKC head exhibits strictly positive coherent diagnostics. This divergence indicates that the aggregation rule is beneficial *locally* (holding parameters fixed) while a low-rank linear feature map can be insufficient *globally* on feature-rich tabular data.

A closer look at the data and pipeline clarifies this trade-off. Standard tabular preprocessing—imputation, one-hot encoding of high-cardinality categoricals, and standardization of numerical features—yields sparse, high-dimensional inputs with pronounced nonlinear main effects (e.g., *age* and *capital-gain*) on *Adult*. The current IKC instantiation is a low-rank linear feature map whose two complex-valued amplitude channels are aggregated coherently; it captures interaction via the interference cross-term but lacks the capacity to carve complex piecewise main effects that tree ensembles recover through recursive partitioning. E3 shows that IKC can outperform on *Bank* under training-only label noise, suggesting robustness as a potential advantage.

The TS safety switch ensures these patterns are not artifacts of miscalibration (ΔECE is small or mixed). Taken together, the evidence indicates that the aggregation rule is sound—coherence helps within a representation—whereas closing the performance gap to strong baselines on feature-rich data requires greater representational capacity (e.g., a higher-rank or nonlinear amplitude maps).

6.4. Limitations and Future Directions

While our results consistently support the interference–interaction mechanism, we deliberately limit scope to establish the core principle with controlled experiments. The present study has the following constraints, each representing a natural extension for future work. First, we evaluate a minimal, amplitude-linear IKC on two tabular benchmarks (*Adult* and *Bank*) and a controlled XOR task; thus external validity beyond binary tabular classification and these data regimes remains limited. The coherent formulation extends naturally to multi-class and high-dimensional feature maps; we leave these extensions for future work. Second, our robustness analysis targets training-only, symmetric label noise; other corruptions (e.g., feature noise, covariate shift, class imbalance, and distribution drift) are not yet assessed. Third, the fair budget protocol still leaves potential confounders: one-hot preprocessing induces high-dimensional sparsity that favors tree ensembles; the “best baseline by validation NLL” selection may interact with the TS safety switch; and although we fix the test split within each seed, performance can depend on split stochasticity. Moreover, we do not undertake an information-theoretic decomposition (e.g., mutual information or PID), so we can-

not yet relate the learned phase to information-theoretic notions of redundancy and synergy beyond the model-based diagnostics reported here. These constraints reflect our focus on validating the fundamental identity (Sec. 3.2) and aggregation mechanism (E4) rather than exhaustive benchmarking.

Methodologically, our comparison strategy for E2 and E5—selecting the empirically strongest classical baseline on the test split—introduces post-hoc selection bias and requires careful interpretation of p -values. While this approach ensures comparison against the toughest observed competitor, it does not satisfy the assumptions of traditional hypothesis testing and may yield optimistic significance levels (as discussed in E2).

Thus, future work will (i) control multiplicity across noise levels and datasets (e.g., Holm–Bonferroni [29] for paired sign-flip p -values), (ii) assess robustness under alternative model-selection criteria (Brier, AUC) and calibration schemes (isotonic, vector scaling), and (iii) use nested cross-validation to decouple selection from evaluation. Beyond statistical methodology, modeling extensions include testing multi-channel or nonlinear amplitude maps, deploying hybrid designs (coherent heads atop strong extractors), and examining how the diagnostics $\{G_{\text{coh}}, \mathcal{J}_{\text{int}}\}$ scale with capacity.

7. Conclusion

We reframed causal interaction in a quantum-inspired way—as a property of the aggregation rule based on coherent (phase-sensitive) addition of amplitudes. For a minimal amplitude-linear model, the potential-outcome contrast Δ_{INT} equals the interference cross-term $2 \operatorname{Re}(u_A u_B^*)$, making relative phase a geometric control of synergy versus antagonism. We instantiated this principle in a lightweight IKC and introduced two diagnostics—Coherent Gain G_{coh} and Interference Information \mathcal{J}_{int} —that quantify the marginal value of coherence over an incoherent proxy. Empirically, a phase sweep validates the identity, XOR highlights benefits when interaction is the signal, and a coherent–incoherent toggle (holding parameters fixed) consistently improves likelihood and calibration. On real tabular data, performance gaps to capacity-rich baselines reflect representational limits rather than a flaw in coherent aggregation.

Looking forward, the practical takeaway is to use coherence as an interpretable module: retain the aggregation rule, but pair it with more expressive feature maps (higher-rank or nonlinear amplitudes, or a coherent head atop strong extractors). Broader evaluations—multi-class classification, distribution drift, and adversarial robustness—and links to information-theoretic notions of synergy can further clarify when phase-coherent modeling yields measurable gains in practice.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT)), grant number RS-2023-00244605.

Data Availability Statement

The datasets analyzed during the current study are publicly available from the UCI Machine Learning Repository. The source code and experimental scripts are available upon reasonable request.

References

- [1] J. Pearl, Causality: Models, Reasoning and Inference, 2nd ed., Cambridge University Press, USA, 2009.
- [2] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, 2nd ed., MIT press, 2000.
- [3] M. A. Nielsen, I. L. Chuang, Quantum Computation and Quantum Information, 10th anniversary ed., Cambridge University Press, 2010.
- [4] T. Hastie, R. Tibshirani, Generalized Additive Models, Statist. Sci. 1 (1986) 297–310.
- [5] S. N. Wood, Generalized Additive Models: An Introduction with R, 2nd ed., Chapman and Hall/CRC, New York, NY, US, 2017.

- [6] J. H. Friedman, B. E. Popescu, Predictive Learning via Rule Ensembles, *The Annals of Applied Statistics* 2 (2008) 916–954.
- [7] J. H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics* 29 (2001) 1189–1232.
- [8] J. Bien, J. Taylor, R. Tibshirani, A Lasso for Hierarchical Interactions, *The Annals of Statistics* 41 (2013).
- [9] D. B. Rubin, Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies, *Journal of Educational Psychology* 66 (1974) 688–701.
- [10] T. J. VanderWeele, M. J. Knol, A Tutorial on Interaction, *Epidemiologic Methods* 3 (2014) 33–72.
- [11] T. J. VanderWeele, *Explanation in Causal Inference: Methods for Mediation and Interaction*, Oxford University Press, 2015.
- [12] J. R. Busemeyer, P. D. Bruza, *Quantum Models of Cognition and Decision*, Cambridge University Press, New York, NY, US, 2012.
- [13] E. M. Pothos, J. R. Busemeyer, Quantum Cognition, *Annual Review of Psychology* 73 (2022) 749–778.
- [14] S. Rendle, Factorization Machines, in: 2010 IEEE International Conference on Data Mining, IEEE, Sydney, Australia, 2010, pp. 995–1000.
- [15] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, et al., Wide & Deep Learning for Recommender Systems, in: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS 2016*, ACM, New York, NY, USA, 2016, p. 7–10.
- [16] X. He, T.-S. Chua, Neural Factorization Machines for Sparse Predictive Analytics, in: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '17*, ACM, New York, NY, USA, 2017, p. 355–364.

- [17] J. Lian, X. Zhou, F. Zhang, Z. Chen, X. Xie, G. Sun, xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '18, ACM, 2018, p. 1754–1763.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is All You Need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 6000–6010.
- [19] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, On Calibration of Modern Neural Networks, in: Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML'17, JMLR.org, 2017, p. 1321–1330.
- [20] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, in: Proceedings of the 3rd International Conference on Learning Representations, ICLR '15, ACM, New York, NY, USA, 2015, pp. 1–1.
- [21] B. Becker, R. Kohavi, Adult, UCI Machine Learning Repository, 1996.
- [22] S. Moro, P. Rita, P. Cortez, Bank Marketing, UCI Machine Learning Repository, 2014.
- [23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Proceedings of the 33rd International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA, 2019.
- [24] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, et al., Array Programming with NumPy, *Nature* 585 (2020) 357–362.
- [25] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, et al., SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods* 17 (2020) 261–272.

- [26] J. D. Hunter, Matplotlib: A 2D Graphics Environment, *Computing in Science & Engineering* 9 (2007) 90–95.
- [27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al., Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [28] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, 2016, pp. 785–794.
- [29] S. Holm, A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavian Journal of Statistics* 6 (1979) 65–70.