REDUNDANCY AND SYNERGY IN MULTIVARIATE GAUSSIANS VIA THE BLACKWELL ORDER

Anonymous authors
Paper under double-blind review

ABSTRACT

The goal of the partial information decomposition (PID) is to quantify the redundant and synergistic information that multiple sources provide about a target. PID has many applications in machine learning, neuroscience, and other fields, but defining and computing it for high-dimensional continuous systems remains challenging. Here, we define a PID for multivariate Gaussian systems using the Blackwell order, which formalizes when one channel is more informative than another. We show that, among existing redundancy measures that apply to any number of sources, Blackwell redundancy is the only one that satisfies a natural set of desiderata. We also prove that Gaussian channels are optimal for extracting both redundant and union information, yielding an exact and efficient PID as well as an intuitive geometric interpretation in terms of nested ellipsoids. Our union information and synergy coincide with the well-known BROJA measures, and we derive closed-form expressions for both in the case of two sources. We demonstrate the scalability of our method on systems with up to a thousand dimensions or sources. Finally, we illustrate the approach on an optimal control problem, uncovering redundant and synergistic interactions between sensor and memory.

1 INTRODUCTION

In many domains, one seeks to understand how information about a target variable Y is distributed among a set of source variables $X_1, \ldots, X_n$. For example, in neuroscience, one may ask how information about a stimulus is encoded across different brain regions (Luppi et al., 2024); in machine learning, one may ask how different features, layers, or modalities combine to produce the model's output (Liang et al., 2023); and so on. To solve this problem, the *partial information decomposition* (PID) (Williams & Beer, 2010) aims to decompose the information that the sources have about the target into *redundancy* (present in each individual source), *union information* (present in at least one source), *unique information* (information in one source beyond that shared by all sources), and *synergy* (present only in the joint outcome of the sources). In general, the PID is inspired by an analogy between information theory and set theory: if the information provided by each source is imagined as a set, then redundancy is the size of the intersection of these sets, while synergy is the difference between the size of the joint set and that of the union (Williams, 2011; Kolchinsky, 2022).

The PID requires an extension of standard information theory because existing quantities (such as mutual information) conflate synergy and redundancy (Williams & Beer, 2010). However, this extension faces conceptual and practical challenges. Conceptually, it has been difficult to define the PID in a way that is nonnegative, operationally meaningful, and well-defined for any number of sources. Nonetheless, recent work has shown that these difficulties can be overcome by making the analogy with set theory formal (Kolchinsky, 2022), as long as there is a definition of when one channel is less informative about the target than another. The *Blackwell order* provides a principled answer: one channel is less informative if it can be produced by stochastic processing of the other; equivalently, if it is no better for any decision problem (Blackwell, 1953). The resulting "Blackwell PID" leads to nonnegative measures of redundant, union, and synergistic information for any number of sources (Kolchinsky, 2022).

From a practical perspective, it is difficult to compute PID quantities, especially in continuous systems with many sources or high-dimensional variables. As an important first step, we address this challenge in the setting of multivariate Gaussians, which play a fundamental role in the study of

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

time-series prediction (autoregressive models), filtering (Kalman, 1960), deep learning (Saxe et al., 2014), latent-Gaussian modeling (Zhao et al., 2025), and various other settings.

We demonstrate that the Blackwell order can be used to define a principled PID for multivariate Gaussians, one that has an elegant geometric interpretation (Fig. 1). We show that, among existing redundancy measures that apply to any number of sources, Blackwell redundancy is the only nontrivial one that satisfies a natural set of desiderata (Sec. 4). Among other results, we prove that Gaussian channels are optimal for extracting both redundant and union information. In the case of two sources, we uncover interpretable closed-form expressions for union information and synergy. More generally, for any number of sources, we show that redundancy and union information can be efficiently computed by solving a convex optimization problem, which easily scales to a thousand dimensions or sources. We illustrate our approach on an optimal control problem, quantifying redundancy and synergy between sensing and memory. Extensions beyond Gaussians are also discussed.

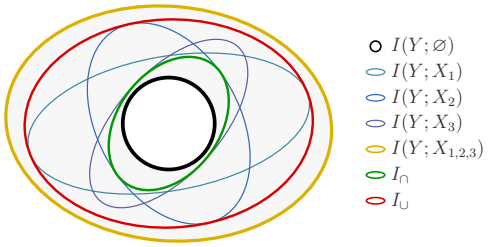


Figure 1: PID geometry for multivariate Gaussians, shown here for three sources X_1, X_2, X_3 and target $Y \in \mathbb{R}^2$. Each source X_i is represented by a blue ellipse whose log-area is the mutual information $I(X_i;Y)$; the black circle of unit area is the no-information channel. Redundancy $I_\cap$ corresponds to the largest (green) ellipse that is enclosed in every source and encloses the black circle; union information $I_\cup$ to the smallest (red) ellipse enclosing all sources. Synergy is the difference of log-areas between the joint-source ellipse (yellow) and the union ellipse.

Proofs and additional material are provided in the Supplementary Material.

2 Background: The Blackwell PID

We briefly review the Blackwell PID framework; see Kolchinsky (2022) for more details.

Recall that, in set theory, the intersection $\bigcap_i A_i$ of sets $A_1, \ldots, A_n$ is the largest set contained in each A_i; their union $\bigcup_i A_i$ is the smallest set that contains each A_i. The sizes of the intersection and the union can therefore be expressed variationally as

$$\left|\bigcap_i A_i\right| = \sup_{B: B \subseteq A_i \ \forall i} |B|, \qquad \left|\bigcup_i A_i\right| = \inf_{B: A_i \subseteq B \ \forall i} |B|. \tag{1}$$

These definitions require only a notion of set size and the subset relation.

Kolchinsky (2022) showed that information-theoretic redundancy (i.e., intersection) and union information can be defined in a similar way. The analogue of a set is a *channel* $P_{A|Y}$ (a conditional distribution of A given the target), the analogue of set size is the mutual information $I(A;Y)$, and the analogue of set inclusion is an ordering relation $P_{B|Y} \sqsubseteq_Y P_{C|Y}$, which specifies when channel B is "less informative" about Y than channel C. The redundancy of source channels $P_{X_1|Y}, \ldots, P_{X_n|Y}$ about target Y is the most information in any channel that is less informative than every source; union information is the least information in any channel that is more informative than every source:

$$I_\cap := \sup_{P_{Q|Y}: P_{Q|Y} \sqsubseteq_Y P_{X_i|Y} \ \forall i} I(Q;Y), \qquad I_\cup := \inf_{P_{Q|Y}: P_{X_i|Y} \sqsubseteq_Y P_{Q|Y} \ \forall i} I(Q;Y). \tag{2}$$

Importantly, the optimizing channels $P_{Q^*|Y}$ identify the content of redundant and union information, not only their amount. Using the above quantities, we may also define the *unique information* in source i as $I(Y;X_i) - I_\cap$, the *excluded information* (union information absent from source i) as $I_\cup - I(Y;X_i)$, and *synergy* as the information in the joint sources beyond their union, $I_{\text{syn}} := I(Y;X_1,\ldots,X_n) - I_\cup$. All these terms are nonnegative whenever the differences are defined.

The definitions in Eq. (2) rely on a choice of ordering relation. Perhaps the best-known and most principled choice is the *Blackwell order* (Blackwell, 1953). A channel $P_{B|Y}$ is Blackwell less informative than $P_{C|Y}$, written $P_{B|Y} \sqsubseteq_Y P_{C|Y}$, if there is a Markov kernel κ from C to B such that $P_{B|Y}(E|y) = \int \kappa(E|c)\, P_{C|Y}(dc|y)$ for every measurable set E and P_Y-almost every y. In other

words, B can be simulated by *garbling* the output of C, without further access to Y. The Blackwell order has a fundamental operational characterization in decision theory (Blackwell, 1953): $P_{B|Y} \sqsubseteq_Y P_{C|Y}$ iff an agent with C can do at least as well as an agent with B on every decision problem about Y. Recently, the Blackwell order has attracted growing attention in machine learning applications (Feng et al., 2026; Guo et al., 2026; Halder et al., 2025; Kaissis et al., 2024).

Combining the Blackwell order with Eq. (2) gives the *Blackwell PID*: Blackwell redundancy $I_\cap$ is the most information one can have about Y without being able to outperform any single source on any decision problem; Blackwell union information $I_\cup$ is the least information needed in order to perform at least as well as every source. These measures are defined for any number of sources, uniquely satisfy a natural set of axioms, and behave sensibly on examples; importantly, they also possess simple operational interpretations in terms of decision theory. In addition, Blackwell redundancy has recently been related to the information bottleneck and privacy preservation (maximizing predictive information while hiding conditional source identity) (Kolchinsky, 2024).

Note that Blackwell redundancy and union information are formal duals, but their values are not tied together by inclusion–exclusion (Möbius inversion). In fact, it has been argued that inclusion–exclusion should not apply to the PID, otherwise leading to inconsistencies (Kolchinsky, 2022).

3 Other prior work

An information-theoretic definition of union information and synergy, commonly referred to as "BROJA", was proposed by Bertschinger et al. (2014) and Griffith & Koch (2014). The union information is defined as the smallest joint information compatible with the source–target marginals,

$$I_\cup^{\mathrm{BR}} := \min_{\tilde{P}:\, \tilde{P}_{Y,X_i} = P_{Y,X_i}\ \forall i} I_{\tilde{P}}(Y; X_1, \ldots, X_n); \tag{3}$$

BROJA synergy is defined as $I_{\mathrm{syn}}^{\mathrm{BR}} := I(Y; X_1, \ldots, X_n) - I_\cup^{\mathrm{BR}}$. Although they are defined in different ways, it turns out that Blackwell and BROJA union information/synergy are numerically equal, $I_\cup = I_\cup^{\mathrm{BR}}, I_{\mathrm{syn}} = I_{\mathrm{syn}}^{\mathrm{BR}}$ (Kolchinsky, 2022).

Barrett (2015) first considered the PID in the Gaussian setting. Given a scalar target and two sources, he showed that a broad class of redundancy measures reduces to "minimum mutual information" (MMI) redundancy, defined as $\min_i I(Y; X_i)$. For multivariate targets, the MMI result no longer applies, because different sources can capture information about different directions of the target, which MMI cannot distinguish.

Recent work has considered BROJA PID for high-dimensional Gaussian and other continuous variables. Venkatesh et al. (2023) developed a Gaussian-restricted BROJA solver based on projected gradients, leaving the optimality of Gaussian solutions open. Zhao et al. (2025) subsequently proved Gaussian optimality, introduced the faster Thin-PID solver, and used normalizing flows to extend the method to latent-Gaussian representations. Goswami & Merkley (2024) derive analytical PIDs for special affine families using Blackwell comparability, Ma & Yu (2025, Appendix F) consider Gaussian-restricted union optimization for three sources and a scalar target, and Pakman et al. (2021) estimate BROJA unique information for continuous variables using copulas and variational optimization. Liang et al. (2023) introduce estimators of BROJA PID quantities that scale to high-dimensional multimodal data, and use them to analyze interactions in multimodal datasets and models. Liardi et al. (2025) extend BROJA union information to multiple targets by preserving only the pairwise marginals between individual source and target components.

As we discuss below, several alternative Gaussian PIDs have also been proposed.

4 Why Blackwell redundancy?

Before proceeding, we motivate Blackwell redundancy, the most novel term in the Blackwell PID.

In the past, defining a measure of redundancy that works for more than two sources has proven remarkably difficult. In fact, many existing redundancy measures are defined only for two sources, including BROJA (Bertschinger et al., 2014) and measures based on deficiency (Venkatesh & Schamberg, 2022), dependency constraints (Kay & Ince, 2018), information geometry (Kay, 2024), and

conditional independence (Lyu et al., 2026). To our knowledge, besides Blackwell redundancy, there are five other redundancy measures that are defined for any number of sources and can be evaluated for Gaussian variables: MMI (Barrett, 2015), $I_{\min}$ (Williams & Beer, 2010), I_{PM} (Milzman & Lyzinski, 2022), I_{ccs} (Ince, 2017), and I_{sx} (Ehrlich et al., 2024).[1]

As we now show, among these measures, Blackwell redundancy $I_\cap$ is the only one that satisfies three natural properties: (1) nonnegativity, (2) monotonicity, so redundancy does not increase when additional sources are added (i.e., $A \cap B \cap C$ cannot be larger than $A \cap B$), and (3) vanishing redundancy if the sources carry information about independent features of the target; in particular, redundancy should vanish if the target $Y = (Y_1, Y_2)$ contains two independent features $Y_1 \perp Y_2$, and each source depends on only one of them, $P_{X_1|Y} = P_{X_1|Y_1}$ and $P_{X_2|Y} = P_{X_2|Y_2}$. The first two conditions are core axioms of the PID (Williams, 2011, Ch. 4), while the third extends the "Independent Identity" property (Ince, 2017) to noisy observations of independent target features.

We construct two Gaussian examples to test these properties (see Sec. A.5). Let $Y_1, Y_2, \xi_1, \xi_2, \xi_3$ be independent standard Gaussians. In the COPY example, the target has two independent features, $Y = (Y_1, Y_2)$, and the two sources are $X_1 = Y_1 + \xi_1$ and $X_2 = Y_2 + \xi_2$. Each source observes a different feature, so condition (3) requires $I_\cap(Y; X_1, X_2) = 0$. The COPY+ example adds a third source, $X_3 = Y_1 + Y_2 + \xi_3$. Adding a source should not increase redundancy, so the three-source redundancy must remain zero. Table 1 reports redundancy (in nats). MMI, $I_\cap^{\min}$, $I_\cap^{\mathrm{PM}}$, and $I_\cap^{\mathrm{sx}}$ already give positive redundancy for COPY, violating condition (3). $I_\cap^{\mathrm{ccs}}$ gives zero for COPY but positive redundancy for COPY+, violating monotonicity. Among the measures in the table, only Blackwell redundancy gives the expected zero in both examples.

Table 1: Redundancy in nats

	COPY	COPY+
Expected	**0**	**0**
$I_\cap$ (ours)	**0**	**0**
MMI	0.35	0.35
$I_\cap^{\mathrm{ccs}}$	**0**	0.02
$I_\cap^{\min}$	0.19	0.17
$I_\cap^{\mathrm{PM}}$	0.35	0.40
$I_\cap^{\mathrm{sx}}$	0.35	0.42

5 Blackwell PID for Gaussians

We now present the Blackwell PID for multivariate Gaussians. Suppose that the target $Y \in \mathbb{R}^{d_Y}$ is multivariate Gaussian with mean μ_Y and positive-definite covariance $\Sigma_Y \succ 0$. Suppose also that each source $X_i \in \mathbb{R}^{d_i}$ is jointly Gaussian with the target, so that $P_{X_i|Y}(\cdot|y) = \mathcal{N}\big(\mu_{X_i} + H_i(y - \mu_Y), \Sigma_{X_i|Y}\big)$ with $H_i \in \mathbb{R}^{d_i \times d_Y}$ and $\Sigma_{X_i|Y} \succeq 0$. Since redundancy and union information depend only on the source–target marginals, it is enough that each pair (X_i, Y) is jointly Gaussian; synergy additionally involves the joint distribution of all sources and the target.

Importantly, the Blackwell order can be written as a matrix inequality between posterior covariances. (Here $\preceq$ is the Loewner order, where $V \preceq W$ means that $W - V$ is positive semidefinite.)

Theorem 1. *Let A, B be sources with a common target $Y \in \mathbb{R}^{d_Y}$ of arbitrary distribution, whose posteriors are Gaussian with constant covariance: $P_{Y|A=a} = \mathcal{N}(m_A(a), \Sigma_{Y|A})$ and $P_{Y|B=b} = \mathcal{N}(m_B(b), \Sigma_{Y|B})$ almost surely, where m_A, m_B are measurable and $\Sigma_{Y|A}, \Sigma_{Y|B} \succeq 0$. Then*

$$P_{A|Y} \sqsubseteq_Y P_{B|Y} \quad \iff \quad \Sigma_{Y|A} \succeq \Sigma_{Y|B}. \tag{4}$$

For jointly Gaussian channels, Theorem 1 reduces to a classical result by Hansen & Torgersen (1974); see also Venkatesh & Schamberg (2022, Theorem 2). In this result, we do not assume joint Gaussianity, only that the target may be written as a (possibly nonlinear) function of each source plus Gaussian noise. In Sec. 8, we use this to discuss generalizations beyond Gaussian assumptions.

Theorem 1 has a simple geometric interpretation. Associate with each channel $P_{A|Y}$ the ellipsoid $\mathcal{E}_A := \{z \in \mathbb{R}^{d_Y} : z^\top \Sigma_{Y|A} z \leq 1\}$. Then the Blackwell order is ellipsoid inclusion: $P_{A|Y} \sqsubseteq_Y P_{B|Y}$ if and only if $\mathcal{E}_A \subseteq \mathcal{E}_B$. The no-information channel corresponds to $\mathcal{E}_\varnothing := \{z : z^\top \Sigma_Y z \leq 1\}$, which is contained in every $\mathcal{E}_A$, and mutual information is a log-volume

[1] The Gács–Körner redundancy of Griffith et al. (2014), based on the common information of Gács & Körner (1973), is also defined for any number of sources. We exclude it because it is trivial in generic Gaussian settings, vanishing whenever the sources have a nondegenerate joint Gaussian distribution.

ratio, $I(A;Y) = \frac{1}{2}\ln(|\Sigma_Y|/|\Sigma_{Y|A}|) = \ln(\mathrm{Vol}\,\mathcal{E}_A/\mathrm{Vol}\,\mathcal{E}_\varnothing)$. Fig. 1 illustrates this geometry for a two-dimensional target in whitened coordinates, with areas normalized so that $\mathcal{E}_\varnothing$ has unit area.

In the following, it will be convenient to work in whitened target coordinates, using the *normalized posterior covariance*

$$C_A := \Sigma_Y^{-1/2}\Sigma_{Y|A}\Sigma_Y^{-1/2}\,. \tag{5}$$

This matrix satisfies $0 \preceq C_A \preceq I$ and $I(A;Y) = -\frac{1}{2}\ln|C_A|$. For the source channels $P_{X_i|Y}$, we abbreviate $C_i := C_{X_i}$. By Theorem 1, $P_{A|Y} \sqsubseteq_Y P_{B|Y}$ iff $C_A \succeq C_B$. Conversely, every matrix $0 \preceq R \preceq I$ is the normalized posterior covariance of some Gaussian channel (see Sec. A.1):

$$P_{Q_R|Y}(\cdot|y) = \mathcal{N}\Big((I-R)^{1/2}\Sigma_Y^{-1/2}(y-\mu_Y),\, R\Big)\,. \tag{6}$$

We write $\mathsf{GC}(d)$ for the set of Gaussian channels $P_{Q|Y}$ with $Q \in \mathbb{R}^d$, allowing singular conditional covariances, and $\mathbb{S}_+^{d_Y}$ for the set of positive semidefinite $d_Y \times d_Y$ matrices.

5.1 Redundancy

We now consider Blackwell redundancy for multivariate Gaussians. The definition in Eq. (2) optimizes over arbitrary channels, which may be non-Gaussian and infinite-dimensional. Our first result shows that the optimizer may be taken as Gaussian.

Theorem 2. *Suppose that each source–target pair (X_i, Y) is jointly Gaussian. Then the Blackwell redundancy $I_\cap$ in Eq.* (2) *may be computed by restricting the optimization to Gaussian channels of output dimension at most $d_Q = \dim \bigcap_i \mathrm{range}(I - C_i) \le \min(d_Y, \min_i d_i)$:*

$$I_\cap = \max_{P_{Q|Y} \in \mathsf{GC}(d_Q)} I(Q;Y) \quad \textit{where} \quad P_{Q|Y} \sqsubseteq_Y P_{X_i|Y} \;\; \forall i. \tag{7}$$

Combining Theorems 1 and 2 allows us to express redundancy as an optimization over covariance matrices. It is convenient to work with the inverse matrix $M := C_Q^{-1}$, which satisfies $I(Q;Y) = \frac{1}{2}\ln|M|$ and $M \succeq I$. The Blackwell constraint $C_Q \succeq C_i$ then reads $C_i^{1/2} M C_i^{1/2} \preceq I$, valid even when C_i is singular. $I_\cap$ can then be computed by solving the convex matrix optimization problem:

$$I_\cap = \sup_{M \in \mathbb{S}_+^{d_Y}} \frac{1}{2}\ln|M| \quad \text{where} \quad I \preceq M, \qquad C_i^{1/2} M C_i^{1/2} \preceq I \quad \forall i\,. \tag{8}$$

If all $C_i \succ 0$, or more generally if $\bigcap_i \ker C_i = \{0\}$, then Eq. (8) has a unique optimizer $M_\cap^*$ and the redundancy is finite. The optimal redundant channel is given by Eq. (6):

$$P_{Q_\cap^*|Y}(\cdot|y) = \mathcal{N}\big((I - M_\cap^{*-1})^{1/2}\Sigma_Y^{-1/2}(y-\mu_Y),\, M_\cap^{*-1}\big). \tag{9}$$

This channel specifies the content of the redundant information: the directions of the target on which all sources agree, and how precisely they are jointly observed. (If the kernel condition is not satisfied, then every source observes a common target direction with infinite precision and $I_\cap = +\infty$.)

Recall the geometric interpretation of the Blackwell order as ellipsoid inclusion. Then, $I_\cap$ corresponds to the largest ellipsoid that encloses the no-information ellipsoid $\mathcal{E}_\varnothing$ and is enclosed in every source ellipsoid (Fig. 1). This is the Gaussian analogue of set intersection.

When Y is scalar, $C_i = c_i$ are numbers and Eq. (8) gives the well-known "minimum mutual information" (MMI) redundancy $I_\cap = \min_i I(Y;X_i)$ (Barrett, 2015). This equivalence does not hold for the more interesting case of multivariate targets, where different sources may provide information about different target directions.

The no-information constraint. The constraint $I \preceq M$ in Eq. (8) arises because no channel can leave more posterior uncertainty about the target than observing nothing. Somewhat surprisingly, this constraint is not implied by the source constraints. In terms of the geometric picture in Fig. 1, the largest ellipsoid contained in all source ellipsoids need not contain the no-information ellipsoid, even though each source ellipsoid does. This is related to the fact that neither the Blackwell order nor the Loewner order is a lattice (Bertschinger & Rauh, 2014; Kadison, 1951), so common lower bounds need not have a greatest element.

For two sources, Blackwell redundancy and union information need not satisfy the inclusion–exclusion formula $I_\cap = I(Y;X_1) + I(Y;X_2) - I_\cup$ (Kolchinsky, 2022). In Gaussian systems with $C_1, C_2 \succ 0$, this formula holds precisely when dropping the no-information constraint $I \preceq M$ does not change the optimum of Eq. (8). In that case, the optimal matrices of the redundancy and union problems obey a matrix form of inclusion–exclusion,

$$M_\cap^{*-1} + W_\cup^* = C_1 + C_2, \tag{10}$$

where $W_\cup^*$ is the union optimizer of Sec. 5.2, so the redundant channel can be read off from the union channel; see Sec. A.4 for details.

5.2 Union information and synergy

Next, we turn to Blackwell union information for multivariate Gaussians. We first show that, as for redundancy, it suffices to optimize over Gaussian channels.

Theorem 3. *Suppose that each source–target pair (X_i, Y) is jointly Gaussian. Then the Blackwell union information $I_\cup$ in Eq. (2) may be computed by restricting the optimization to Gaussian channels of output dimension at most $d_Q = \dim \sum_i \operatorname{range}(I - C_i) \le \min(d_Y, \sum_i d_i)$:*

$$I_\cup = \min_{P_{Q|Y} \in \mathsf{GC}(d_Q)} I(Q;Y) \quad \textit{where} \quad P_{X_i|Y} \sqsubseteq_Y P_{Q|Y} \ \ \forall i. \tag{11}$$

Writing $W := C_Q \equiv \Sigma_Y^{-1/2} \Sigma_{Y|Q} \Sigma_Y^{-1/2}$, the Blackwell constraint $P_{X_i|Y} \sqsubseteq_Y P_{Q|Y}$ becomes $W \preceq C_i$, so union information is given by the convex matrix optimization problem:

$$I_\cup = \min_{W \in \mathbb{S}_+^{d_Y}} -\frac{1}{2} \ln |W| \quad \text{where} \quad W \preceq C_i \quad \forall i. \tag{12}$$

The optimizer $W_\cup^*$ is unique as long as all $C_i \succ 0$. An optimal union channel is obtained from Eq. (6), now with $R = W_\cup^*$:

$$P_{Q_\cup^*|Y}(\cdot|y) = \mathcal{N}\big((I - W_\cup^*)^{1/2} \Sigma_Y^{-1/2} (y - \mu_Y),\, W_\cup^*\big). \tag{13}$$

Eq. (12) is the geometric dual of the redundancy problem: we seek the smallest ellipsoid that encloses all source ellipsoids (Fig. 1). This may be interpreted as the Gaussian analogue of set union.

Remarkably, for two sources the union problem can be solved in closed form, so no iterative optimization is needed at all (see Sec. A.4).

Theorem 4. *Let V and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_{d_Y})$ solve the generalized eigenvalue problem $C_2 V = C_1 V \Lambda$ with $V^\top C_1 V = I$. Then Eq. (12) is solved by $W_\cup^* = V^{-\top} \operatorname{diag}(\min(1, \lambda_j)) V^{-1}$ and*

$$I_\cup = \frac{1}{2}\big(I(Y;X_1) + I(Y;X_2)\big) + \frac{1}{4} \sum_{j=1}^{d_Y} |\ln \lambda_j| \,. \tag{14}$$

Synergy is the information available from the joint sources beyond the union channel, $I_{\text{syn}} = I(Y;X_1,\ldots,X_n) - I_\cup \ge 0$. If the full joint distribution of $(Y, X_1, \ldots, X_n)$ is Gaussian, then $I_{\text{syn}} = -\frac{1}{2} \ln |C_{X_1,\ldots,X_n}| - I_\cup$, the difference between the log-volumes of the joint-source ellipsoid and the union ellipsoid (Fig. 1).

Since Blackwell union information and synergy equal their BROJA counterparts (Eq. (3)), Theorem 3 also establishes that Gaussian channels are optimal for BROJA union information with any number of Gaussian sources. Zhao et al. (2025) derived a similar result in the case of two sources. In addition, our results provide a closed-form solution for BROJA union/synergy in the case of two sources, and more generally they replace the BROJA optimization over high-dimensional joint distributions (Zhao et al., 2025) by an optimization over a single $d_Y \times d_Y$ matrix.

For a scalar target, Eq. (12) gives $I_\cup = \max_i I(Y;X_i)$, so union information is the information in the strongest source X_{i^*}, and synergy is the conditional mutual information $I(Y;X_1,\ldots,X_n|X_{i^*})$ (Barrett, 2015). Union information and synergy that do not reduce to these standard mutual information expressions require a multidimensional target.

6 NUMERICAL OPTIMIZATION

We now describe efficient algorithms for computing the Gaussian Blackwell PID and evaluate their performance.

Gaussian redundancy and union, Eqs. (8) and (12), are defined in terms of convex log-determinant problems in a single $d_Y \times d_Y$ matrix, so they can be solved with off-the-shelf solvers such as CVXPY (Diamond & Boyd, 2016). However, we obtained much better performance by combining a Newton–PCG optimizer and log-barrier methods for the constraints (see Sec. C for details). Our implementation uses JAX to just-in-time compile the full optimization loop.

The size of our optimization problem is set by the target dimension d_Y. For two vector sources, existing Gaussian BROJA solvers optimize a $d_1 \times d_2$ cross-noise covariance (Venkatesh et al., 2023; Zhao et al., 2025), so working in target space can substantially reduce the computational cost when the target dimension is modest relative to the source dimensions (Fig. 2).

In addition, when the sources are low-dimensional relative to the target, each source observes only a low-dimensional subspace of the target and our optimization problems can be reparameterized in terms of smaller matrices. Specifically, for redundancy, the constraints $I \preceq M \preceq C_i^{-1}$ imply that $M - I \succeq 0$ and that its range is contained in the subspace $\bigcap_i \mathrm{range}(C_i^{-1} - I)$, with dimension $q \leq \min(d_Y, \min_i d_i)$. Writing $P \in \mathbb{R}^{d_Y \times q}$ for an orthonormal basis of this subspace, every feasible matrix is $M = I + PKP^\top$ with $K \in \mathbb{S}_+^q$, and Eq. (8) becomes a problem in the $q \times q$ matrix K. For union, write the whitened channels as $X_i = H_i Y + N_i$,[2] so that source i depends on the target only through the projection of Y onto the row space of H_i. Together, the sources depend on the target only through the subspace $\sum_i \mathrm{range}(H_i^\top)$, with dimension $r \leq \min(d_Y, \sum_i d_i)$. Directions orthogonal to this subspace can be dropped: if $P_\cup \in \mathbb{R}^{d_Y \times r}$ is an orthonormal basis of the observed subspace, then $P_\cup^\top Y$ is an r-dimensional target and $H_i Y = H_i P_\cup (P_\cup^\top Y)$, so replacing each H_i by $H_i P_\cup$ gives an r-dimensional problem with the same union information.

In the case of two sources, union and redundancy can often be computed much faster. Everything is driven by the generalized eigenvalues λ_j of the pair (C_2, C_1), which avoids forming matrix square roots or inverses. To compute union information for two sources, we evaluate the closed form of Theorem 4, $I_\cup = -\frac{1}{2}(\ln|C_1| + \sum_j \ln\min(1, \lambda_j))$, which requires only the eigenvalues, with no eigenvectors and no reconstruction of the optimal matrix $W_\cup^*$. Synergy then follows by subtraction. To compute redundancy for two sources, we use the logic of Eq. (10): from the closed-form union optimizer $W_\cup^*$ of Theorem 4, we form the candidate $R = C_1 + C_2 - W_\cup^*$. If $R \preceq I$, the no-information constraint is not binding, so $M_\cap^* = R^{-1}$ and $I_\cap = I(Y;X_1) + I(Y;X_2) - I_\cup$. Otherwise, we solve Eq. (8) with the Newton–PCG optimizer.

Figure 2 shows runtimes on random Gaussian channels $X_i = H_i Y + N_i$, with dimensions and source counts up to 10^3 and matched instances across methods (Sec. C.4). For redundancy, we compare Newton–PCG with and without the two-source fast path. For union, both Newton formulations use the same JAX optimization driver and each stops based on its own certificate. The source-covariance baseline implements the Gaussian BROJA formulation considered by Venkatesh et al. (2023), Zhao et al. (2025), and Ma & Yu (2025). The benefit of working in target space is clearest when both sources grow: at $d_Y = 8$ and $d_1 = d_2 = 10^3$ in panel (h), source-space Newton takes 47.8 s, compared with 0.038 s for our Newton solver and 0.036 s for the closed form. With 10^3 sources of dimension eight, our redundancy and union solvers take 0.13 s and 0.27 s, respectively, while source-space union reaches the 100-s limit. Comparisons with the original Thin-PID RProp algorithm and a certificate-assisted variant appear in Fig. S1.

7 EXAMPLE: OPTIMAL CONTROL

We now illustrate our approach by characterizing information processing in a simple optimal control model. Model details are provided in Sec. B.

We consider a linear–quadratic–Gaussian (LQG) controller that confines a two-dimensional particle $\mathbf{r} \in \mathbb{R}^2$. The particle diffuses as $d\mathbf{r} = (H\mathbf{r} + \mathbf{u})\,dt + d\mathbf{W}$, where $\mathbf{u}$ is the applied control and

[2] Whitened coordinates mean that Y has identity covariance and each noise N_i is standard Gaussian. This affine transformation leaves all information quantities unchanged.

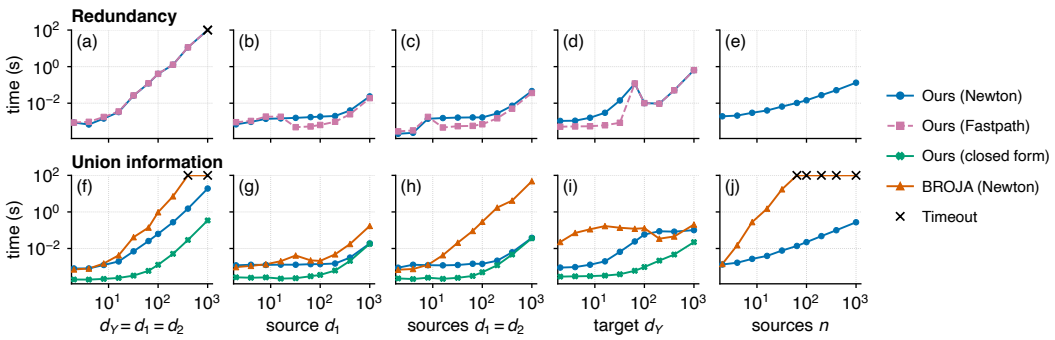


Figure 2: Runtimes across 5 trials for Blackwell redundancy (top) and union information (bottom), certified to 0.01 nats. Ours (Newton): target-space Newton–PCG; Ours (Fastpath): two-source redundancy shortcut with Newton fallback; Ours (closed form): evaluates the two-source union formula. BROJA (Newton): our Newton–PCG implementation of the BROJA source-covariance optimization (Venkatesh et al., 2023; Zhao et al., 2025; Ma & Yu, 2025). (a,f) $d_Y = d_1 = d_2$; (b,g) $d_Y = d_2 = 8$; (c,h) $d_Y = 8$, $d_1 = d_2$; (d,i) $d_1 = d_2 = 64$; (e,j) $d_Y = d_i = 8$, varying the source count. Crosses mark a 100-s timeout in at least one trial. Timings use one CPU thread, Apple M4 Pro (Sec. C.4).

H represents an elliptical trap with rotational drift. At each sampling interval Δt, the controller makes a noisy position measurement $\mathbf{s}_t = \mathbf{r}_t + \mathbf{n}_t$, where $\mathbf{n}_t \sim \mathcal{N}(0, \nu S)$ for an anisotropic covariance matrix S and noise scale ν. This measurement is combined with the memory $\mathbf{m}_t$, the Kalman prediction of the current position from earlier measurements, to compute the applied control $\mathbf{u}_t = -F\big[L\mathbf{s}_t + (I - L)\mathbf{m}_t\big] + \boldsymbol{\eta}_t$, where L is the Kalman gain, F is the feedback gain, and $\boldsymbol{\eta}_t$ is actuator noise. The memory is then updated.

To study information processing in the controller, we consider the current sensor reading $X_1 = \mathbf{s}_t$ and the memory $X_2 = \mathbf{m}_t$ as two sources and the control force $Y = \mathbf{u}_t$ as the target, evaluated under the steady-state Gaussian distribution. Mutual information quantifies how informative each source (or both together) is about the control, but not whether and how the controller combines them in computing $\mathbf{u}$.

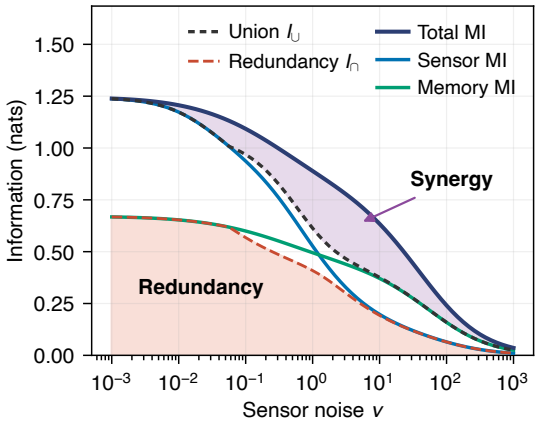


Figure 3: PID analysis of how the applied control (target) integrates two sources, sensor measurements and memory, as a function of sensor noise ν. Shading shows redundancy $I_\cap$ and synergy $I(Y; X_1, X_2) - I_\cup$. For more details, see Sec. B.

To make progress, we apply the Blackwell PID framework. We sweep the sensor noise strength ν at a fixed sampling interval, recomputing the optimal controller and the Blackwell PID quantities. Figure 3 shows a transition from redundant to synergistic to memory-dominated control as ν increases. At low sensor noise, the controller mainly uses the current measurement. Memory, which propagates earlier readings through the known dynamics, predicts the current sensor reading well but adds little further information: whatever memory knows about the control, the sensor knows as well. Memory is then Blackwell less informative than the sensor, its information is entirely redundant, and synergy is negligible. As ν grows, the sensor becomes noisy and the controller shifts weight toward memory. The two sources now carry different information: the measurement contains fluctuations that memory cannot anticipate, while memory contains evidence accumulated over earlier measurements. Because the control depends on the combination of the two, part of the information about it is available only jointly. Redundancy falls and synergy rises, peaking at 0.31 nats (39% of the total) near $\nu \simeq 2.6$.

At high noise the measurement is nearly ignored and the control follows memory; the sensor is now Blackwell less informative than memory, so its information is entirely redundant and the remainder of the union information is unique to memory.

8 Beyond Gaussians

In this paper, we focus on developing the Blackwell PID for multivariate Gaussians, the fundamental starting point for continuous systems. Extensions to non-Gaussian systems are a natural next direction and, as we note here, some of our results already apply in this setting. (Proofs are given in Sec. A.6.)

Theorem 1, which relates Blackwell and matrix orders, does not require Gaussian sources or targets, only that the target can be predicted from each source with additive homoscedastic Gaussian noise:

$$Y = m_i(X_i) + Z_i, \qquad Z_i \sim \mathcal{N}(0, \Sigma_{Y|X_i}), \qquad Z_i \perp X_i \quad \forall i, \tag{15}$$

where m_i may be nonlinear. For the formulas below, assume finite target covariance $\Sigma_Y \succ 0$ and $\Sigma_{Y|X_i} \succ 0$ for every i. Let $I_\cap^{\mathrm{G}}$ and $I_\cup^{\mathrm{G}}$ denote the values of the Gaussian programs (8) and (12), evaluated using the normalized conditional covariances $C_i = \Sigma_Y^{-1/2} \Sigma_{Y|X_i} \Sigma_Y^{-1/2}$.

Blackwell union information extends directly. Let $\tilde{Y}$ be Gaussian with the same mean and covariance as Y. Then $I_\cup = I_\cup^{\mathrm{G}} - D(P_Y \| P_{\tilde{Y}}) \leq I_\cup^{\mathrm{G}}$, where the divergence quantifies the target's deviation from Gaussianity. The optimal matrix $W_\cup^*$ is unchanged. Synergy follows as $I_{\mathrm{syn}} := I(Y; X_1, \ldots, X_n) - I_\cup$; computing it still requires the full joint mutual information.

Blackwell redundancy obeys the upper bound $I_\cap \leq I_\cap^{\mathrm{G}} - D(P_Y \| P_{\tilde{Y}}) \leq I_\cap^{\mathrm{G}}$. Moreover, if Y is Gaussian, both redundancy and union information are given exactly by $I_\cap^{\mathrm{G}}$ and $I_\cup^{\mathrm{G}}$, even if the sources do not have Gaussian marginals.

A complementary and more general approach, which we do not pursue here, is to learn invertible transformations into a Gaussian latent space, as in Flow-PID (Zhao et al., 2025). Separate invertible transformations of the sources and the target preserve mutual information and Blackwell comparisons, so our redundancy and union programs apply exactly if the transformed source–target pairs are Gaussian in the latent space; otherwise, they provide approximations. PIDReg (Ma & Yu, 2025) also applies Gaussian PID to learned representations, without requiring invertible encoders.

9 Discussion

We have shown that the Blackwell PID admits an exact and tractable characterization for arbitrarily many Gaussian sources and target dimensions, leading to a simple geometric interpretation and efficient numerical methods. We are also able to identify optimal channels, which extract the content of redundant and union information. This provides a powerful framework for studying how information is encoded across distributed representations.

Although we have not explored the content of redundant and union information here, such analyses could be useful in various data-driven applications. For example, in high-dimensional neuroimaging data (Luppi et al., 2024), they could be used to identify redundant modes shared across brain regions or timepoints. The framework also applies directly to linear-Gaussian dynamical systems, such as vector autoregressive and state-space models, where it can identify how information about the future is distributed across the past of different components (Creutzig et al., 2009).

Another direction is to study how the organization of information changes during learning. Deep linear networks (Saxe et al., 2014) with Gaussian inputs and noise provide a controlled setting in which to follow redundancy and synergy across layers and training trajectories, and the optimal channels could reveal whether learning causes different units to represent the same target features or to specialize in different ones. The control example also motivates studying how different kinds of information relate to task performance. Such questions could, in turn, guide the use of PID quantities as learning objectives.

REPRODUCIBILITY STATEMENT

All proofs are given in the Supplementary Material, Sec. A. Details of the optimization algorithm, including the Newton–PCG solver, are provided in Sec. C. The optimal-control model is described in Sec. B. The solver and timing benchmark code will be released upon publication, together with code reproducing the remaining figures and tables.

AI USE STATEMENT

AI assistance, including OpenAI Codex and Anthropic Claude, was used to assist with mathematical derivations, implementation of numerical solvers, generation of figures, and manuscript editing. The final version of the manuscript, including all mathematical proofs, was verified by the authors. Responsibility for all scientific claims, results, and manuscript text rests with the authors.

REFERENCES

Adam B. Barrett. Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. *Physical Review E*, 91(5):052802, 2015.

Nils Bertschinger and Johannes Rauh. The Blackwell relation defines no lattice. In *2014 IEEE International Symposium on Information Theory*, pp. 2479–2483. IEEE, 2014.

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pp. 265–272, 1953.

Felix Creutzig, Amir Globerson, and Naftali Tishby. Past-future information bottleneck in dynamical systems. *Physical Review E*, 79(4):041925, 2009. doi: 10.1103/PhysRevE.79.041925.

Amir Dembo. Information inequalities and uncertainty principles. *Dept. Statistics, Stanford Univ., Stanford, CA, Tech. Rep*, 75, 1990.

Amir Dembo, Thomas M. Cover, and Joy A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, 1991.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

David A. Ehrlich, Kyle Schick-Poland, Abdullah Makkeh, Felix Lanfermann, Patricia Wollstadt, and Michael Wibral. Partial information decomposition for continuous variables based on shared exclusions: Analytical formulation and estimation. *Physical Review E*, 110:014115, 2024. doi: 10.1103/PhysRevE.110.014115.

Alexandros Eskenazis and Lampros Gavalakis. On the entropy and information of Gaussian mixtures. *Mathematika*, 70(2):e12246, 2024. doi: 10.1112/mtk.12246.

Yiding Feng, Liuhan Qian, and Wei Tang. Is this predictor more informative than another? A decision-theoretical comparison. In *ACM Conference on Economics and Computation*, 2026. URL `https://arxiv.org/abs/2507.12094`.

Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.

Chaitanya Goswami and Amanda Merkley. Analytically deriving partial information decomposition for affine systems of stable and convolution-closed distributions. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-2754.

Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. In *Guided Self-Organization: Inception*, pp. 159–190. Springer, 2014.

Virgil Griffith, Edwin KP Chong, Ryan G James, Christopher J Ellison, and James P Crutchfield. Intersection information based on common randomness. *Entropy*, 16(4):1985–2000, 2014.

Ziyang Guo, Yifan Wu, Jason Hartline, and Jessica Hullman. The value of information in Human-AI decision-making. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/45824c7580778ad1ba384982b9e6b39f-Abstract-Conference.html.

Barproda Halder, Faisal Hamman, Pasan Dissanayake, Qiuyi Zhang, Ilia Sucholutsky, and Sanghamitra Dutta. Towards formalizing spuriousness of biased datasets using partial information decomposition. *Transactions on Machine Learning Research*, 2025. URL https://arxiv.org/abs/2407.00482.

Ole Havard Hansen and Erik N. Torgersen. Comparison of linear normal experiments. *The Annals of Statistics*, 2(2):367–373, 1974. doi: 10.1214/aos/1176342672.

Robin A. A. Ince. Measuring multivariate redundant information with pointwise common change in surprisal. *Entropy*, 19(7):318, 2017. doi: 10.3390/e19070318.

Richard V. Kadison. Order properties of bounded self-adjoint operators. *Proceedings of the American Mathematical Society*, 2(3):505–510, 1951. doi: 10.1090/S0002-9939-1951-0042064-2.

Georgios Kaissis, Stefan Kolek, Borja Balle, Jamie Hayes, and Daniel Rueckert. Beyond the calibration point: Mechanism comparison in differential privacy. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 22840–22860. PMLR, 2024. URL https://proceedings.mlr.press/v235/kaissis24a.html.

Rudolph E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960. doi: 10.1115/1.3662552.

Jim W Kay. A partial information decomposition for multivariate Gaussian systems based on information geometry. *Entropy*, 26(7):542, 2024.

Jim W. Kay and Robin A. A. Ince. Exact partial information decompositions for Gaussian systems based on dependency constraints. *Entropy*, 20(4):240, 2018. doi: 10.3390/e20040240.

Artemy Kolchinsky. A novel approach to the partial information decomposition. *Entropy*, 24(3):403, 2022. doi: 10.3390/e24030403.

Artemy Kolchinsky. Partial information decomposition: Redundancy as information bottleneck. *Entropy*, 26(7):546, 2024.

Paul Pu Liang, Yun Cheng, Xiang Fan, Chun Kai Ling, Suzanne Nie, Richard Chen, Zihao Deng, Nicholas Allen, Randy Auerbach, Faisal Mahmood, Ruslan Salakhutdinov, and Louis-Philippe Morency. Quantifying & modeling multimodal interactions: An information decomposition framework. *Advances in Neural Information Processing Systems*, 36:27351–27393, 2023.

Alberto Liardi, George Blackburne, Hardik Rajpal, Fernando E Rosas, and Pedro AM Mediano. A scalable estimator of higher-order information in complex dynamical systems. *arXiv preprint arXiv:2506.18498*, 2025.

Andrea I. Luppi, Fernando E. Rosas, Pedro A. M. Mediano, David K. Menon, and Emmanuel A. Stamatakis. Information decomposition and the informational architecture of the brain. *Trends in Cognitive Sciences*, 28(4):352–368, 2024. doi: 10.1016/j.tics.2023.11.005.

Aobo Lyu, Andrew Clark, and Netanel Raviv. Closed-form Gaussian estimators for multi-source partial information decomposition. *arXiv preprint arXiv:2605.09919*, 2026. URL https://arxiv.org/abs/2605.09919.

Zhaozhao Ma and Shujian Yu. Explainable multimodal regression via information decomposition. *arXiv preprint arXiv:2512.22102*, 2025. URL https://arxiv.org/abs/2512.22102.

Jesse Milzman and Vince Lyzinski. Signed and unsigned partial information decompositions of continuous network interactions. *Journal of Complex Networks*, 10(5):cnac026, 2022. doi: 10.1093/comnet/cnac026.

Ari Pakman, Amin Nejatbakhsh, Dar Gilboa, Abdullah Makkeh, Luca Mazzucato, Michael Wibral, and Elad Schneidman. Estimating the unique information of continuous variables. In *Advances in Neural Information Processing Systems*, volume 34, pp. 20295–20307, 2021.

Andrew M Saxe, James L McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. In *International Conference on Learning Representations*, 2014.

Praveen Venkatesh and Gabriel Schamberg. Partial information decomposition via deficiency for multivariate Gaussians. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pp. 2892–2897. IEEE, 2022.

Praveen Venkatesh, Corbett Bennett, Sam Gale, Tamina Ramirez, Greggory Heller, Severine Durand, Shawn Olsen, and Stefan Mihalas. Gaussian partial information decomposition: Bias correction and application to high-dimensional data. *Advances in Neural Information Processing Systems*, 36:74602–74635, 2023.

Paul L. Williams. *Information dynamics: Its theory and application to embodied cognitive systems*. Ph.D. thesis, Indiana University, 2011.

Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

Wenyuan Zhao, Adithya Balachandran, Chao Tian, and Paul Pu Liang. Partial information decomposition via normalizing flows in latent Gaussian distributions. *Advances in Neural Information Processing Systems*, 38, 2025.