MDPI

Decomposing and Tracing Mutual Information by Quantifying Reachable Decision Regions

Tobias Mages *[ID] and Christian Rohner [ID]

Department of Information Technology, Uppsala University, 752 36 Uppsala, Sweden; christian.rohner@it.uu.se
* Correspondence: tobias.mages@it.uu.se

Abstract: The idea of a partial information decomposition (PID) gained significant attention for attributing the components of mutual information from multiple variables about a target to being unique, redundant/shared or synergetic. Since the original measure for this analysis was criticized, several alternatives have been proposed but have failed to satisfy the desired axioms, an inclusion–exclusion principle or have resulted in negative partial information components. For constructing a measure, we interpret the achievable type I/II error pairs for predicting each state of a target variable (reachable decision regions) as notions of pointwise uncertainty. For this representation of uncertainty, we construct a distributive lattice with mutual information as consistent valuation and obtain an algebra for the constructed measure. The resulting definition satisfies the original axioms, an inclusion–exclusion principle and provides a non-negative decomposition for an arbitrary number of variables. We demonstrate practical applications of this approach by tracing the flow of information through Markov chains. This can be used to model and analyze the flow of information in communication networks or data processing systems.

Keywords: partial information decomposition; redundancy; synergy; information flow analysis

1. Introduction

A Partial Information Decomposition (PID) aims to attribute the provided information about a discrete target variable T from a set of predictor or viewable variables $\mathbf{V} = \{V_1, \ldots, V_n\}$ to each individual variable V_i. The partial contributions to the information about T may be provided by all variables (redundant or shared), by a specific variable (unique) or only be available through a combination of variables (synergetic/complementing) [1]. This decomposition is particularly applicable when studying complex systems. For example, it can be used to study logical circuits, neural networks [2] or the propagation of information over multiple paths through a network. The concept of synergy has been applied to develop data privacy techniques [3,4], and we think that the concept of redundancy may be suitable to study a notion of robustness in data processing systems.

Unfortunately and to the best of our knowledge, there does not exist a non-negative decomposition of mutual information for an arbitrary number of variables that satisfies the commutativity, monotonicity and self-redundancy axioms except the original measure of Williams and Beer [5]. However, this measure has been criticized for not distinguishing "the *same* information and the *same amount* of information" [6–9].

Here, we propose an alternative non-negative partial information decomposition that satisfies Williams and Beer's axioms [5] for an arbitrary number of variables. It provides an intuitive operational interpretation and results in an algebra like probability theory. To demonstrate that the approach distinguishes the same information from the same amount of information, we highlight its application in tracing the flow of information through a Markov chain, as visualized in Figure 1.

$$T \longrightarrow \mathbf{V} = (V_1, V_2) \longrightarrow \mathbf{Q} = (Q_1, Q_2) \longrightarrow \mathbf{R} = (R_1, R_2) \longrightarrow \hat{T}$$

$$H(T) \quad \geq \quad I(T; \mathbf{V}) \quad \geq \quad I(T; \mathbf{Q}) \quad \geq \quad I(T; \mathbf{R}) \quad \geq \quad I(T; \hat{T})$$

Shared   Unique 1   Unique 2   Synergetic
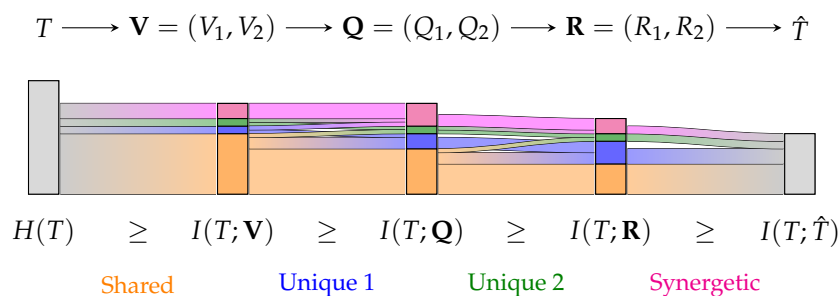
Figure 1. Visualization of a partial information decomposition with information flow analysis of a Markov chain as Sankey diagram. A partial information decomposition enables attributing the provided information about T to being shared (**orange**), unique (**blue/green**) or synergetic/complementing (**pink**). While this already offers practical insights for studying complex systems, the ability to trace the flow of partial information may create a valuable tool to model and analyze many applications.

This work is structured in three parts: Section 2 provides an overview of the related work and background information. Section 3 presents a representation of pointwise uncertainty, constructs a distributive lattice and demonstrates that mutual information is the expected value of its consistent valuation. Section 4 discusses applications of the resulting measure to PIDs and the tracing of information through Markov chains. We provide an overview of the used notation at the end of the paper.

2. Related Work

We briefly summarize partial orders and the four main publications which led to our proposed decomposition approach. This includes the PID by Williams and Beer [5], the quantification of unique information by Bertschinger et al. [10] and Griffith and Koch [11], the Blackwell order based on Bertschinger and Rauh [12], the evaluation of binary decision problems using Receiver Operating Characteristics and consistent lattice valuations by Knuth [13].

2.1. Partial Orders and Lattices

This section provides a brief overview of the relevant definitions on partial orders and lattices for the context of this work based on [9,13]. A binary ordering relation $\preccurlyeq$ on a set $\mathbf{L}$ is called a *preorder* if it is reflexive and transitive. If the ordering relation additionally satisfies an antisymmetry, then $(\mathbf{L}, \preccurlyeq)$ is called a *partially ordered set (poset)*. For $\alpha, \beta, \gamma \in \mathbf{L}$:

$$\alpha \preccurlyeq \alpha \qquad \text{(reflexivity)}$$
$$\text{if } \alpha \preccurlyeq \beta \text{ and } \beta \preccurlyeq \gamma \text{ then } \alpha \preccurlyeq \gamma \qquad \text{(transitivity)}$$
$$\text{if } \alpha \preccurlyeq \beta \text{ and } \beta \preccurlyeq \alpha \text{ then } \alpha = \beta \qquad \text{(antisymmetry)}$$

Two elements satisfy $\alpha \preccurlyeq \beta$, $\beta \preccurlyeq \alpha$ or may be incomparable, meaning $\alpha \not\preccurlyeq \beta$ and $\beta \not\preccurlyeq \alpha$. A partially ordered set has a *bottom element* $\bot \in \mathbf{L}$ if $\bot \preccurlyeq \alpha$ for all $\alpha \in \mathbf{L}$ and a *top element* $\top \in \mathbf{L}$ if $\alpha \preccurlyeq \top$ for all $\alpha \in \mathbf{L}$. For each element α, it can be defined a down-set ($\downarrow\alpha$) and up-set ($\uparrow\alpha$) as well as a strict down-set ($\dot{\downarrow}\alpha$) and strict up-set ($\dot{\uparrow}\alpha$) as shown below:

$$\downarrow\alpha = \{\beta \in \mathbf{L} \mid \beta \preccurlyeq \alpha\} \qquad \text{(down-set)}$$
$$\dot{\downarrow}\alpha = \{\beta \in \mathbf{L} \mid \beta \preccurlyeq \alpha \text{ and } \alpha \not\preccurlyeq \beta\} \qquad \text{(strict down-set)}$$
$$\uparrow\alpha = \{\beta \in \mathbf{L} \mid \alpha \preccurlyeq \beta\} \qquad \text{(up-set)}$$
$$\dot{\uparrow}\alpha = \{\beta \in \mathbf{L} \mid \alpha \preccurlyeq \beta \text{ and } \beta \not\preccurlyeq \alpha\} \qquad \text{(strict up-set)}$$

A *lattice* is a partially ordered set $(\mathbf{L}, \preccurlyeq)$ for which every pair of elements $\{\alpha, b\} \subseteq \mathbf{L}$ has a unique least upper bound $\alpha \curlyvee \beta = \sup\{\alpha, \beta\}$, referred to as *joint*, and a unique greatest lower bound $\alpha \curlywedge \beta = \inf\{\alpha, \beta\}$, referred to as *meet*. This creates an algebra $(\mathbf{L}, \curlywedge, \curlyvee)$ with

the binary operators $\curlywedge$ and $\curlyvee$ that satisfies indempotency, commutativity, associativity and absorption. The consistency relates the ordering relation and algebra with each other. A *distributive lattice* additionally satisfies distributivity.

$$
\begin{array}{lll}
\alpha \curlyvee \alpha = \alpha & \alpha \curlywedge \alpha = \alpha & \text{(indempotency)} \\
\alpha \curlyvee \beta = \beta \curlyvee \alpha & \alpha \curlywedge \beta = \beta \curlywedge \alpha & \text{(commutativity)} \\
\alpha \curlyvee (\beta \curlyvee \gamma) = (\alpha \curlyvee \beta) \curlyvee \gamma & \alpha \curlywedge (\beta \curlywedge \gamma) = (\alpha \curlywedge \beta) \curlywedge \gamma & \text{(associativity)} \\
\alpha \curlyvee (\alpha \curlywedge \beta) = \alpha & \alpha \curlywedge (\alpha \curlyvee \beta) = \alpha & \text{(absorption)} \\
\alpha \preccurlyeq \beta \Rightarrow \alpha \curlyvee \beta = \beta & \alpha \preccurlyeq \beta \Rightarrow \alpha \curlywedge \beta = \alpha & \text{(consistency)} \\
\hline
\alpha \curlyvee (\beta \curlywedge \gamma) = (\alpha \curlyvee \beta) \curlywedge (\alpha \curlyvee \gamma) & \alpha \curlywedge (\beta \curlyvee \gamma) = (\alpha \curlywedge \beta) \curlyvee (\alpha \curlywedge \gamma) & \text{(distributivity)}
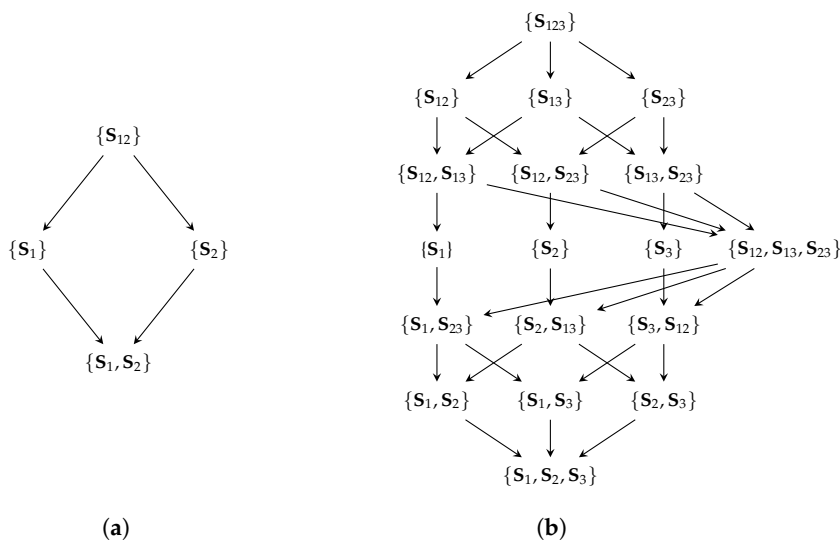\end{array}
$$

2.2. Partial Information Decomposition

This section summarizes Williams and Beer's general approach to partial information decompositions [5]. A more detailed discussion of the literature and required background can be found in [9] (p. 6–20).

Williams and Beer [5] define *sources* $\mathbf{S}_i \in \mathcal{P}_1(\mathbf{V})$ as all combinations of viewable variables ($\mathcal{P}_1(\mathbf{V})$ referring to the power set of $\mathbf{V}$ without the empty set) and use Equation (1a) to construct all distinct interactions between them $\alpha \in \mathcal{A}(\mathbf{V})$, which are referred to as *partial information atoms*. Equation (1b) provides a partial order of atoms to construct a *redundancy lattice* $(\mathcal{A}(\mathbf{V}), \preccurlyeq)$. As a convention, we indicate the visible variables contained in a source by its index, such as $\mathbf{S}_{12} = \{V_1, V_2\}$. The example of the redundancy lattice for two and three visible variables is shown in Figure 2.

$$
\mathcal{A}(\mathbf{V}) = \{\alpha \in \mathcal{P}_1(\mathcal{P}_1(\mathbf{V})) \mid \forall \mathbf{S}_i, \mathbf{S}_j \in \alpha, \mathbf{S}_i \not\subset \mathbf{S}_j\}, \tag{1a}
$$

$$
\forall \alpha, \beta \in \mathcal{A}(\mathbf{V}), (\alpha \preccurlyeq \beta \Leftrightarrow \forall \mathbf{S}_j \in \beta, \exists \mathbf{S}_i \in \alpha \mid \mathbf{S}_i \subseteq \mathbf{S}_j). \tag{1b}
$$

(**a**)　　　　　　　　　　　　　　　　　　　(**b**)

Figure 2. The redundancy lattices for two (**a**) and three (**b**) visible variables. The redundancy lattice specifies the expected inclusion relation between atoms. The following function $I_\cap$ shall measure the shared information for a sets of variables such that the element $\{\mathbf{S}_1, \mathbf{S}_2\}$ represents the shared information between $\mathbf{S}_1$ and $\mathbf{S}_2$ about the target variable T.

A measure of *redundant information* $I_\cap$ shall be defined for this lattice as "[...] cumulative information function which in effect integrates the contribution from each node as one moves up through the nodes of the lattice" [9] (p. 15). Williams and Beer [5] then use the Möbius inverse (Equation (2)) to identify the *partial information* $I_\delta(\alpha; T)$ as the contribution of atom $\alpha \in \mathcal{A}(\mathbf{V})$ and therefore the desired unique/redundant/synergetic component.

A PID is said to be *non-negative* if the resulting partial contributions are guaranteed to be non-negative.

$$I_\delta(\alpha; T) = I_\cap(\alpha; T) - \sum_{\beta \in \downarrow \alpha} I_\delta(\beta; T). \tag{2}$$

Williams and Beer [5] highlight three axioms that a measure of redundancy should satisfy.

Axiom 1 (Commutativity). *Invariant to the order of sources (σ permuting the order of indices):*

$$I_\cap(\mathbf{S}_1, \ldots, \mathbf{S}_i; T) = I_\cap(\mathbf{S}_{\sigma(1)}, \ldots, \mathbf{S}_{\sigma(i)}; T)$$

Axiom 2 (Monotonicity). *Additional sources can only decrease redundant information:*

$$I_\cap(\mathbf{S}_1, \ldots, \mathbf{S}_{i-1}; T) \geq I_\cap(\mathbf{S}_1, \ldots, \mathbf{S}_i; T)$$

Axiom 3 (Self-redundancy). *For a single source, redundancy equals mutual information:*

$$I_\cap(\mathbf{S}_i; T) = I(\mathbf{S}_i; T)$$

Finally, Williams and Beer [5] proposed $I_{\min}$ (Equation (3)) as a measure of redundancy and demonstrated that it satisfies the required axioms.

$$I(\mathbf{S}_i; T = t) = \sum_{s \in \mathcal{S}_i} p(s \mid t) \left[\log \frac{1}{p(t)} - \log \frac{1}{p(t \mid s)} \right] \tag{3a}$$

$$I_{\min}(\mathbf{S}_1, \ldots, \mathbf{S}_k; T) = \sum_{t \in \mathcal{T}} p(t) \min_{i \in 1..k} I(\mathbf{S}_i; T = t). \tag{3b}$$

However, the measure has been criticized for not distinguishing "the *same* information and the *same amount* of information" [6–9] due to its use of a pointwise minimum (for each $t \in \mathcal{T}$) over the sources.

2.3. Quantifying Unique Information

A non-negative decomposition for the case of two viewable variables $\mathbf{V} = \{V_1, V_2\}$ was proposed by Bertschinger et al. [10] (defining unique information) as well as an equivalent decomposition by Griffith and Koch [11] (defining union information) as shown in Equation (4) (modified notation). The function $\vartheta(V_1, V_2; T)$ acts as an information measure of the union for V_1 and V_2 (the minimal information that any two variables with the same marginal distributions can achieve), which is then used to compute the partial contributions using an inclusion–exclusion principle. Bertschinger et al. [10] motivated the decomposition from the operational interpretation that if a variable provides unique information, there must be a way to utilize this information in a decision problem for some reward function. Additionally, they argue that unique information should only depend on the marginal distributions $P_{(T,V_1)}$ and $P_{(T,V_2)}$.

$$
\begin{aligned}
\vartheta(V_1, V_2; T) &= \min \ I(F; G_1, G_2) \ \text{s.t.} \ P_{(F,G_1)} = P_{(T,V_1)} \ \text{and} \ P_{(F,G_2)} = P_{(T,V_2)} \\
S(V_1, V_2; T) &= I(V_1; T) + I(V_2; T) - \vartheta(V_1, V_2; T) && \text{(Shared)} \\
U(V_1; T) &= \vartheta(V_1, V_2; T) - I(V_1; T) && \text{(Unique)} \\
C(V_1, V_2; T) &= I(T; V_1, V_2) - \vartheta(V_1, V_2; T) && \text{(Complementing)}
\end{aligned} \tag{4}
$$

We highlight this decomposition since our approach can be interpreted as its pointwise extension (see Section 4.1).

2.4. Blackwell Order

A channel κ can be represented as a (row) stochastic matrix wherein each element is non-negative and all rows sum to one (see Figure 3). In this work, we consider the

sources to be the indirect observation of the target variable through a channel $T \xrightarrow{\kappa_i} \mathbf{S}_i$ while taking the joint distribution of the visible variables within $\mathbf{S}_i$. As a result, κ_i is obtained from the conditional probability distribution $\kappa_i = T \to \mathbf{S}_i = P_{(\mathbf{S}_i|T)}$. As for sources, we list the contained visible variables as an index such that $i = 12$ corresponds to $\kappa_{12} = P_{(\mathbf{S}_{12}|T)} = P_{(V_1,V_2|T)}$.
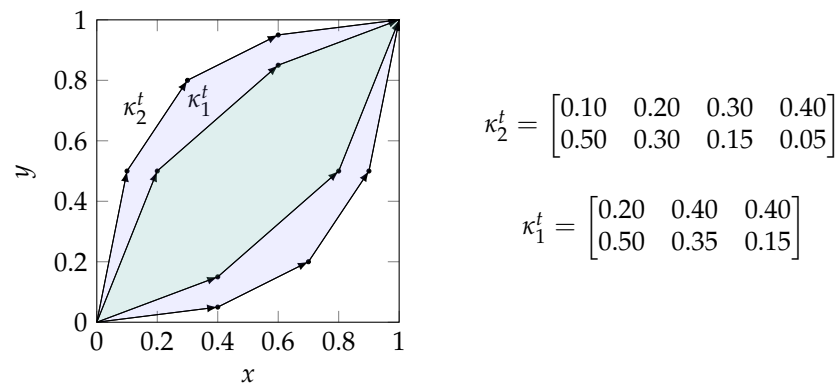
The Blackwell order $\kappa_1 \sqsubseteq \kappa_2$ is a preorder of channels, as shown in Equation (5) [14]. It highlights that a channel equivalent to κ_1 can be obtained by garbling the output of κ_2 (a chaining of channels as seen in Equation (5)). Therefore, there exists a decision strategy based on κ_2 for any reward function that performs at least as well as all strategies based on κ_1 [12].

$$\kappa_1 \sqsubseteq \kappa_2 \iff \kappa_1 = \kappa_2 \cdot \lambda \qquad \text{for some channel } \lambda \tag{5}$$

Bertschinger and Rauh [12] showed that the Blackwell order does not define a lattice in general since it does not provide a unique meet and joint element beyond binary inputs. However, binary input channels provide a special case for which the Blackwell order is equivalent to the *zonotope order* and defines a lattice. We use the notation κ^t to indicate that a channel has a binary input ($|\mathcal{T}| = 2$) or corresponds to the one-vs-rest encoding for one state t if $|\mathcal{T}| > 2$. In this case, the row stochastic matrix representing a channel contains a set of vectors $\vec{v}_s$ as shown in Equation (6). A zonotope Z_{κ^t} (Equation (6b)) corresponds to "the image of the unit cube [...] under the linear map corresponding to $[\kappa^t]$" [12] (p. 2), and the resulting zonotope order $\kappa_1^t \sqsubseteq \kappa_2^t \Leftrightarrow Z_{\kappa_1^t} \subseteq Z_{\kappa_2^t}$ is a preorder that is identical to the Blackwell order in the special case of binary input channels [12] as visualized in Figure 3. In the resulting lattice, the joint of two channels can be obtained as the convex hull $Z_{\kappa_1^t \sqcup \kappa_2^t}$ of the zonotopes $Z_{\kappa_1^t}$ and $Z_{\kappa_2^t}$, and the meet element $Z_{\kappa_1^t \sqcap \kappa_2^t}$ corresponds to their intersection.

$$\kappa_i^t = \begin{bmatrix} p(\mathbf{S}_i = s_1 \mid T = t) & p(\mathbf{S}_i = s_2 \mid T = t) & \dots & p(\mathbf{S}_i = s_n \mid T = t) \\ p(\mathbf{S}_i = s_1 \mid T \neq t) & p(\mathbf{S}_i = s_2 \mid T \neq t) & \dots & p(\mathbf{S}_i = s_n \mid T \neq t) \end{bmatrix} \tag{6a}$$

$$Z_{\kappa_i^t} = \left\{ \sum_{s \in \mathcal{S}_i} x_s \cdot \vec{v}_s \mid 0 \leq x_s \leq 1 \right\} \qquad \text{where} \quad \vec{v}_s = \begin{pmatrix} p(\mathbf{S}_i = s \mid T = t) \\ p(\mathbf{S}_i = s \mid T \neq t) \end{pmatrix} \tag{6b}$$

$$\kappa_2^t = \begin{bmatrix} 0.10 & 0.20 & 0.30 & 0.40 \\ 0.50 & 0.30 & 0.15 & 0.05 \end{bmatrix}$$

$$\kappa_1^t = \begin{bmatrix} 0.20 & 0.40 & 0.40 \\ 0.50 & 0.35 & 0.15 \end{bmatrix}$$

Figure 3. Visualization of the zonotope order for binary input channels. The channel κ_1^t is Blackwell inferior to κ_2^t ($\kappa_1^t \sqsubseteq \kappa_2^t$) since the corresponding zonotope $Z_{\kappa_1^t}$ (**green**) is a subset of $Z_{\kappa_2^t}$ (**purple**). As a result, the meet and joint elements of this example are: $\kappa_1^t \sqcap \kappa_2^t = \kappa_1^t$ and $\kappa_1^t \sqcup \kappa_2^t = \kappa_2^t$.

2.5. Receiver Operating Characteristic Curves

While any classification system can be represented as channel, this section focuses on binary decision problems or the one-vs-rest encoding of others ($T^t = t \Leftrightarrow T = t$). The binary label $t \in \mathcal{T}^t$ is used to obtain a sample $s \in \mathcal{S}_i$, which is processed by a classification system C to its output $o \in \mathcal{O}$ with $o = C(s)$, and applying a decision strategy d shall result in an approximation of the label $\hat{t} \in \mathcal{T}^t$ with $\hat{t} = d(o)$. This forms the Markov chain: $T^t \to \mathbf{S}_i \to O \to \hat{T}^t$. A common method of analyzing binary decision/classification systems is the Receiver Operating Characteristic (ROC). A ROC plot typically represents a

classifier C with a continuous, discrete or categorical output range (by assigning distinct arbitrary values to each category) for a binary decision problem by a curve in a True-Positive Rate (TPR)/False-Positive Rate (FPR) diagram for varying decision thresholds τ_x with the decision rule for a sample s being $C(s) \leq \tau_x \Leftrightarrow$ False [15]. The resulting points are typically connected using a step function, as shown in red in Figure 4a. As a result of using a single decision threshold, the points of the ROC curve monotonically increase from $(0,0)$ to $(1,1)$; however, they are in general neither concave nor convex [16].

(**a**) (**b**)
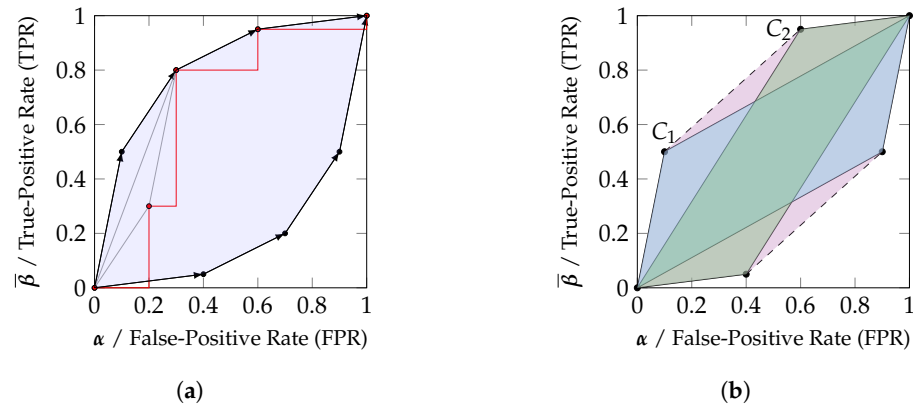
Figure 4. Relating zonotopes and their convex hull to achievable decision regions. (**a**) A ROC curve (red) can be used to estimate the parameters of a channel, and the randomized combination of thresholds (Equation (7)) corresponds to an interpolation in the visualization (gray). The reachable decision region when utilizing all thresholds can be constructed using a likelihood ratio test, which corresponds to reordering the vectors by decreasing slope (in this case, swapping the first two steps) and taking the convex hull of reachable points. This reachable decision region is the zonotope of the channel. (**b**) The convex hull of any set of zonotopes is reachable by their randomized combination. Given two classifiers C_1 (blue) and C_2 (green), there always exists a randomized combination that can reach any position in their convex hull (purple).

We want to highlight the distinction between a ROC curve and the achievable performance pairs (TPR, FPR) based on the classifier. Any performance pair within the convex hull of the obtained points for constructing the ROC curve can be achieved since the decision strategy of Equation (7) results in an interpolation of the points corresponding to $\tau_1 \leq \tau_2$ with the parameter $0 \leq h \leq 1$ in the TPR/FPR diagram. Therefore, while a ROC curve *is not* convex in general, the achievable performance region *is* convex in general.

$$\begin{aligned} C(s) \leq \tau_1 &\Rightarrow \text{False,} \\ \tau_1 < C(s) \leq \tau_2 &\Rightarrow \text{Bernoulli}(h), \\ \tau_2 < C(s) &\Rightarrow \text{True.} \end{aligned} \tag{7}$$

When utilizing the set of all available thresholds on the classification output, we can identify the reachable decision regions within the TPR/FPR diagram using the likelihood ratio test, which is well known to be optimal for binary decision problems: Neyman–Pearson theory [17] states that the likelihood ratio test (Equation (8)) provides the minimal type II error (minimal β, maximal TPR$= \overline{\beta} = 1 - \beta$) for a bounded type I error (FPR, α).

$$\begin{aligned} \frac{P(\mathbf{S}_i = s | T = t)}{P(\mathbf{S}_i = s | T \neq t)} < \tau &\Rightarrow \text{False,} \\ \frac{P(\mathbf{S}_i = s | T = t)}{P(\mathbf{S}_i = s | T \neq t)} = \tau &\Rightarrow \text{Bernoulli}(h), \\ \frac{P(\mathbf{S}_i = s | T = t)}{P(\mathbf{S}_i = s | T \neq t)} > \tau &\Rightarrow \text{True.} \end{aligned} \tag{8}$$

Notice that the decision criterion is determined by the slope of each vector in the row stochastic matrix that represents the binary input channel (Equation (6a)). This effective

reordering of vectors based on their slope when varying the parameters τ and h results in the upper half of the zonotope discussed in Section 2.4 and as visualized in Figure 4a. The lower half of the zonotope is obtained from negating the outcome of the likelihood ratio test. Therefore, the zonotope representation of a channel corresponds to the achievable performance region in a TPR/FPR diagram of a classifier at binary decision problems. When reconsidering Figure 3, the channels $\kappa_1^t = T^t \to O_1$ and $\kappa_2^t = T^t \to O_2$ may correspond to two classifiers C_1 and C_2 whose channel parameters have been estimated from a ROC curve, and the achievable performance regions correspond to the zonotopes in a TPR/FPR diagram. Since the likelihood ratio test is optimal for binary decision problems, there cannot exist a decision strategy that would achieve a performance outside the zonotope. At the same time, the likelihood ratio test can be randomized to reach any desired position within the zonotope.

Finally, notice that the convex hull of any two classification systems is reachable by their randomized combination. We can view each classifier as an observation from a channel κ_1^t/κ_2^t about T^t and know that there always exists a garbling λ of the joint channel κ_{12}^t to obtain their convex hull $\kappa_1^t \sqcup \kappa_2^t = \kappa_{12}^t \lambda$. Using a likelihood ratio test on $\kappa_1^t \sqcup \kappa_2^t$, any position within the convex hull is reachable as a randomized combination of both classifiers. This has been visualized in Figure 4b. Due to this reason, we will say in Section 3.1 that the convex hull should be fully attributed to the marginal channels κ_1^t and κ_2^t.

2.6. Lattice Valuations

This section summarizes the properties of consistent lattice valuations based on Knuth [13]. The *quantification* of a lattice $(\mathbf{L}, \preccurlyeq)$ or $(\mathbf{L}, \curlywedge, \curlyvee)$ with $\alpha \curlywedge \beta = \alpha \Leftrightarrow \alpha \preccurlyeq \beta$ for elements of the set $\alpha, \beta \in \mathbf{L}$ is a function $q : \mathbf{L} \to \mathbb{R}$, which assigns reals to each element. A quantification is called a *valuation* if any two elements maintain an ordering relation: $\alpha \preccurlyeq \beta$ implies that $q(\alpha) \leq q(\beta)$. A quantification q is *consistent* if it satisfies a sum rule (inclusion–exclusion principle): $q(\alpha \curlyvee \beta) = q(\alpha) + q(\beta) - q(a \curlywedge \beta)$. If the bottom element of the lattice ($\perp$) is evaluated to zero $q(\perp) = 0$, then the valuation of the Cartesian product of two lattices $q((\alpha; \beta)) = q(\alpha) \cdot q(\beta)$ remains consistent with the individual lattices. Finally, a *bi-quantification* can be defined as $b(\alpha, \beta) = q(\alpha \curlywedge \beta)/q(\beta)$. Similar to Knuth [13], we will use the notation $q([\alpha; \beta]) \equiv b(\alpha, \beta)$ which can be thought of as quantifying a degree of inclusion for α within β. The distributive lattice then creates an algebra like probability theory for the consistent valuation, as summarized in Equation (9) [13].

$$
\begin{aligned}
q(\alpha \curlyvee \beta) &= q(\alpha) + q(\beta) - q(\alpha \curlywedge \beta) & \text{(Sum rule)} \\
q([\alpha \curlyvee \beta; \gamma]) &= q([\alpha; \gamma]) + q([\beta; \gamma]) - q([\alpha \curlywedge \beta; \gamma]) & \text{(Sum rule)} \\
q((\alpha; \beta)) &= q(\alpha) \cdot q(\beta) & \text{(Direct product rule)} \\
q(([\alpha; \beta]; [\tau; \upsilon])) &= q([\alpha; \tau]) \cdot q([\beta; \upsilon]) & \text{(Direct product rule)} \\
q([\beta \curlywedge \gamma; \alpha]) &= q([\gamma; \alpha \curlywedge \beta]) \cdot q([\beta; \alpha]) & \text{(Product rule)} \\
q([\gamma; \alpha \curlywedge \beta]) &= \frac{q([\beta; \alpha \curlywedge \gamma]) \cdot q([\gamma; \alpha])}{q([\beta; \alpha])} & \text{(Bayes' Theorem)}
\end{aligned} \tag{9}
$$

3. Quantifying Reachable Decision Regions

We start by studying the decomposition of binary decision problems from an interpretational perspective (Section 3.1). This provides the basis for constructing a distributive lattice in Section 3.2 and demonstrating the structure of a consistent valuation function. Section 3.3 highlights that mutual information is such a consistent valuation and extends the concept from binary decision problems to target variables with an arbitrary finite number of states. The resulting definition of shared information for the PID will be discussed as an application in Section 4.1 together with the tracing of information flows in Section 4.2.

We define an equivalence relation ($\sim$) for binary input channels κ^t, which allows for the removal of zero vectors, the permutation of columns (**P** representing a permutation matrix) and the splitting/merging of columns with identical likelihood ratios (vectors of

identical slope, $\ell \in \mathbb{R}$), as shown in Equation (10). These operations are invertible using garblings and do not affect the underlying zonotope.

$$\kappa^t \sim \begin{bmatrix} \kappa^t & \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{bmatrix}; \tag{10a}$$

$$\kappa^t \sim \kappa^t \mathbf{P}; \tag{10b}$$

$$\begin{bmatrix} (1+\ell)\vec{v}_1 & \vec{v}_2 & \dots \end{bmatrix} \sim \begin{bmatrix} \vec{v}_1 & \ell\vec{v}_1 & \vec{v}_2 & \dots \end{bmatrix}. \tag{10c}$$

Based on this definition, block matrices cancel at an inverted sign ($\ell = -1$) if we allow negative columns, as shown in Equation (11), where $\mathbf{M}_1$ and $\mathbf{M}_2$ are some $2 \times n$ matrix.

$$\mathbf{M}_1 \sim \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 & -\mathbf{M}_2 \end{bmatrix} \tag{11}$$

*3.1. Motivation and Operational Interpretation*

The aim of this section is to provide a first intuition based on a visual example for the methodology that will be used in Section 3.2 to construct a distributive lattice of the reachable decision regions and its consistent valuation. We only consider binary variables $\mathcal{T}^t = \{t, \bar{t}\}$ or the one-vs-rest encoding of others ($T^t = t \Leftrightarrow T = t$).

In the used example, the desired variable can be observed indirectly using the two variables V_1 and V_2. The visible variables are considered to be the output of the channels $T^t \xrightarrow{\kappa_1^t} V_1$, $T^t \xrightarrow{\kappa_2^t} V_2$ and $T^t \xrightarrow{\kappa_{12}^t} (V_1, V_2)$ and correspond to the zonotopes shown in Figure 5. We consider each reachable decision point (a pair of TPR and FPR) to represent a different notion of uncertainty about the state of the target variable. We want to attribute the reachable decision regions to each channel for constructing a lattice, as shown in Figure 6, with the following operational interpretation:

- *Synergy:* Corresponds to the partial contribution of $\kappa_{12}^t = T^t \to (V_1, V_2)$ and represents the decision region which is only accessible due to the (in-)dependence of both variables.
- *Joint:* The joint element $\kappa_1^t \vee \kappa_2^t = (T^t \to V_1) \vee (T^t \to V_2)$ corresponds to the joint under the Blackwell order and represents the decision region which is always accessible if the marginal distributions (V_1, T^t) and (V_2, T^t) can be obtained. Therefore, we say that its information shall be fully attributed to V_1 and V_2 such that is has no partial contribution. For binary target variables, this definition is equivalent to the notion of union information by Bertschinger et al. [10] and Griffith and Koch [11]. However, we extend the analysis beyond binary target variables with a different approach in Section 3.3.
- *Unique:* Corresponds to the partial contribution of $\kappa_1^t = T^t \to V_1$ or $\kappa_2^t = T^t \to V_2$ and represents the decision region that is lost when losing the variable. It only depends on their marginal distributions (V_1, T^t) and (V_2, T^t).
- *Shared:* Corresponds to the cumulative contribution of $\kappa_1^t \wedge \kappa_2^t = (T^t \to V_1) \wedge (T^t \to V_2)$ and represents the decision region which is lost when losing either V_1 or V_2. Since it only depends on the marginal distributions, we interpret it as being part of both variables. The shared decision region can be split in two components: the decision region that is part of both individual variables and the component that is part of the convex hull but neither individual one. The latter component only exists if both variables provide unique information.
- *Redundant:* The largest decision region $\kappa_1^t \sqcap \kappa_2^t = (T^t \to V_1) \sqcap (T^t \to V_2)$ which can be accessed from both V_1 and V_2. It corresponds to the meet under the Blackwell order and the part of shared information that can be represented by some random variable (pointwise extractable component of shared information). The redundant and shared regions are equal unless both variables provide some unique information.
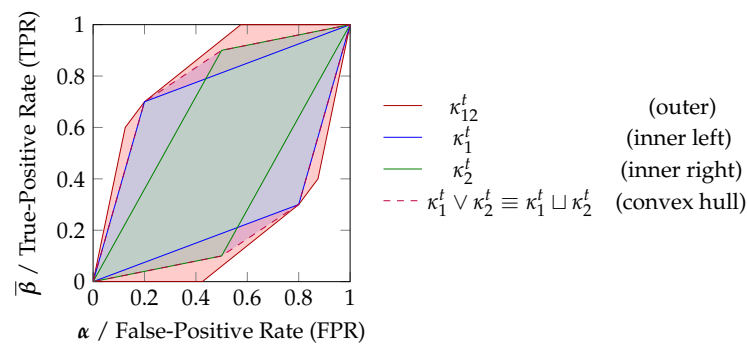
Figure 5. Relating the zonotope representations to TPR/FPR plots. The zonotopes correspond to the regions of a TPR/FPR plot that are reachable by some decision strategy. Regions outside of the zonotopes are known to be unreachable since the likelihood ratio test is optimal for binary decision problems. The convex hull of both zonotopes $\kappa_1^t \vee \kappa_2^t$ is the (unique) lower bound of any joint distribution under the Blackwell order.

Due to the invariance of re-ordering columns under the defined equivalence relation, κ^t represents a set of likelihood vectors. All cumulative and partial decision regions of Figure 6 can be constructed using a convex hull operator (joint) and matrix concatenations under the defined equivalence relation ($\sim$). For example, the shared decision region (meet) can be expressed through an inclusion–exclusion principle with the joint operator $\kappa_1^t \wedge \kappa_2^t \sim \begin{bmatrix} \kappa_1^t & \kappa_2^t & -\kappa_1^t \vee \kappa_2^t \end{bmatrix}$. This operator is not closed on channels since it introduces negative likelihood vectors. Therefore, we distinguish the notation between channels (κ^t) and atoms (α^t). These matrices α^t sum to one similar to channels but may contain negative columns. Their partial contributions $\alpha^{\delta t}$ sum to zero.

- The unique contribution of V_2: $\alpha^{\delta t} \sim \begin{bmatrix} (\kappa_1^t \vee \kappa_2^t) & -\kappa_1^t \end{bmatrix}$
- The shared cumulative region of V_1 and V_2: $\beta^t \sim \begin{bmatrix} \kappa_1^t & \kappa_2^t & -(\kappa_1^t \vee \kappa_2^t) \end{bmatrix} \sim \kappa_1^t \wedge \kappa_2^t$
- The shared partial contribution: $\beta^{\delta t} \sim \begin{bmatrix} \beta^t & -(\kappa_1^t \sqcap \kappa_2^t) \end{bmatrix}$
- Each cumulative region corresponds to the combination of partial contributions in its down-set. Notice that the partial contribution of the shared region is canceled by a section of each unique contribution due to an opposing sign:

$$\kappa_1^t \sim \begin{bmatrix} \alpha^{\delta t} & \beta^{\delta t} & (\kappa_1^t \sqcap \kappa_2^t) \end{bmatrix}$$

In Section 3.3, we demonstrate a valuation function f that can quantify all cumulative and partial atoms of this lattice while ensuring their non-negativity and consistency with the defined equivalence relation ($\sim$). We will refer to a more detailed example on the valuation of partial decision regions in Appendix C in the context of the following section.

Why does the decomposition of reachable decision regions as shown in Figure 6 provide a meaningful operational interpretation? Because combining the partial contributions of the up-set for a variable results in the decision region that becomes inaccessible when the variable is lost, while combining the partial contributions of the down-set results in the decision region that is accessible through the variable. For example, losing access to variable V_2 results in losing access to the decision regions provided uniquely by V_2 and its synergy with V_1 (the up-set on the lattice). Additionally, the cumulative component corresponds to the combination of all partial contributions in its down-set since opposing vectors cancel under the defined equivalence relation ($\sim$) such as the shared and unique contributions. Therefore, we define a consistent valuation of this lattice in Section 3.2 by quantifying decision regions based on their spanning vectors and highlight that the expected value for each $t \in \mathcal{T}$ corresponds to the definition of mutual information.

Sections 3.2 and 3.3 focus only on defining the meet and joint operators ($\wedge/\vee$) with their consistent valuation. To obtain the pointwise redundant and synergetic components for a PID, we can later add the corresponding channels when constructing the pointwise lattices $\mathbf{V} = \{V_1, V_2, (V_1, V_2), V_1 \sqcap V_2\}$ with the ordering of Figure 6 from the meet and joint operators.
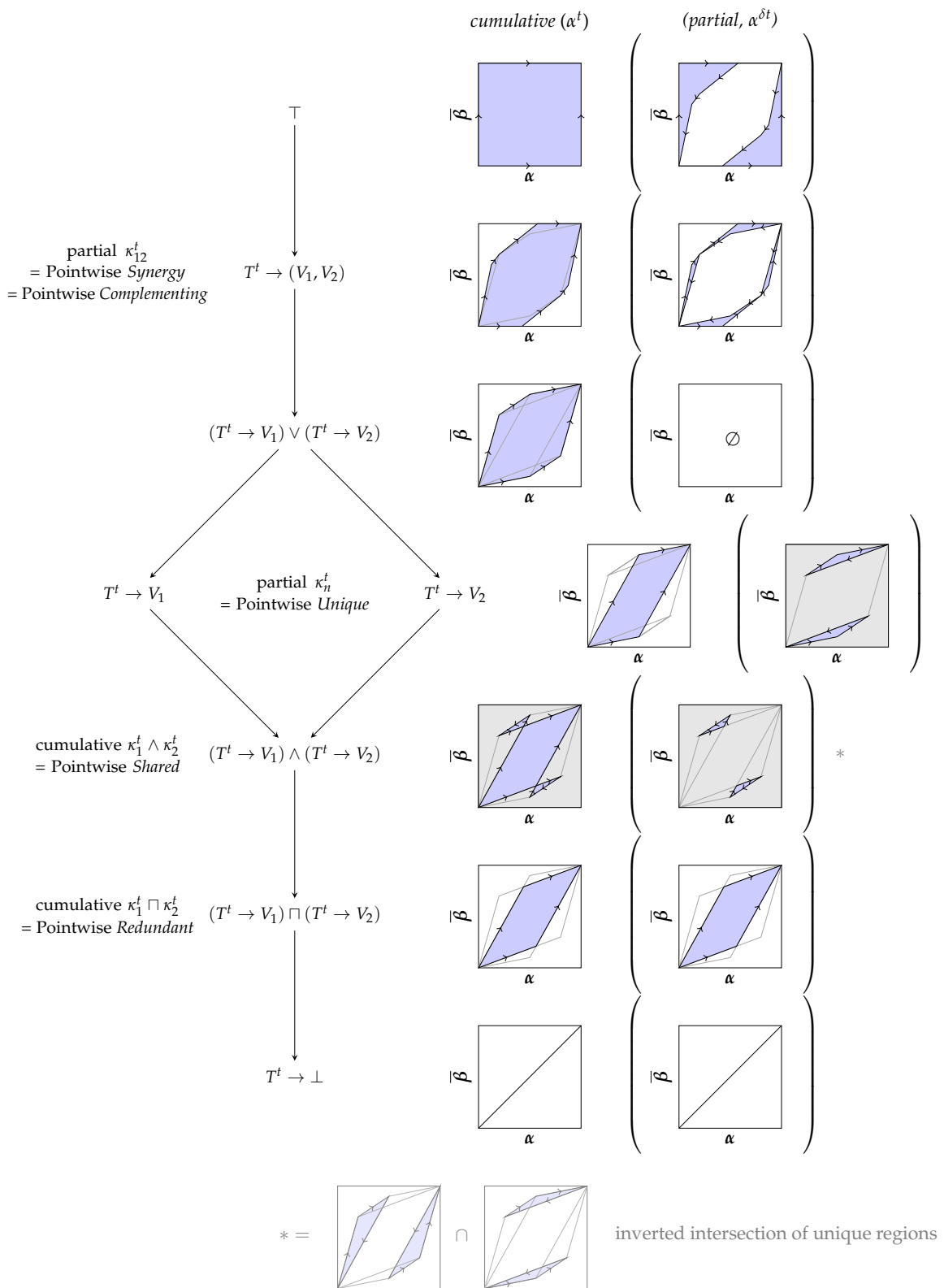
Figure 6. Decomposing the achievable decision regions for binary decision problems from an operational perspective. Each node is visualized by its cumulative and partial decision region. The partial decision region is shown within round brackets. The cumulative region corresponds to the matrix concatenation of the partial regions in its down-set under the defined equivalence relation. Three key elements are highlighted using a grey background.

3.2. Decomposition Lattice and Its Valuation

This section first defines the meet and joint operators ($\wedge$, $\vee$) and then constructs a consistent valuation for the resulting distributive lattice. For constructing a pointwise channel lattice based on the redundancy lattice, we notate the map of functions as shown in Equation (12) and consider the function $k^t(\mathbf{S}_i) = T^t \rightarrow \mathbf{S}_i = \kappa_i^t$ to obtain the pointwise channel κ_i^t of a source $\mathbf{S}_i$.

$$
\begin{aligned}
f\langle \mathbf{P} \rangle &= \{f(x) \mid x \in \mathbf{P}\}, \\
f\langle\langle \mathbf{P} \rangle\rangle &= \{f\langle \mathbf{x} \rangle \mid \mathbf{x} \in \mathbf{P}\}, \\
f\langle\langle\langle \mathbf{P} \rangle\rangle\rangle &= \{f\langle\langle \mathbf{x} \rangle\rangle \mid \mathbf{x} \in \mathbf{P}\}.
\end{aligned}
\tag{12}
$$

The intersections shall correspond to some meet operation and the union to some joint operation on the pointwise channels, as shown in Equation (13), while maintaining the ordering relation of Williams and Beer [5]. This section aims to define suitable meet and joint operations together with a function for their consistent valuation. Each atom $\alpha^t, \beta^t \in \mathcal{B}^t(\mathbf{V})$ now represents an expression of channels κ^t with the operators $\vee/\wedge$, as shown in Appendix A. For example, the element $\{\mathbf{S}_{12}, \mathbf{S}_3\}$ is converted to the expression $(\kappa_1^t \vee \kappa_2^t) \wedge \kappa_3^t$.

$$
\mathcal{B}^t(\mathbf{V}) = \bigwedge \langle \; \bigvee \langle\langle \, k^t \langle\langle\langle \mathcal{A}(\mathbf{V}) \rangle\rangle\rangle \, \rangle\rangle \; \rangle.
\tag{13}
$$

As seen in Section 3.1, we want to define the joint for a set of channels to be equivalent to their convex hull, matching the Blackwell order. This also ensures that the joint operation is closed on channels.

$$
\kappa_1^t \vee \kappa_2^t \equiv \kappa_1^t \sqcup \kappa_2^t \qquad \text{(joint is closed on channels)}
\tag{14}
$$

Since opposing vectors cancel under the defined equivalence relation, we can use a notion of the Möbius inverse to define the set of vectors spanning a partial decision region $\alpha^{\delta t}$ for an atom $\alpha^t \in \mathcal{B}^t(\mathbf{V})$, as shown in Equation (15), written as a recursive block matrix and using the strict down-set of the ordering based on the underlying redundancy lattice.

$$
\alpha^{\delta t} \equiv \begin{bmatrix} \alpha^t & -[\beta^{\delta t} \mid \beta^t \in \dot{\downarrow}\alpha^t] \end{bmatrix}
\tag{15}
$$

The definition of the meet operator ($\wedge$) and the extension of the joint operator ($\vee$) from channels to atoms is now obtained from the constraint that the partial contribution for the joint of two incomparable atoms ($\alpha^t, \beta^t \in \mathcal{B}^t(\mathbf{V})$, $\alpha^t \vee \beta^t \not\succ \alpha^t$ and $\alpha^t \vee \beta^t \not\succ \beta^t$) shall be zero, as shown in Equation (16).

$$
\alpha^t \vee \beta^t \not\succ \alpha^t \text{ and } \alpha^t \vee \beta^t \not\succ \beta^t \quad \Rightarrow \quad (\alpha^t \vee \beta^t)^{\delta t} \equiv \begin{bmatrix} 0 \\ 0 \end{bmatrix}
\tag{16}
$$

This creates the desired inclusion–exclusion principle and results in the equivalences of the meet for two and three atoms, as shown in Equation (17). Their resulting partial channels ($\alpha^{\delta t}$) correspond to the set of vectors spanning the desired unique and shared decision regions of Figure 6.

$$
\alpha^t \wedge \beta^t \quad \sim \begin{bmatrix} \alpha^t & \beta^t & -\alpha^t \vee \beta^t \end{bmatrix}
\tag{17a}
$$

$$
\alpha^t \wedge (\beta^t \wedge \gamma^t) \sim \begin{bmatrix} \alpha^t & \beta^t & \gamma^t & -\alpha^t \vee \beta^t & -\alpha^t \vee \gamma^t & -\beta^t \vee \gamma^t & \alpha^t \vee \beta^t \vee \gamma^t \end{bmatrix}
\tag{17b}
$$

From their construction, the meet and joint operators provide a distributive lattice for a set of channels under the defined equivalence relation as shown in Appendix B by satisfying idempotency, commutativity, associativity, absorption and distributivity. This can be used to define a corresponding ordering relation (Equation (18)).

$$
\alpha^t \preceq \beta^t \equiv \alpha^t \wedge \beta^t \sim \alpha^t \Leftrightarrow \alpha^t \vee \beta^t \sim \beta^t
\tag{18}
$$

To obtain a consistent valuation of this lattice, we consider a function $f(\alpha^t)$, as shown in Equation (19). First, this function has to be invariant under the defined equivalence relation, and second, it has to match the ordering of the constructed lattice.

$$f(\alpha^t) = \sum_{\vec{v} \in \alpha^t} r(\vec{v}) \quad \text{where } r \text{ is } convex \text{ and satisfies} \quad r(\ell\vec{v}) = \ell r(\vec{v}) \text{ and } r\left(\left[\begin{smallmatrix} \ell \\ \ell \end{smallmatrix}\right]\right) = 0 \quad (19)$$

The function f shall apply a (convex) function $r(\vec{v})$ to each vector of the matrix of an atom $\vec{v} \in \alpha^t$. The function is invariant under the equivalence relation ($\sim$, Equation (10)):

- Zero vectors do not affect the quantification: $r\left(\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]\right) = 0$
- The structure of f ensures invariance under reordering columns: $f(\kappa^t) = f(\kappa^t \mathbf{P})$
- The property $r(\ell\vec{v}) = \ell r(\vec{v})$ with $\ell \in \mathbb{R}$ ensures invariance under splitting/merging columns of identical likelihood ratios:
  $f([(1 + \ell)\vec{v}_1]) = (1 + \ell)r(\vec{v}_1) = r(\vec{v}_1) + \ell r(\vec{v}_1) = f([\vec{v}_1 \quad \ell\vec{v}_1])$

The function f is a consistent valuation of the ordering relation ($\preceq$, Equation (18)) from the constructed lattice:

- The convexity of r ensures that the quantification $f(\alpha^t)$ is a valuation as shown in Appendix C: $\beta^t \preceq \alpha^t \Rightarrow f(\beta^t) \leq f(\alpha^t)$
- The function f provides a sum-rule: $f(\alpha^t \wedge \beta^t) = f([\begin{smallmatrix} \alpha^t & \beta^t & -\alpha^t \vee \beta^t \end{smallmatrix}]) = f(\alpha^t) + f(\beta^t) - f(\alpha^t \vee \beta^t)$
- The function f quantifies the bottom element correctly: $f(\bot) = r\left(\left[\begin{smallmatrix} \ell \\ \ell \end{smallmatrix}\right]\right) = 0$

A parameterized function that forms a consistent lattice valuation with $0 \leq p \leq 1$ and that will be used in Section 3.3 is shown in Equation (20) (the convexity of r_p is shown in Appendix D).

$$f_p(\alpha^t) = \sum_{\vec{v} \in \alpha^t} r_p(\vec{v}) \tag{20a}$$

$$r_p(\vec{v}) = r_p\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = x \log\left(\frac{x}{px + (1-p)y}\right) \tag{20b}$$

This section demonstrated the construction of a distributive lattice and its consistent valuation, resulting in an algebra as shown in Equation (9).

3.3. Decomposing Mutual Information

This section demonstrates that mutual information is the expected value of a consistent valuation for the constructed pointwise lattices and discusses the resulting algebra. To show this, we define the parameter p and pointwise channel κ_i^t for the consistent valuation (Equation (20)) using a one-vs-rest encoding (Equation (21)).

$$p = P_{(T=t)} \quad \text{(parameter)}$$

$$\kappa_i^t = \begin{bmatrix} P_{(\mathbf{S}_i | T=t)} \\ P_{(\mathbf{S}_i | T \neq t)} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_m \\ y_1 & y_2 & \dots & y_m \end{bmatrix} \quad \text{(binary input channel)} \tag{21}$$

The expected value of the resulting valuation in Equation (20) is equivalent to the definition of mutual information, as shown in Equation (22). Therefore, we can interpret mutual information as being the expected value of quantifying the reachable decision regions for each state of the target variable that represent a concept of pointwise uncertainty.

$$I(T; \mathbf{S}_i) = \sum_{s \in \mathcal{S}_i} \sum_{t \in \mathcal{T}} P_{(\mathbf{S}_i, T)}(s, t) \log\left(\frac{P_{(\mathbf{S}_i, T)}(s, t)}{P_{\mathbf{S}_i}(s) P_T(t)}\right) \tag{22a}$$

$$= \mathbb{E}_T\left[\sum_{s \in \mathcal{S}_i} \underbrace{P_{(\mathbf{S}_i|T=t)}(s)}_{x_j} \log\left(\frac{\overbrace{P_{(\mathbf{S}_i|T=t)}(s)}^{x_j}}{\underbrace{P_{(T=t)}}_{p} \underbrace{P_{(\mathbf{S}_i|T=t)}(s)}_{x_j} + \underbrace{(1 - P_{(T=t)})}_{1-p} \underbrace{P_{(\mathbf{S}_i|T \neq t)}(s)}_{y_j}}\right)\right] \tag{22b}$$

The expected value for a set of consistent lattice valuations corresponds to a weighted sum such that the resulting lattice remains consistent. Therefore, we can combine the pointwise lattices to extend the definition of mutual information for meet and joint elements, which we will think of as intersections and unions. Let α represent an expression of sources with the operators $\vee$ and $\wedge$. Then, we can obtain its valuation from the pointwise lattices using the function $\hat{I}_T$, as shown in Equation (23). Notice that we do not define the operators for random variables but only use the notation for selecting the corresponding element on the underlying pointwise lattices. For example, we write $\alpha = (\mathbf{S}_{12} \wedge \mathbf{S}_3) \vee \mathbf{S}_4$ to refer to the pointwise atom $\alpha^t = (\kappa_{12}^t \wedge \kappa_3^t) \vee \kappa_4^t$ on each pointwise lattice.

The special case of atoms that consist of a single source corresponds by construction to the definition of mutual information. However, we propose normalizing the measure, as shown in Equation (23), to capture a degree of inclusion between zero and one. This is possible for discrete variables and will lead to an easier intuition for the later definition of bi-valuations and product spaces by ensuring the same output range for these measures. As a possible interpretation for the special role of the target variable, we like to think of T as the considered origin of information within the system, which then propagates through channels to other variables.

$$\hat{I}_T(\alpha) \equiv \frac{\mathbb{E}_T\left[f_{P_T(t)}(\alpha^t)\right]}{\mathbb{E}_T\left[f_{P_T(t)}(\top)\right]} \quad = \frac{\mathbb{E}_T\left[f_{P_T(t)}(\alpha^t)\right]}{H(T)}$$

$$\hat{I}_T(T) = 1 \quad\quad\quad\quad = \frac{H(T)}{H(T)} \tag{23}$$

$$\hat{I}_T(\mathbf{S}_i) = \hat{I}_T(T \wedge \mathbf{S}_i) \quad = \frac{I(T; \mathbf{S}_i)}{H(T)}$$

We obtain the following algebra with the bi-valuation $\hat{I}_T([\alpha; \beta])$ that quantifies a degree of inclusion from α within the context of β. We can think of $\hat{I}_T([\alpha; \beta])$ as asking how much of the information from β about T is shared with α.

$$\hat{I}_T(\alpha \vee \beta) = \hat{I}_T(\alpha) + \hat{I}_T(\beta) - \hat{I}_T(\alpha \wedge \beta) \tag{Sum rule} \tag{24a}$$

$$\hat{I}_T([\alpha; \beta]) \equiv \frac{\hat{I}_T(\alpha \wedge \beta)}{\hat{I}_T(\beta)} \tag{Bi-Valuation} \tag{24b}$$

$$\hat{I}_T([\alpha \vee \beta; \gamma]) = \hat{I}_T([\alpha; \gamma]) + \hat{I}_T([\beta; \gamma]) - \hat{I}_T([\alpha \wedge \beta; \gamma]) \tag{Conditioned sum rule} \tag{24c}$$

$$\hat{I}_T([\beta \wedge \gamma; \alpha]) = \hat{I}_T([\gamma; \alpha \wedge \beta]) \cdot \hat{I}_T([\beta; \alpha]) \tag{Product rule} \tag{24d}$$

$$\hat{I}_T([\beta; \alpha \wedge \gamma]) = \frac{\hat{I}_T([\gamma; \alpha \wedge \beta]) \cdot \hat{I}_T([\beta; \alpha])}{\hat{I}_T([\gamma; \alpha])} \tag{Bayes' Theorem} \tag{24e}$$

Since the definitions satisfy an inclusion–exclusion principle, we obtain the interpretation of classical measures as proposed by Williams and Beer [5]: conditional mutual information $I(T; V_1 \mid V_2)$ measures the unique contribution of V_1 plus its synergy with V_2, and interaction information $I(T; V_1; V_2)$ measures the difference between synergy and shared information, which explains its possible negativity.

As highlighted by Knuth [13], the lattice product (the Cartesian product with ordering $(\alpha; \beta) \preceq (\tau; \upsilon) \Leftrightarrow \alpha \preceq \tau$ and $\beta \preceq \upsilon$) can be valuated using a product rule to maintain consistency with the ordering of the individual lattices. This creates an opportunity to define information product spaces for multiple reference variables. Since we normalized the measures, the valuation of the product space will also be normalized to the range from zero to one. The subscript notation $T_1 \times T_2$ shall indicate the product of the lattice constructed for T_1 with the product of the lattice constructed for T_2.

$$\hat{I}_{(T_1 \times T_2)}((\alpha; \beta)) = \hat{I}_{T_1}(\alpha) \cdot \hat{I}_{T_2}(\beta) \qquad \text{(Valuation Product rule)} \qquad \text{(25a)}$$

$$\hat{I}_{(T_1 \times T_2)}(([\alpha; \tau]; [\beta; \upsilon])) = \hat{I}_{T_1}([\alpha; \beta]) \cdot \hat{I}_{T_2}([\tau; \upsilon]) \qquad \text{(Bi-Valuation Product rule)} \qquad \text{(25b)}$$

The lattice product is distributive over the joint for disjoint elements [13], which leads to the equivalence in Equation (26). Unfortunately, it appears that only the bottom element is disjoint with other atoms in the constructed lattice.

$$\forall t: \alpha^t \wedge \beta^t \sim \bot \quad \Rightarrow \quad \hat{I}_{(T_1 \times T_2)}((\alpha \vee \beta; \tau)) = \hat{I}_{(T_1 \times T_2)}((\alpha; \tau) \vee (\beta; \tau)) \qquad \text{(26)}$$

Finally, we would like to provide an intuition for this approach based on possible operational scenarios:

1. Consider having characterized four radio links and obtained the conditional distributions $P_{V_1|T}$, $P_{(V_2,V_3)|T}$ and $P_{V_4|T}$. We are interested in their joint channel capacity; however, lack the required joint distribution. In this case, we can use their joint $\sup_{P_T(t)} \hat{I}_T(\mathbf{S}_1 \vee \mathbf{S}_{23} \vee \mathbf{S}_4)$ to obtain a (pointwise) lower bound on their joint channel capacity.
2. Consider having two datasets $\{T_1, V_1, V_2, V_3\}$ and $\{T_2, V_2, V_3, V_4\}$ that provide different types of labels (T_x) and associated features (V_y), where some events were recorded in both datasets. In such cases, one may choose to study the cases $T_1 \rightarrow (V_1, V_2, V_3)$, $T_2 \rightarrow (V_2, V_3, V_4)$ and $(T_1, T_2) \rightarrow (V_1, V_2, V_3, V_4)$ for events appearing in both datasets, which could then be combined into a product lattice $\hat{I}_{(T_1 \times T_2 \times (T_1, T_2))}$.

4. Applications

This section focuses on applications of the obtained measure from Section 3.3. We first apply the meet operator to the redundancy lattice for constructing a PID. Since an atom of the redundancy lattice $\alpha \in \mathcal{A}(\mathbf{V})$ corresponds to a set of sources for which the shared information shall be measured, we use the notation $\bigwedge \alpha$ to obtain an expression for the function $\hat{I}_T$. Section 4.2 additionally utilizes the properties of a Markov chain to demonstrate how the flow of partial information can be traced through system models.

4.1. Partial Information Decomposition

Based on Section 3.3, we can define a measure of shared information $\hat{I}(\alpha; T)$ for the elements of the redundancy lattice $\alpha \in \mathcal{A}(\mathbf{V})$ in the framework of Williams and Beer [5], as shown in Equation (27). The measure satisfies the three axioms of Williams and Beer [5] (commutativity from the equivalence relation and structure of f_p, monotonicity from being a lattice valuation and self-redundancy from removing the normalization), and the decomposition is non-negative since the joint channel κ_{12}^t is superior to the joint of two channels $\kappa_1^t \vee \kappa_2^t$ for all $t \in \mathcal{T}$. The partial contribution $\hat{I}^\delta(\alpha; T)$ corresponds to the expected value of the quantified partial decision regions $\alpha^{\delta t}$.

This provides the interpretation of Section 3.1, where combining the partial contributions of the up-set corresponds to the expected value of quantifying the decision regions that are lost when losing the variable, while combining the partial contributions of the down-set

corresponds to the expected value of quantifying the accessible decision region from this variable. Additionally, we obtain a pointwise version of the property by Bertschinger et al. [10]: if a variable provides unique information, then there is a way to utilize this information for a reward function to some target variable state. Finally, it can be seen that taking the minimal quantification of the different decision regions as done by Williams and Beer [5] leads to a lack in distinguishing distinct reachable decision regions or, as phrased in the literature: a lack of distinguishing "the *same* information and the *same amount* of information" [6–9].

$$\forall \alpha \in \mathcal{A}(\mathbf{V}), \quad \hat{I}(\alpha; T) = \hat{I}_T\left(\bigwedge \alpha\right) \cdot H(T), \tag{27a}$$

$$\hat{I}^{\delta}(\alpha; T) = \hat{I}(\alpha; T) - \sum_{\beta \in \downarrow\alpha} \hat{I}^{\delta}(\beta; T) = \mathbb{E}_T\left[f_{P_T(t)}\left(\alpha^{\delta t}\right)\right] \tag{27b}$$

An identical definition of $\hat{I}(\alpha; T)$ can be obtained only based on the Blackwell order, as shown in Equation (28). Let $\alpha \in \mathcal{A}(\mathbf{V})$ be a set of sources and let T^t represent a binary target variable ($\mathcal{T}^t = \{t, \bar{t}\}$) such that $T^t = t \Leftrightarrow T = t$. We can expand the meet operator used in Equation (27a) using the sum-rule and utilize the distributivity for arriving at the joint of two channels, which matches the Blackwell order (Equation (28b)). We write $\mathbf{S}_i \sqcup_{T^t} \mathbf{S}_j$ to refer to the joint of $\mathbf{S}_i$ and $\mathbf{S}_j$ under the Blackwell order with respect to variable T^t. This results in the recursive definition of $i(\alpha; T^t)$ that corresponds to the definition of mutual information for a single source (Equation (28a)). This expansion of Equation (27a) is particularly helpful since it eliminates the operators $\wedge/\vee$ for a simplified implementation.

$$i(\{\mathbf{S}_i\}; T^t) = \sum_{s \in \mathcal{S}_i} P_{(\mathbf{S}_i|T^t=t)}(s) \log\left(\frac{P_{(\mathbf{S}_i|T^t=t)}(s)}{P_{(T^t=t)}P_{(\mathbf{S}_i|T^t=t)}(s) + (1 - P_{(T^t=t)})P_{(\mathbf{S}_i|T^t \neq t)}(s)}\right) \tag{28a}$$

$$i(\{\mathbf{S}_i\} \cup \beta; T^t) = i(\{\mathbf{S}_i\}; T^t) + i(\beta; T^t) - i(\{\mathbf{S}_i \sqcup_{T^t} \mathbf{S}_j \mid \mathbf{S}_j \in \beta\}; T^t) \tag{28b}$$

$$\hat{I}(\alpha; T) = \mathbb{E}_T\left[i(\alpha, T^t)\right] \tag{28c}$$

Our decomposition is equivalent to the measures of Bertschinger et al. [10], Griffith and Koch [11] and Williams and Beer [5] in two special cases:

- For a binary target variable $\mathcal{T} = \{t, \bar{t}\}$ with two observable variables V_1 and V_2, our approach is identical to Bertschinger et al. [10] and Griffith and Koch [11] since $\kappa_1 \sqcup \kappa_2 \sim \kappa_1^t \vee \kappa_2^t \sim \kappa_1^{\bar{t}} \vee \kappa_2^{\bar{t}}$. Beyond binary target variables, the resulting definitions differ due to the pointwise construction (see Appendix E).

- If from a pointwise perspective (T^t), some variable is Blackwell superior to the other (not necessarily the same each time), then our method is identical to Williams and Beer [5] since the defined meet operation will equal their minimum $\kappa_1^t \sqcup \kappa_2^t \sim \kappa_2^t \Rightarrow f_p(\kappa_1^t) \leq f_p(\kappa_2^t) \Rightarrow \min(f_p(\kappa_1^t), f_p(\kappa_2^t)) = f_p(\kappa_1^t \wedge \kappa_2^t) = f_p(\kappa_1^t)$ and equivalently for the function $i(\alpha, T^t)$.

A decomposition of typical examples can be found in Appendix E. We also provide an implementation of the PID based on our approach [18].

4.2. Information Flow Analysis

Due to the achieved inclusion–exclusion principle, the data processing inequality of mutual information and the achieved non-negativity of partial information for an arbitrary number of variables, it is possible to trace the flow of information through Markov chains. The measure $\hat{I}_T$ appears suitable for this analysis due to the chaining properties of the underlying pointwise channels that are quantified. The analysis can be applied among others for analyzing communication networks or designing data processing systems.

The flow of information in Markov chains has been studied by Niu and Quinn [19], who considered chaining individual variables $X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_n$ and performed a decomposition on $\mathbf{V} = \{X_1, X_2, \ldots, X_n\}$. In contrast to this, we consider Markov chains

that map sets of random variables from one step to the next. In this case, it is possible to perform an information decomposition at each step of the Markov chain and identify how the partial information components propagate from one set of variables to the next.

Let $T \to \mathbf{V} \to \mathbf{Q}$ be a Markov chain with the atoms $\alpha \in \mathcal{A}(\mathbf{V})$ and $\beta \in \mathcal{A}(\mathbf{Q})$, through which we trace the flow of partial information from α to β about T. We can measure the shared information between both atoms α and β, as shown in Equation (29a), to obtain how much information their cumulative components share $\hat{J}^{\cap \to \cap}(\alpha \to \beta; T)$. Similar to the PID, we remove the normalization for the self-redundancy axiom. To identify how much of the cumulative information of β is obtained from the partial information of α, we subtract the strict down-set of α on the lattice $(\mathcal{A}(\mathbf{V}), \preccurlyeq)$ as shown in Equation (29b) to obtain $\hat{J}^{\delta \to \cap}(\alpha \to \beta; T)$. To compute how much of the partial information of α is shared with the partial contribution of β, we similarly remove the flow from the partial information of α into the strict down-set of β on the lattice $(\mathcal{A}(\mathbf{Q}), \preccurlyeq)$, as shown in Equation (29c), to obtain $\hat{J}^{\delta \to \delta}(\alpha \to \beta; T)$. This can be used to trace the origin of information for each atom $\beta \in \mathcal{A}(\mathbf{Q})$ to the previous elements $\alpha \in \mathcal{A}(\mathbf{V})$.

The approach is not limited to one step and can be extended for tracing the flow through Markov chains of arbitrary length $\hat{J}^{\delta \to \delta \to \delta \cdots}(\alpha \to \beta \to \gamma \ldots; T)$. However, we only trace one step in this demonstration for simplicity.

$$\hat{J}^{\cap \to \cap}(\alpha \to \beta; T) = \hat{I}_T\left(\bigwedge \alpha \wedge \bigwedge \beta\right) \cdot H(T) \tag{29a}$$

$$\hat{J}^{\delta \to \cap}(\alpha \to \beta; T) = \hat{J}^{\cap \to \cap}(\alpha \to \beta; T) - \sum_{\gamma \in \downarrow \alpha} \hat{J}^{\delta \to \cap}(\gamma \to \beta; T) \tag{29b}$$

$$\hat{J}^{\delta \to \delta}(\alpha \to \beta; T) = \hat{J}^{\delta \to \cap}(\alpha \to \beta; T) - \sum_{\gamma \in \downarrow \beta} \hat{J}_T^{\delta \to \delta}(\alpha \to \gamma; T) \tag{29c}$$
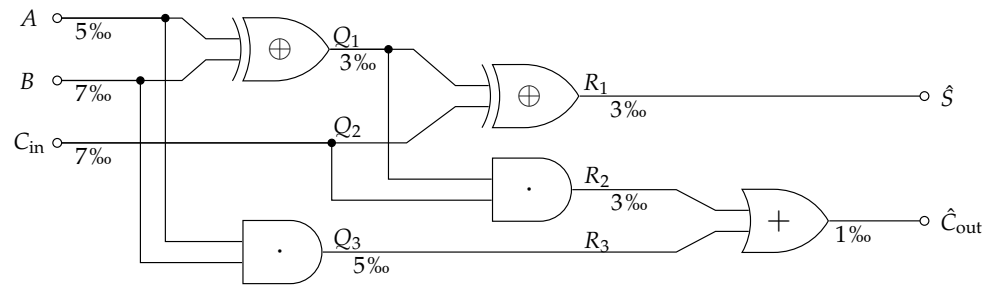
We demonstrate the Information Flow Analysis using a full-adder as a small logic circuit with the input variables $\mathbf{V} = \{A, B, C_{\text{in}}\}$ and the output $\mathbf{T} = \{S, C_{\text{out}}\}$ as shown in Equation (30). Any ideal implementation of this computation results in the same channel from $\mathbf{V}$ to $\mathbf{T}$. Therefore, they create an identical flow of the partial information from $\mathbf{V}$ to the partial information of $\mathbf{T}$. However, the specific implementation will determine how (over which intermediate representations and paths) the partial information is transported.

$$
\begin{aligned}
S &= A \oplus B \oplus C_{\text{in}} \\
C_{\text{out}} &= A \cdot B + A \cdot C_{\text{in}} + B \cdot C_{\text{in}} \\
&= (A \cdot B) + C_{\text{in}} \cdot (A \oplus B)) \qquad \text{(typical implementation)} \\
T &= (S, C_{\text{out}})
\end{aligned}
\tag{30}
$$

To make the example more interesting, we consider the implementation of a noisy full-adder, as shown in Figure 7, which allows for bit-flips on wires. We indicate the probability of a bit-flip below each line and imagine this value correlates to the wire length and proximity to others. Now, changing the implementation or even the layout of the same circuit would have an impact on the overall channel.

To perform the analysis, we first have to define the target variable: What it is that we want to measure information about? In this case, we select the joint distribution of the desired computation output T as the target variable and define the noisy computation result to be $\hat{\mathbf{T}} = \{\hat{S}, \hat{C}_{\text{out}}\}$, as shown in Figure 7. We obtain both variables from their definition by assuming that the input variables $\mathbf{V}$ are independently and uniformly distributed and that bit-flips occurred independently. However, it is worth noting that noise dependencies can be modeled in the joint distribution. This fully characterizes the Markov chain shown in Equation (31).

$$T = (S, C_{\text{out}}) \to \mathbf{T} = \{S, C_{\text{out}}\} \to \mathbf{V} = \{A, B, C_{\text{in}}\} \to \mathbf{Q} = \{Q_1, Q_2, Q_3\} \to \mathbf{R} = \{R_1, R_2, R_3\} \to \hat{\mathbf{T}} = \{\hat{S}, \hat{C}_{\text{out}}\} \tag{31}$$

Figure 7. Noisy full-adder example for the Information Flow Analysis demonstration. The probability of a bit-flip is indicated below the wires. If a wire has two labels, the first label corresponds to the wire input and the second label to its output.

We group two variables at each stage to reduce the number of interactions in the visualization. The resulting information flow of the full-adder is shown as a Sankey diagram in Figure 8. Each bar corresponds to the mutual information of a stage in the Markov chain with the input T. The bars' colors indicate the partial information decomposition of Equation (27). The information flow over one step using Equation (29) is indicated by the width of a line between the partial contributions of two stages. To follow the flow of a particular component over more than one step—for example, to see how the shared information of $\mathbf{T}$ propagates to the shared information of $\hat{\mathbf{T}}$—the analysis can be performed by tracing multiple steps after extending Equation (29).
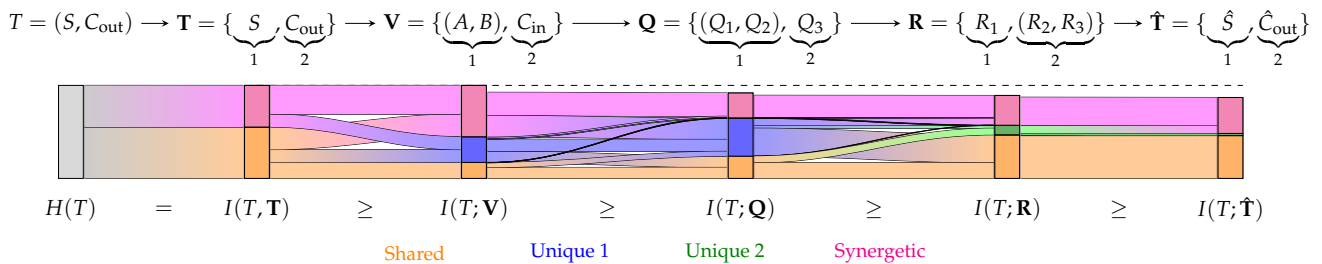
Figure 8. Sankey diagram of the Information Flow Analysis for the noisy full-adder in Figure 7. Each bar corresponds to one stage in the Markov chain, and its height corresponds to this stage's mutual information with the target T. Each bar is decomposed into the information that the considered variables provide shared (orange), unique (blue/green) or synergetic (pink) about the target. If a stage is represented by a single variable or joint distribution, no further decomposition is performed (gray). We trace the information between variables over one step using the sub-chains $T \to T \to \mathbf{T}$, $T \to \mathbf{T} \to \mathbf{V}$, $T \to \mathbf{V} \to \mathbf{Q}$, $T \to \mathbf{Q} \to \mathbf{R}$ and $T \to \mathbf{R} \to \hat{\mathbf{T}}$ using Equation (29). The resulting flows between each bar visualize how the partial information propagates for one step in the Markov chain. For following the flow of a particular partial component over more than one step in the Sankey diagram, Equation (29) can be extended.

The results (Figure 8) show that the decomposition does not attribute unique information to S or C_{out} about their own joint distribution. The reason for this is shown in Equation (32): both variables provide an equivalent channel for each state of their joint distribution and, thus, an equivalent uncertainty about each state of T. Phrased differently, both variables provide access to the identical decision regions for each state of their joint distribution and can therefore not provide unique information (no advantage for any reward function to any $t \in \mathcal{T}$). If this result feels counter-intuitive, we would also recommend the discussion of the two-bit-copy problem and identity axiom by Finn [9] (p. 16ff.) and Finn and Lizier [20]. The same effect can also be seen when viewing each variable in $\mathbf{V}$ individually (not shown in Figure 8), which causes neither of them to provide unique information on their own about the joint target distribution T.

$$(T^{(0,0)} \to C_{\text{out}}) \sim (T^{(0,0)} \to S) \sim \begin{bmatrix} 1 & 0 \\ 3/7 & 4/7 \end{bmatrix} \sim (T^{(1,1)} \to C_{\text{out}}) \sim (T^{(1,1)} \to S)$$

$$(T^{(0,1)} \to C_{\text{out}}) \sim (T^{(0,1)} \to S) \sim \begin{bmatrix} 1 & 0 \\ 1/5 & 4/5 \end{bmatrix} \sim (T^{(1,0)} \to C_{\text{out}}) \sim (T^{(1,0)} \to S)$$

(32)

The Information Flow Analysis is particularly useful in practice since it can be performed on an arbitrary resolution of the system model to handle its complexity. For example, a small full-adder can be analyzed on the level of gates and wires represented by channels. However, the full-adder is itself a channel that can be used to analyze an *n*-bit adder on the level of full-adders.

Further applications of the Information Flow Analysis could include the identification of which inputs are most critical for the computational result and where information is being lost. It can also be explored if a notion of robustness in data processing systems could be meaningfully defined based on how much pointwise redundant or shared information of the input $\mathbf{V}$ can be traced to its output $\hat{\mathbf{T}}$. This might indicate a notion of robustness based on whether or not it is possible to compensate for the unavailability of input sources through a system modification.

Finally, the target variable does not have to be the desired computational outcome as has been done in the demonstration. When thinking about secure multi-party computations, it might be of interest to identify the flow of information from the perspective of some sensitive or private variable (T) to understand the impact of disclosing the final computation result. The possible applications of such an analysis are as diverse as those of information theory.

5. Discussion

We propose the interpretation that the reachable decision regions correspond to different notions of uncertainty about each state of the target variable and that mutual information corresponds to the expected value of quantifying these decision regions. This allows partial information to represent the expected value of quantifying partial decision regions (Equations (27) and (28)), which can be used to attribute mutual information to the visible variables and their interactions (pointwise redundant/shared/unique/synergetic). Since the proposed quantification results in the consistent valuation of a distributive lattice, it creates a novel algebra for mutual information with possible practical applications (Equations (24) and (25)). Finally, the approach allows for tracing information components through Markov chains (Equation (29)), which can be used to model and study a wide range of scenarios. The presented method is directly applicable to discrete and categorical source variables due to their equivalent construction for the reachable decision regions (zonotopes). However, we recommend that the target variable should be categorical since the measure does not consider a notion of distance between target states (achievable estimation proximity). This would be an interesting direction for future work due to its practical application for introducing semantic meaning to sets of variables. An intuitive example is a target variable with 256 states that is used to represent an 8-bit unsigned integer as the computation result. For this reason, we wonder if it is possible to introduce a notion of distance to the analysis such that the classical definition of mutual information becomes the special case for encoding categorical targets.

A recent work by Kolchinsky [21] removes the assumption that an inclusion–exclusion principle relates the intersection and union of information and demands their extractability. This has the disadvantage that a similar algebra or tracing of information would no longer be possible. We tried to address this point by distinguishing the *pointwise redundant* from the *pointwise shared* element and also obtain no inclusion–exclusion principle for the pointwise redundancy. We focus in this work on the pointwise shared element due to the resulting properties and operational interpretation from the accessibility and losses of reachable decision regions. Moreover, the relation between the used meet and joint operators provides consistent results from performing the decomposition using the meet

operator on a redundancy lattice, as done in this work, or a decomposition using the joint operator on a synergy or loss lattice [22].

Further notions of redundancy and synergy can be studied within this framework if they are extractable, meaning they can be represented by some random variable. Depending on the desired interpretation, the representing variable can be constructed for T and added to the set of visible variables or can be constructed for each pointwise variable T^t and added to the pointwise lattices. We showed an example of the latter in Section 3.1 by adding the pointwise redundant element to the lattice, which we interpret as pointwise extractable components of shared information to quantify the decision regions that can be obtained from each source.

Since our approach satisfies the original axioms of Williams and Beer [5] and results in non-negative partial contributions for an arbitrary number of variables, it cannot satisfy the proposed identity axiom of Harder et al. [8]. This can also be seen by the decomposition examples in Appendix E (Table A2 and Figure A3). We do not consider this a limitation since all four axioms cannot be satisfied without obtaining negative partial information [23], which creates difficulties for interpreting results.

Finally, our approach does not appear to satisfy a target/left chain rule as proposed by Bertschinger et al. [7]. While our approach provides an algebra that can be used to handle multiple target variables, we think that further work on understanding the relations when decomposing with multiple target variables is needed. In particular, it would be helpful for the analysis of complex systems if the flow of already analyzed sub-chains could be reused and their interactions could be predicted.

6. Conclusions

We use the approach of Bertschinger et al. [10] and Griffith and Koch [11] to construct a pointwise partial information decomposition that provides non-negative results for an arbitrary number of variables and target states. The measure obtains an algebra from the resulting lattice structure and enables the analysis of complex multivariate systems in practice. To our knowledge, this is the first alternative to the original measure of Williams and Beer [5] that satisfies their three proposed axioms and results in a non-negative decomposition for an arbitrary number of variables.

Author Contributions: T.M. and C.R conceived the idea; T.M. prepared the original draft; C.R. reviewed and edited the draft. All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following abbreviations are used:

PID	Partial Information Decomposition
ROC	Receiver Operating Characteristic
TPR	True-Positive Rate ($\bar{\beta}$)
FPR	False-Positive Rate (α)

We use the following notation conventions:

$T, \mathcal{T}, t, T^t$　　　T (upper case) represents the target variable with an event t (lower case) of its event space (calligraphic), $t \in \mathcal{T}$. T^t represents a pointwise (binary) target variable which takes state one if $T = t$ and state two if $T \neq t$ (T^t represents the one-vs-rest encoding of state t);

$\mathbf{V}, V_i, \mathcal{V}_i, v$　　　$\mathbf{V}$ represents a set of visible/observable/predictor variables V_i with $v \in \mathcal{V}_i$;

$\mathbf{S}_i, \mathcal{S}_i$　　　*sources* represent a set of visible variables, where the index i lists the contained visible variables, such as $\mathbf{S}_{12} = \{V_1, V_2\}$. The event $s \in \mathcal{S}_i$ corresponds to an event of the corresponding joint variable, e.g., (V_1, V_2).

We represent channels (κ, λ) as row stochastic matrices with the following indexing:

$\mathbf{P}$　　　represents a permutation matrix;

κ_i　　　represents a channel from the target to a source $T \xrightarrow{\kappa_i} \mathbf{S}_i$ using the joint distribution of the variables within the source, such as $T \xrightarrow{\kappa_{12}} (V_1, V_2)$;

κ_i^t　　　represents a pointwise channel from the target to a source $T^t \xrightarrow{\kappa_i^t} \mathbf{S}_i$, such as $T^t \xrightarrow{\kappa_{12}^t} (V_1, V_2)$;

$Z_{\kappa_i^t}$　　　binary input channels κ_i^t can be represented as (row) stochastic matrix, which contain a likelihood vector $\vec{v}_s = \begin{pmatrix} p(\mathbf{S}_i = s | T = t) \\ p(\mathbf{S}_i = s | T \neq t) \end{pmatrix}$ for each state $s \in \mathcal{S}_i$. $Z_{\kappa_i^t}$ represents the zonotope for this set of vectors;

$\kappa_1^t \vee \kappa_2^t$　　　represents the binary input channel corresponding to the convex hull of $Z_{\kappa_1^t}$ and $Z_{\kappa_2^t}$ (Blackwell order joint of binary input channels $\kappa_1^t \vee \kappa_2^t \equiv \kappa_1^t \sqcup \kappa_2^t$);

$\kappa_1^t \wedge \kappa_2^t$　　　represents the meet element for constructing a distributive lattice with the joint operator $\kappa_1^t \vee \kappa_2^t$;

$\kappa_1^t \sqcap \kappa_2^t$　　　represents the binary input channel corresponding to the intersection of $Z_{\kappa_1^t}$ and $Z_{\kappa_2^t}$ (Blackwell order meet of binary input channels);

α, β　　　*atoms* represent an expression of random variables with the operators $(\vee / \wedge)$. In Sections 2.2 and 4, they represent sets of sources;

α^t, β^t　　　represent an expression of pointwise channels with the operators $(\vee / \wedge)$;

$\alpha^{\delta t}, \beta^{\delta t}$　　　represent a partial pointwise channel corresponding to α^t.

We use the following convention for operations, functions and brackets:

$\mathcal{P}_1(\cdot)$　　　represents the power set without the empty set;

$\{V_1, V_2\}$　　　curly brackets with comma separation represent a set;

$[\, M_1 \; M_2 \,]$　　　square brackets without comma separation represent a matrix, and the listing of matrices in this manner represents their concatenation;

$q([\alpha; \beta])$　　　square brackets with semicolon separation are used to refer to the bi-valuation $b(\alpha, \beta)$ of a consistent lattice valuation $q(\alpha)$. In a similar manner to Knuth [13], we use the notation $q([\alpha; \beta]) \equiv b(\alpha, \beta)$;

$(\alpha; \beta)$　　　round brackets with semicolon separation represent an element of a Cartesian product $\mathbf{L}_1 \times \mathbf{L}_2$, where $\alpha \in \mathbf{L}_1$ and $\beta \in \mathbf{L}_2$;

$f\langle \mathbf{L} \rangle$　　　angled brackets indicate that a function f shall be mapped to each element of the set $\mathbf{L}$. We may nest this notation, such as $f\langle\langle \mathbf{L} \rangle\rangle$, to indicate a map to each element of the sets within $\mathbf{L}$;

$\boldsymbol{\alpha}$　　　False-Positive Rate, type I error;

$\overline{\boldsymbol{\beta}}$　　　True-Positive Rate, $1-$ type II error.

We distinguish between a *joint channel* $T \xrightarrow{\kappa_{12}} (V_1, V_2)$ and the *joint of two channels* $\kappa_1 \vee \kappa_2$. To avoid confusion, we write the first case as "joint channel (κ)" and the latter case as "joint of channels $(\kappa_i \vee \kappa_j)$" throughout this work.

Appendix A

The considered lattice relates the meet and joint elements $(\wedge / \vee)$ through an inclusion–exclusion principle. Here, the partial contribution for the joint of any two incomparable elements $(\alpha^t, \beta^t \in \mathcal{B}^t(\mathbf{V}), \alpha^t \vee \beta^t \not\sim \alpha^t$ and $\alpha^t \vee \beta^t \not\sim \beta^t)$ shall be zero, which is indicated using a gray font in Figure A1.
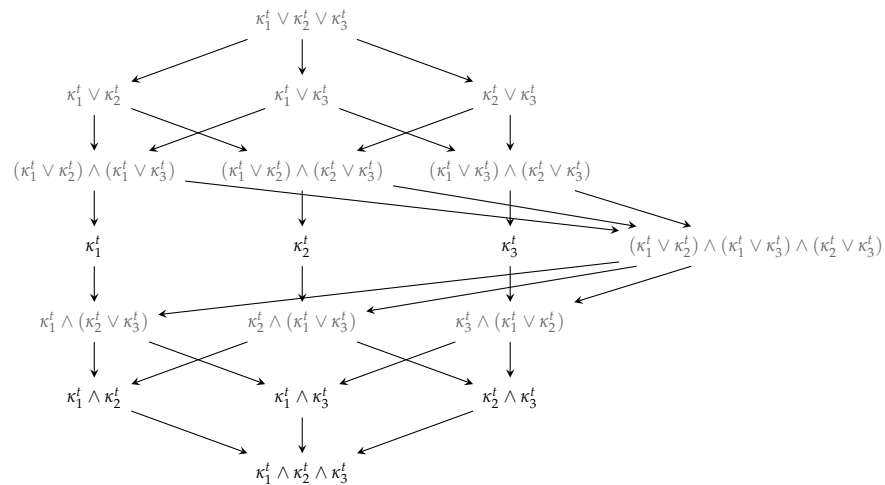
Figure A1. The considered lattice relating the meet and joint operators. The joint of any two incomparable elements ($\alpha^t, \beta^t \in \mathcal{B}^t(\mathbf{V}), \alpha^t \vee \beta^t \not\sim \alpha^t$ and $\alpha^t \vee \beta^t \not\sim \beta^t$) shall have no partial contribution to create an inclusion–exclusion principle between the operators and is highlighted using a gray font.

Appendix B

This section demonstrates that the defined meet and joint operators of Section 3.2 provide a distributive lattice under the defined equivalence relation ($\sim$, Equation (10)).

Lemma A1. *The meet and joint operators ($\wedge$, $\vee$) define a distributive lattice for a set of channels under the defined equivalence relation ($\sim$).*

Proof. The definitions of the meet and joint satisfy associativity, commutativity, idempotency, absorption and distributivity on channels under the defined equivalence relation:

1. *Idempotency:* $\kappa_1^t \vee \kappa_1^t \sim \kappa_1^t$ and $\kappa_1^t \wedge \kappa_1^t \sim \kappa_1^t$.

$$\kappa_1^t \vee \kappa_1^t \sim \kappa_1^t \sqcup \kappa_1^t \sim \kappa_1^t;$$
$$\kappa_1^t \wedge \kappa_1^t \sim \begin{bmatrix} \kappa_1^t & \kappa_1^t & -\kappa_1^t \vee \kappa_1^t \end{bmatrix} \sim \begin{bmatrix} \kappa_1^t & \kappa_1^t & -\kappa_1^t \end{bmatrix} \sim \kappa_1^t.$$

2. *Commutativity:* $\kappa_1^t \vee \kappa_2^t \sim \kappa_2^t \vee \kappa_1^t$ and $\kappa_1^t \wedge \kappa_2^t \sim \kappa_2^t \wedge \kappa_1^t$.

$$\kappa_1^t \vee \kappa_2^t \sim \kappa_1^t \sqcup \kappa_2^t \sim \kappa_2^t \sqcup \kappa_1^t \sim \kappa_2^t \vee \kappa_1^t;$$
$$\kappa_1^t \wedge \kappa_2^t \sim \begin{bmatrix} \kappa_1^t & \kappa_2^t & -\kappa_1^t \vee \kappa_2^t \end{bmatrix} \sim \begin{bmatrix} \kappa_2^t & \kappa_1^t & -\kappa_2^t \vee \kappa_1^t \end{bmatrix} \sim \kappa_2^t \wedge \kappa_1^t.$$

3. *Associativity:* $\kappa_1^t \vee (\kappa_2^t \vee \kappa_3^t) \sim (\kappa_1^t \vee \kappa_2^t) \vee \kappa_3^t$ and $\kappa_1^t \wedge (\kappa_2^t \wedge \kappa_3^t) \sim (\kappa_1^t \wedge \kappa_2^t) \wedge \kappa_3^t$.

$$\kappa_1^t \vee (\kappa_2^t \vee \kappa_3^t) \sim \kappa_1^t \sqcup (\kappa_2^t \sqcup \kappa_3^t) \sim (\kappa_1^t \sqcup \kappa_2^t) \sqcup \kappa_3^t \sim (\kappa_1^t \vee \kappa_2^t) \vee \kappa_3^t;$$
$$\kappa_1^t \wedge (\kappa_2^t \wedge \kappa_3^t) \sim \begin{bmatrix} \kappa_1^t & \kappa_2^t & \kappa_3^t & -\kappa_1^t \vee \kappa_2^t & -\kappa_1^t \vee \kappa_3^t & -\kappa_2^t \vee \kappa_3^t & \kappa_1^t \vee \kappa_2^t \vee \kappa_3^t \end{bmatrix}$$
$$\sim \begin{bmatrix} \kappa_3^t & \kappa_2^t & \kappa_1^t & -\kappa_3^t \vee \kappa_2^t & -\kappa_3^t \vee \kappa_1^t & -\kappa_2^t \vee \kappa_1^t & \kappa_3^t \vee \kappa_2^t \vee \kappa_1^t \end{bmatrix}$$
$$\sim \kappa_3^t \wedge (\kappa_2^t \wedge \kappa_1^t) \sim (\kappa_1^t \wedge \kappa_2^t) \wedge \kappa_3^t.$$

4. *Absorption:* $\kappa_1^t \wedge (\kappa_1^t \vee \kappa_2^t) \sim \kappa_1^t$ and $\kappa_1^t \vee (\kappa_1^t \wedge \kappa_2^t) \sim \kappa_1^t$.

$$\kappa_1^t \wedge (\kappa_1^t \vee \kappa_2^t) \sim \begin{bmatrix} \kappa_1^t & \kappa_1^t \vee \kappa_2^t & -\kappa_1^t \vee \kappa_1^t \vee \kappa_2^t \end{bmatrix}$$
$$\sim \begin{bmatrix} \kappa_1^t & \kappa_1^t \vee \kappa_2^t & -\kappa_1^t \vee \kappa_2^t \end{bmatrix} \sim \kappa_1^t;$$
$$\kappa_1^t \vee (\kappa_1^t \wedge \kappa_2^t) \sim \begin{bmatrix} \kappa_1^t & \kappa_1^t \wedge \kappa_2^t & -\kappa_1^t \wedge \kappa_1^t \wedge \kappa_2^t \end{bmatrix}$$
$$\sim \begin{bmatrix} \kappa_1^t & \kappa_1^t \wedge \kappa_2^t & -\kappa_1^t \wedge \kappa_2^t \end{bmatrix} \sim \kappa_1^t.$$

5. *Distributivity:* $\kappa_1^t \vee (\kappa_2^t \wedge \kappa_3^t) \sim (\kappa_1^t \vee \kappa_2^t) \wedge (\kappa_1^t \vee \kappa_3^t)$ and $\kappa_1^t \wedge (\kappa_2^t \vee \kappa_3^t) \sim (\kappa_1^t \wedge \kappa_2^t) \vee (\kappa_1^t \wedge \kappa_3^t)$.

$$
\begin{aligned}
\kappa_1^t \vee (\kappa_2^t \wedge \kappa_3^t) &\sim \begin{bmatrix} \kappa_2^t \wedge \kappa_3^t & \kappa_1^t & -\kappa_1^t \wedge (\kappa_2^t \wedge \kappa_3^t) \end{bmatrix} \\
&\sim \begin{bmatrix} \kappa_2^t \wedge \kappa_3^t & -\kappa_2^t & -\kappa_3^t & \kappa_1^t \vee \kappa_2^t & \kappa_1^t \vee \kappa_3^t & \kappa_2^t \vee \kappa_3^t & -\kappa_1^t \vee \kappa_2^t \vee \kappa_3^t \end{bmatrix} \\
&\sim \begin{bmatrix} -\kappa_2^t \vee \kappa_3^t & \kappa_1^t \vee \kappa_2^t & \kappa_1^t \vee \kappa_3^t & \kappa_2^t \vee \kappa_3^t & -\kappa_1^t \vee \kappa_2^t \vee \kappa_3^t \end{bmatrix} \\
&\sim \begin{bmatrix} \kappa_1^t \vee \kappa_2^t & \kappa_1^t \vee \kappa_3^t & -\kappa_1^t \vee \kappa_2^t \vee \kappa_3^t \end{bmatrix} \\
&\sim \begin{bmatrix} \kappa_1^t \vee \kappa_2^t & \kappa_1^t \vee \kappa_3^t & -(\kappa_1^t \vee \kappa_2^t) \vee (\kappa_1^t \vee \kappa_3^t) \end{bmatrix} \\
&\sim (\kappa_1^t \vee \kappa_2^t) \wedge (\kappa_1^t \vee \kappa_3^t); \\
\kappa_1^t \wedge (\kappa_2^t \vee \kappa_3^t) &\sim \begin{bmatrix} \kappa_2^t \vee \kappa_3^t & \kappa_1^t & -\kappa_1^t \vee (\kappa_2^t \vee \kappa_3^t) \end{bmatrix} \\
&\sim \begin{bmatrix} \kappa_2^t \vee \kappa_3^t & -\kappa_2^t & -\kappa_3^t & \kappa_1^t \wedge \kappa_2^t & \kappa_1^t \wedge \kappa_3^t & \kappa_2^t \wedge \kappa_3^t & -\kappa_1^t \wedge \kappa_2^t \wedge \kappa_3^t \end{bmatrix} \\
&\sim \begin{bmatrix} -\kappa_2^t \wedge \kappa_3^t & \kappa_1^t \wedge \kappa_2^t & \kappa_1^t \wedge \kappa_3^t & \kappa_2^t \wedge \kappa_3^t & -\kappa_1^t \wedge \kappa_2^t \wedge \kappa_3^t \end{bmatrix} \\
&\sim \begin{bmatrix} \kappa_1^t \wedge \kappa_2^t & \kappa_1^t \wedge \kappa_3^t & -(\kappa_1^t \wedge \kappa_2^t) \wedge (\kappa_1^t \wedge \kappa_3^t) \end{bmatrix} \\
&\sim (\kappa_1^t \wedge \kappa_2^t) \vee (\kappa_1^t \wedge \kappa_3^t).
\end{aligned}
$$

□

Appendix C

This section demonstrates the quantification of a small example and proves that the function f of Equation (19) creates a *consistent valuation* $\alpha^t \wedge \beta^t \sim \beta^t \Rightarrow f(\beta^t) \leq f(\alpha^t)$ for the pointwise lattice $(\mathcal{B}^t(\mathbf{V}), \wedge, \vee)$.

The convexity of the function $r(\vec{v})$ results, in combination with the property that $r(\ell\vec{v}) = \ell r(\vec{v})$ with $\ell \in \mathbb{R}$, in a triangle inequality, as shown in Equation (A1). This ensures that Blackwell superior channels obtain a larger quantification result and thus the non-negativity of channels: $f(\kappa^t \sqcup \lambda^t) \geq f(\kappa^t) \geq f(\begin{bmatrix} 1 \\ 1 \end{bmatrix}) = 0$.

$$
\begin{aligned}
r(t\vec{v}_1 + (1-t)\vec{v}_2) &\leq tr(\vec{v}_1) + (1-t)r(\vec{v}_2) & \text{(convexity, } 0 \leq t \leq 1) \\
r(\vec{v}_1 + \vec{v}_2) &\leq r(\vec{v}_1) + r(\vec{v}_2) & \text{(using } t = 0.5 \text{ and } r(\ell\vec{v}) = \ell r(\vec{v}))
\end{aligned} \tag{A1}
$$

To provide an intuition for the meet operator with a minimal example and highlight its relation to the intersection of zonotopes (redundant region), consider the two channels κ_1^t and κ_2^t of Equation (A2) and as visualized in Figure A2. To simplify the notation, we use the property $\begin{bmatrix} ((1+\ell)\vec{v}_1) \end{bmatrix} \sim \begin{bmatrix} (\vec{v}_1) & (\ell\vec{v}_1) \end{bmatrix}$ to differentiate the vectors $\vec{a}_2$ and $\vec{a}_3$ as well as $\vec{b}_1$ and $\vec{b}_2$.

$$
\begin{aligned}
\kappa_1^t &\sim \begin{bmatrix} (\vec{a}_1) & (\vec{a}_2) & (\vec{a}_3) \end{bmatrix} \\
\kappa_2^t &\sim \begin{bmatrix} (\vec{b}_1) & (\vec{b}_2) & (\vec{b}_3) \end{bmatrix} \\
\kappa_1^t \vee \kappa_2^t &\sim \begin{bmatrix} (\vec{a}_1) & (\vec{a}_2 + \vec{b}_2) & (\vec{b}_3) \end{bmatrix}
\end{aligned} \tag{A2}
$$

The resulting shared and redundant element is shown in Equation (A3). Due to the construction of the meet element through an inclusion–exclusion principle with the joint, the meet element always contains the vectors which span the redundant decision region as the first component.

$$
\begin{aligned}
\kappa_1^t \wedge \kappa_2^t &\sim \begin{bmatrix} (\vec{b}_1) & (\vec{a}_3) & (\vec{b}_2) & (\vec{a}_2) & -(\vec{a}_2 + \vec{b}_2) \end{bmatrix} \\
\kappa_1^t \sqcap \kappa_2^t &\sim \begin{bmatrix} (\vec{b}_1) & (\vec{a}_3) \end{bmatrix}
\end{aligned} \tag{A3}
$$

The second component of the meet element corresponds to the decision region of the joint, which is not part of either individual channel. This component is non-negative due to the triangle inequality.

$$
0 \leq f(\kappa_1^t \wedge \kappa_2^t) - f(\kappa_1^t \sqcap \kappa_2^t) = r(\vec{a}_2) + r(\vec{b}_2) - r(\vec{a}_2 + \vec{b}_2) \tag{A4}
$$

The same argument applies to the meet for an arbitrary number of channels since the inclusion–exclusion principle with the joint elements ensures that the vectors spanning the redundant region are contained in the meet element, and the triangle inequality ensures non-negativity for the additional components.
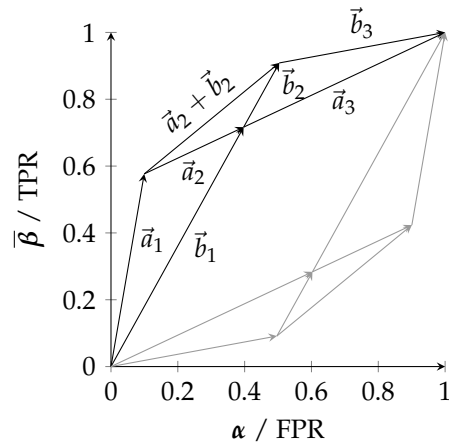
Figure A2. A minimal example to discuss the relation between the shared $(\kappa_1^t \wedge \kappa_2^t)$ and redundant $(\kappa_1^t \sqcap \kappa_2^t)$ decision regions. The channel κ_1^t consists of the vectors $\vec{a}_x$, and the channel κ_2^t consists of the vectors $\vec{b}_x$.

Lemma A2. *The function $f(\alpha^t)$ is a (consistent) valuation $\alpha^t \wedge \beta^t \sim \beta^t \Rightarrow f(\beta^t) \leq f(\alpha^t)$ on the pointwise lattice corresponding to $(\mathcal{B}^t(\mathbf{V}), \wedge, \vee)$, as visualized in Appendix A.*

Proof. Let $\mathbf{S}^t = \{\kappa_1^t, \ldots, \kappa_a^t\}$ represent a set of pointwise channels. The meet element $(\bigwedge_{\lambda^t \in \mathbf{S}^t} \lambda^t)$ is constructed through an inclusion–exclusion principle with the joint (convex hull). This ensures that the set of vectors spanning the zonotope intersection $(\bigsqcap_{\lambda^t \in \mathbf{S}^t} \lambda^t)$ is contained within the meet element. Additionally, the meet contains a second component that is ensured to be positive from the triangle inequality of r: $f(\bigwedge_{\lambda^t \in \mathbf{S}^t} \lambda^t) \geq f(\bigsqcap_{\lambda^t \in \mathbf{S}^t} \lambda^t)$. Since the joint operator is closed on channels and is distributive, we can introduce a channel to enforce a minimal redundant decision region between the channels: $f(\kappa_0^t) \leq f(\bigsqcap_{\lambda^t \in \mathbf{S}^t} \kappa_0^t \sqcup \lambda^t) \leq f(\bigwedge_{\lambda^t \in \mathbf{S}^t} \kappa_0^t \vee \lambda^t) = f(\kappa_0^t \vee \bigwedge_{\lambda^t \in \mathbf{S}^t} \lambda^t)$. Applying the sum-rule shows that $f(\kappa_0^t \wedge \bigwedge_{\lambda^t \in \mathbf{S}^t} \lambda^t) \leq f(\bigwedge_{\lambda^t \in \mathbf{S}^t} \lambda^t)$.

　　We again make use of the distributive property, which allows writing any expression α^t into a conjunctive normal form. Since the joint operator is closed for channels, any expression α^t can be represented as meet for a set of channels $\alpha^t \sim \bigwedge_{\lambda^t \in \{\kappa_{p_1}^t, \ldots, \kappa_{p_i}^t\}} \lambda^t$. This demonstrates that the obtained inequality of the meet operator on channels also applies to atoms $f(\alpha^t \wedge \beta^t) \leq f(\alpha^t)$, such that $\alpha^t \wedge \beta^t \sim \beta^t \Rightarrow f(\beta^t) \leq f(\alpha^t)$. □

Appendix D

　　The considered function $f_p(\kappa^t)$ of Section 3.2 takes the sum of a convex function. The Hessian matrix $\mathbf{H}_r$ of the function $r_p(x, y) = x \log_b\left(\frac{x}{px + (1-p)y}\right)$ is positive-semidefinite in the required domain (symmetric and its eigenvalues e_1 and e_2 are greater than or equal to zero for $x > 0$ and $b > 1$).

$$
\mathbf{H}_r = \frac{1}{\log(b)} \begin{bmatrix} \frac{(p-1)^2 y^2}{x(px+(1-p)y)^2} & -\frac{(p-1)^2 y}{(px+(1-p)y)^2} \\ -\frac{(p-1)^2 y}{(px+(1-p)y)^2} & \frac{(p-1)^2 x}{(px+(1-p)y)^2} \end{bmatrix}
$$

$$
e_1 = 0
$$

$$
e_2 = \frac{(p-1)^2 (x^2 + y^2)}{x \log(b)(px + (1-p)y)^2}
$$

(A5)

Appendix E

We use the examples of Finn and Lizier [20] since they provided an extensive discussion of their motivation. We compare our decomposition results to $I_{\min}$ of Williams and Beer [5] and $I^{\pm}$ of Finn and Lizier [20]. Examples with two sources are additionally compared to I^{BROJA} of Bertschinger et al. [10] and Griffith and Koch [11]. We notate the results for shared information $S(V_1, V_2; T)$, unique information $U(V_x; T)$ and synergetic/complementing information $C(V_1, V_2; T)$. We use the implementation of $I_{\min}$, I^{BROJA} and $I^{\pm}$ provided by the `dit` Python package for discrete information theory [24].

Notice that our approach is identical to Williams and Beer [5] if one of the variables is pointwise (for each T^t, not necessarily the same one each time) Blackwell superior to another, and that our approach is equal to Bertschinger et al. [10] and Griffith and Koch [11] for two visible variables at a *binary* target variable.

We would like to highlight Table A1 for the difference in our approach to Williams and Beer [5]. This is an arbitrary example, where the variables V_1 and V_2 are not Blackwell superior to each other from the perspective of T^t, as visualized in Figure 6. For highlighting the difference in our approach to Bertschinger et al. [10] and Griffith and Koch [11], we require an example where the target variable is not binary, such as the two-bit copy example in Table A2.
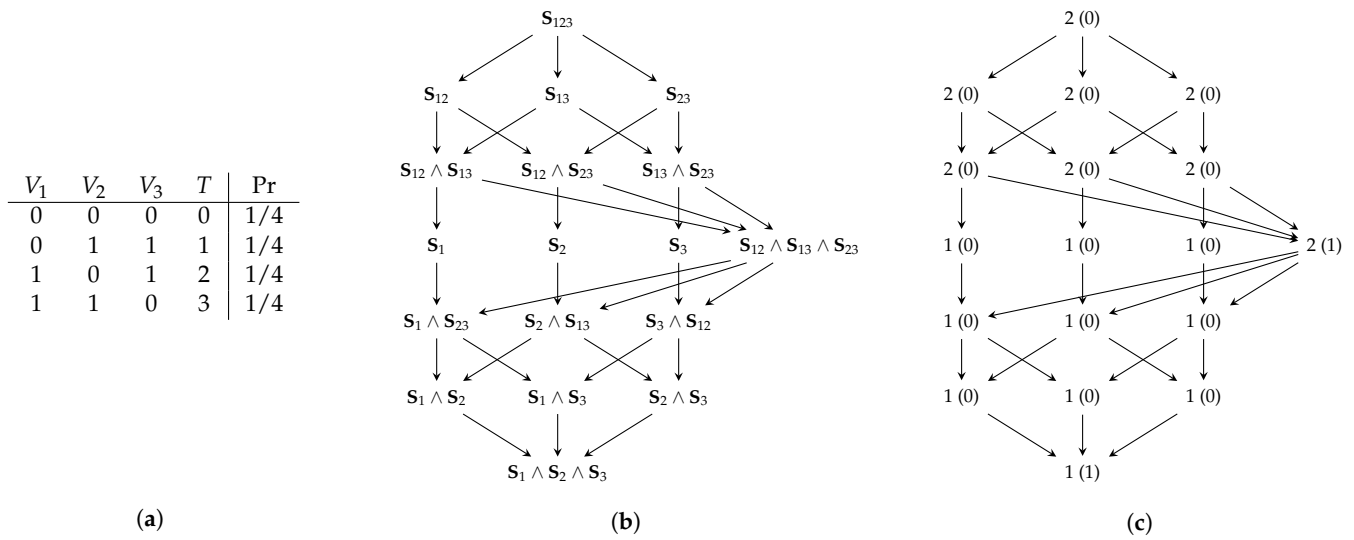
It can be seen that our approach does not satisfy the identity axiom of Harder et al. [8]. This axiom demands the decomposition of the two-bit-copy example (Table A2) to both variables providing one bit unique information and demands negative partial contributions in the three-bit even-parity example (Figure A3) [8,20].

Table A1. Two incomparable channels (visualized in Section 3.1). The table highlights the difference in our approach to Williams and Beer [5] while being identical to Bertschinger et al. [10] since the target variable is binary.

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	0	0	0.0625	$\hat{I}_T \cdot H(T)$	0.1196	0.0272	0.0716	0.1205
0	0	1	0.3	$I_{\min}$ [5]	0.1468	0	0.0444	0.1477
1	0	0	0.0375	$I^{\pm}$ [20]	0.3214	−0.1746	−0.1302	0.3223
1	0	1	0.05	I^{BROJA} [10,11]	0.1196	0.0272	0.0716	0.1205
0	1	0	0.1875					
0	1	1	0.15					
1	1	0	0.2125					

Table A2. Two-bit-copy (TBC) example. The results of our approach differ from Bertschinger et al. [10] and Griffith and Koch [11] since the target variable is not binary.

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	0	0	1/4	$\hat{I}_T \cdot H(T)$	1	0	0	1
0	1	1	1/4	$I_{\min}$ [5]	1	0	0	1
1	0	2	1/4	$I^{\pm}$ [20]	1	0	0	1
1	1	3	1/4	I^{BROJA} [10,11]	0	1	1	0

V_1	V_2	V_3	T	Pr
0	0	0	0	1/4
0	1	1	1	1/4
1	0	1	2	1/4
1	1	0	3	1/4

(**a**)　　　　　　　　(**b**)　　　　　　　　(**c**)

Figure A3. Three-bit even-parity (Tbep) example. The results for $\hat{I}_T \cdot H(T)$, $I_{\min}$ and $I^{\pm}$ are identical. (**a**) Distribution. (**b**) Decomposition lattice. (**c**) Cumulative results (partial).

Table A3. XOR-gate (Xor) example. All compared measures provide the same results.

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	0	0	1/4	$\hat{I}_T \cdot H(T)$	0	0	0	1
0	1	1	1/4	$I_{\min}$ [5]	0	0	0	1
1	0	1	1/4	$I^{\pm}$ [20]	0	0	0	1
1	1	0	1/4	I^{BROJA} [10,11]	0	0	0	1

Table A4. Pointwise unique (PwUnq) example. Our approach provides the same results as Williams and Beer [5] and Bertschinger et al. [10].

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	1	0	1/4	$\hat{I}_T \cdot H(T)$	0.5	0	0	0.5
1	0	0	1/4	$I_{\min}$ [5]	0.5	0	0	0.5
0	2	1	1/4	$I^{\pm}$ [20]	0	0.5	0.5	0
2	0	1	1/4	I^{BROJA} [10,11]	0.5	0	0	0.5

Table A5. Redundant Error (RdnErr) example. Our approach provides the same results as Williams and Beer [5] and Bertschinger et al. [10].

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	0	0	3/8	$\hat{I}_T \cdot H(T)$	0.189	0.811	0	0
1	1	1	3/8	$I_{\min}$ [5]	0.189	0.811	0	0
0	1	0	1/8	$I^{\pm}$ [20]	1	0	−0.811	0.811
1	0	1	1/8	I^{BROJA} [10,11]	0.189	0.811	0	0

Table A6. Unique (Unq) example. Our approach provides the same results as Williams and Beer [5] and Bertschinger et al. [10].

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	0	0	1/4	$\hat{I}_T \cdot H(T)$	0	1	0	0
0	1	0	1/4	$I_{\min}$ [5]	0	1	0	0
1	0	1	1/4	$I^{\pm}$ [20]	1	0	−1	1
1	1	1	1/4	I^{BROJA} [10,11]	0	1	0	0

Table A7. And-gate (And) example. Our approach provides the same results as Williams and Beer [5] and Bertschinger et al. [10].

(a) Distribution				(b) Results				
V_1	V_2	T	Pr	Method	$S(V_1, V_2; T)$	$U(V_1; T)$	$U(V_2; T)$	$C(V_1, V_2; T)$
0	0	0	1/4	$\hat{I}_T \cdot H(T)$	0.311	0	0	0.5
0	1	0	1/4	$I_{\min}$ [5]	0.311	0	0	0.5
1	0	0	1/4	$I^{\pm}$ [20]	0.561	−0.25	−0.25	0.75
1	1	1	1/4	I^{BROJA} [10,11]	0.311	0	0	0.5

References

1. Lizier, J.T.; Bertschinger, N.; Jost, J.; Wibral, M. Information Decomposition of Target Effects from Multi-Source Interactions: Perspectives on Previous, Current and Future Work. *Entropy* **2018**, *20*, 307. [CrossRef] [PubMed]
2. Wibral, M.; Finn, C.; Wollstadt, P.; Lizier, J.T.; Priesemann, V. Quantifying Information Modification in Developing Neural Networks via Partial Information Decomposition. *Entropy* **2017**, *19*, 494. [CrossRef]
3. Rassouli, B.; Rosas, F.E.; Gündüz, D. Data Disclosure Under Perfect Sample Privacy. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 2012–2025. [CrossRef]
4. Rosas, F.E.; Mediano, P.A.M.; Rassouli, B.; Barrett, A.B. An operational information decomposition via synergistic disclosure. *J. Phys. Math. Theor.* **2020**, *53*, 485001. [CrossRef]
5. Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. *arXiv* **2010**, arXiv:1004.2515.
6. Griffith, V.; Chong, E.K.P.; James, R.G.; Ellison, C.J.; Crutchfield, J.P. Intersection Information Based on Common Randomness. *Entropy* **2014**, *16*, 1985–2000. [CrossRef]
7. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J. Shared Information—New Insights and Problems in Decomposing Information in Complex Systems. In *Proceedings of the European Conference on Complex Systems 2012*; Gilbert, T., Kirkilionis, M., Nicolis, G., Eds.; Springer: Cham, Switzerland, 2013; pp. 251–269.
8. Harder, M.; Salge, C.; Polani, D. Bivariate measure of redundant information. *Phys. Rev. E* **2013**, *87*, 012130. [CrossRef] [PubMed]
9. Finn, C. A New Framework for Decomposing Multivariate Information. Ph.D. Thesis, University of Sydney, Sydney, NSW, Australia, 2019.
10. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* **2014**, *16*, 2161–2183. [CrossRef]
11. Griffith, V.; Koch, C. Quantifying Synergistic Mutual Information. In *Guided Self-Organization: Inception*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 159–190. [CrossRef]
12. Bertschinger, N.; Rauh, J. The Blackwell Relation Defines No Lattice. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 2479–2483. [CrossRef]
13. Knuth, K.H. Lattices and Their Consistent Quantification. *Ann. Der Phys.* **2019**, *531*, 1700370. [CrossRef]
14. Blackwell, D. Equivalent Comparisons of Experiments. In *The Annals of Mathematical Statistics*; Institute of Mathematical Statistics: Beachwood, OH, USA, 1953; pp. 265–272.
15. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
16. Schechtman, E.; Schechtman, G. The relationship between Gini terminology and the ROC curve. *Metron* **2019**, *77*, 171–178. [CrossRef]
17. Neyman, J.; Pearson, E.S. IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philos. Trans. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1933**, *231*, 289–337.
18. Mages, T.; Rohner, C. Implementation: PID Quantifying Reachable Decision Regions. 2023. Available online: https://github.com/uu-core/pid-quantifying-reachable-decision-regions (accessed on 1 May 2023).
19. Niu, X.; Quinn, C.J. Information Flow in Markov Chains. In Proceedings of the 2021 60th IEEE Conference on Decision and Control (CDC), Austin, TX, USA, 14–17 December 2021; pp. 3442–3447. [CrossRef]

20. Finn, C.; Lizier, J.T. Pointwise Partial Information Decomposition Using the Specificity and Ambiguity Lattices. *Entropy* **2018**, *20*, 297. [CrossRef] [PubMed]
21. Kolchinsky, A. A Novel Approach to the Partial Information Decomposition. *Entropy* **2022**, *24*, 403. [CrossRef]
22. Chicharro, D.; Panzeri, S. Synergy and Redundancy in Dual Decompositions of Mutual Information Gain and Information Loss. *Entropy* **2017**, *19*, 71. [CrossRef]
23. Rauh, J.; Bertschinger, N.; Olbrich, E.; Jost, J. Reconsidering Unique Information: Towards a Multivariate Information Decomposition. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 2232–2236. [CrossRef]
24. James, R.G.; Ellison, C.J.; Crutchfield, J.P. dit: A Python package for discrete information theory. *J. Open Source Softw.* **2018**, *3*, 738. [CrossRef]