



City Research Online

City St George's, University of London

Citation: Makkeh, A., Chicharro, D., Theis, D. O. & Vicente, R. (2019). MAXENT3D_PID: An estimator for the maximum-entropy trivariate partial information decomposition. *Entropy*, 21(9), 862. doi: 10.3390/e21090862

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/27165/>

Link to published version: <https://doi.org/10.3390/e21090862>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Article

MAXENT3D_PID: An Estimator for the Maximum-Entropy Trivariate Partial Information Decomposition

Abdullah Makkeh ^{1,*}, Daniel Chicharro ^{2,†,‡}, Dirk Oliver Theis ^{1,†} and Raul Vicente ^{1,†}

¹ Institute of Computer Science, University of Tartu, 51014 Tartu, Estonia

² Neural Computation Laboratory, Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, 38068 Rovereto (TN), Italy

* Correspondence: makkeh@ut.ee

† These authors contributed equally to this work.

‡ Current address: Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA.

Received: 28 June 2019; Accepted: 27 August 2019; Published: 3 September 2019

Abstract: Partial information decomposition (PID) separates the contributions of sources about a target into unique, redundant, and synergistic components of information. In essence, PID answers the question of “who knows what” of a system of random variables and hence has applications to a wide spectrum of fields ranging from social to biological sciences. The paper presents MAXENT3D_PID, an algorithm that computes the PID of three sources, based on a recently-proposed maximum entropy measure, using convex optimization (cone programming). We describe the algorithm and its associated software utilization and report the results of various experiments assessing its accuracy. Moreover, the paper shows that a hierarchy of bivariate and trivariate PID allows obtaining the finer quantities of the trivariate partial information measure.

Keywords: multivariate information theory; partial information decomposition; cone programming; synergy; redundancy; PYTHON

1. Introduction: Motivation and Significance

The characterization of dependencies within complex multivariate systems helps to identify the mechanisms operating in the system and understanding their function. Recent work has developed methods to characterize multivariate interactions by separating n -variate dependencies for different orders n [1–5]. In particular, the work of Williams and Beer [6,7] introduced a framework, called partial information decomposition (PID), which quantifies whether different input variables provide redundant, unique, or synergistic information about an output variable when combined with other input variables. Intuitively, inputs are redundant if each carries individually information about the same aspects of the output. Information is unique if it is not carried by any other single (or group of) variables, and synergistic information can only be retrieved combining several inputs.

This information-theoretic approach to study interactions has found many applications to complex systems such as gene networks (e.g., [8–10]), interactive agents (e.g., [11–14]), or neural processing (e.g., [15–17]). More generally, the nature of the information contained in the inputs determines the complexity of extracting it [18,19], how robust it is to disrupt the system [20], or how input dimensionality can be reduced without information loss [21,22].

Despite this great potential, the applicability of the PID framework has been hindered by the lack of agreement on the definition of a suitable measure of redundancy. In particular, Harder et al. [23] indicated that the original measure proposed by [6] only quantifies common amounts of information, instead of shared information that is qualitatively the same. A constellation of

measures has been proposed to implement the PID (e.g., [23–29]), and core properties, such as requiring nonnegativity as a property of the measures, are still the subject of debate [29–32].

A widespread application of the PID framework has also been limited by the lack of multivariate implementations. Some of the proposed measures were only defined for the bivariate case [23,24,33]. Other multivariate measures allow negative components in the PID [26,29], which, although it may be adequate for statistical characterization of dependencies, limits the interpretation of the information-theoretic quantities in terms of information communication [34]. Even though at the level of local information, negativity is regarded as misinformation and can be interpreted, for example, operationally in terms of changes in belief [35], when considering information in the context of communication [36], then interpreting it as the number of messages to be retrieved without error through a noisy channel requires nonnegativity; for example, assessing the information representation about multidimensional sensory stimulus across neurons, in particular the analyses of the information content of neural responses [17,37]. Among the PID measures proposed, the maximum entropy measures of Bertschinger et al. [24] have a preeminent role in the bivariate case because they provide bounds for any other measure consistent with a set of properties shared by many of the proposed measures. Motivated by this special role of the maximum entropy measures, Chicharro [38] extended the maximum entropy approach to measures of the multivariate redundant information, which provide analogous bounds for the multivariate case. However, the work in [38] did not address their numerical implementation.

In this work, we present MAXENT3D_PID, a Python module that computes a trivariate information decomposition following the maximum entropy PID of [38] and exploits the connection with the bivariate decompositions associated with the trivariate ones [28]. This is, to our knowledge, the first available implementation of the maximum-entropy PID framework beyond the bivariate case [39–42], see Appendix B. This implementation is relevant for the theoretical development and practical use of the PID framework.

From a theoretical point of view, this implementation will provide the possibility to test the properties of the PID beyond the bivariate case. This is critical with regard to the nonnegativity property because, while nonnegativity is guaranteed in the bivariate case, for the multivariate case, it has been proven that negative terms can appear in the presence of deterministic dependencies [30,32,43]. However, the violation of nonnegativity has only been proven with isolated counterexamples, and it is not understood which properties of a system's dependencies lead to negative PID measures.

From a practical point of view, the trivariate PID allows studying new types of distributed information that only appear beyond the bivariate case, such as information that is redundant for two inputs and unique with respect to a third [6]. This extension is significant both to study multivariate systems directly, as well as to be exploited for data analysis [21,44]. As mentioned above, the characterization of synergy and redundancy in multivariate systems is relevant for a broad range of fields that encompass social and biological systems. So far, the PID has particularly found applications in neuroscience (e.g., [17,37,45–48]). For data analysis, the quantification of multivariate redundancy can be applied to dimensionality reduction [22] or to better understand how representations emerge in neural networks during learning [49,50]. Altogether, this software promises to contribute significantly to the refinement of the information-theoretic tools it implements and also to foster its widespread application to analyze data from multivariate systems.

2. Models and Software

The section starts by briefly describing the mathematical model of the problem. Then, it discusses the architecture of MAXENT3D_PID. It closes by explaining in details of how to use the software.

2.1. Maximum Entropy Decomposition Measure

Consider \mathbf{X} , \mathbf{Y} , and \mathbf{Z} as the sources and \mathbf{T} as the target of some system. Let P be the joint distribution of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ and $MI(\mathbf{T}; \mathcal{S})$ be the mutual information of \mathbf{T} and \mathcal{S} , where \mathcal{S} is any

nonempty subset of $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. The PID decomposes $MI(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ into finer parts, namely synergistic, unique, redundant unique, and redundant information. These finer parts respect certain identities [6], e.g., a subset of them sums up to $MI(\mathbf{T}, \mathbf{X})$ (all identities are explained in Appendices A and C). Following the maximum entropy approach [24], to obtain this decomposition, it is necessary to solve the following optimization problems:

$$\min_{\Delta_P} MI(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \quad (1a)$$

$$\min_{\Delta_P} MI(\mathbf{T}; \mathbf{X}_1, \mathbf{X}_2) \quad \text{for } \mathbf{X}_1, \mathbf{X}_2 \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} \quad (1b)$$

where:

$$\Delta_P = \{Q \in \Delta : Q(\mathbf{T}, \mathbf{X}) = P(\mathbf{T}, \mathbf{X}), Q(\mathbf{T}, \mathbf{Y}) = P(\mathbf{T}, \mathbf{Y}), \\ Q(\mathbf{T}, \mathbf{Z}) = P(\mathbf{T}, \mathbf{Z})\}$$

and Δ is the set of all joint distributions of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. The four minimization problems in Equation (1a,b) can be formulated as exponential cone programs, a special case of convex optimization. The authors refer to [41] for a nutshell introduction to cone programs, in particular the exponential ones. The full details on how to formulate (1a,b) as exponential cone programs and their convergence properties are explained in [51] (Chapter 5).

MAXENT3D_PID on its own returns the synergistic information and unique information collectively. In addition, with the help of the bivariate solver [39] (used in a specific way), the finer synergistic and unique information can also be extracted. Hence, the presented model obtains all the trivariate PID quantities. The full details for recovering the finer parts can be found in Appendices C and D.

2.2. Software Architecture and Functionality

MAXENT3D_PID is implemented using the standard PYTHON syntax. The module uses an optimization software ECOS [52] to solve several optimization problems needed to compute the trivariate PID. To install the module, the ECOS Python package has to be installed, and then from the GITHUB repository, the files MAXENT3D_PID.py, TRIVARIATE_SYN.py, TRIVARIATE_UNQ.py, and TRIVARIATE_QP.py must be downloaded [53].

MAXENT3D_PID has two Python classes Solve_w_ECOS and QP. Class Solve_w_ECOS receives the marginal distributions of (\mathbf{T}, \mathbf{X}) , (\mathbf{T}, \mathbf{Y}) , and (\mathbf{T}, \mathbf{Z}) as Python dictionaries. These distributions are used by Solve_w_ECOS sub-classes Opt_I and Opt_II to solve the optimization problems of Equation (1a,b) respectively. The class QP is used to recover the solution of any optimization problems of Equation (1a,b) when Solve_w_ECOS fails to obtain a solution of a good quality. Figure 1 gives an overview of how these two classes interact.

2.2.1. The Subclass Opt_I and Opt_II

The sub-classes Opt_I and Opt_II formulate the problems Equation (1a,b), use ECOS to get the optimal values, and compute their violations of the optimality certificates. They return the optimal values and their optimality violations. These violations are quality measures of the obtained PID. Figure 1 describes this process within the class Solve_w_ECOS. Note that both sub-classes Opt_I and Opt_II optimize conditional entropy functionals; however, the different number of arguments leads to a difference in how to fit the problems into the cone program and retrieving the optimal solution; hence the requirement of splitting them into different classes.

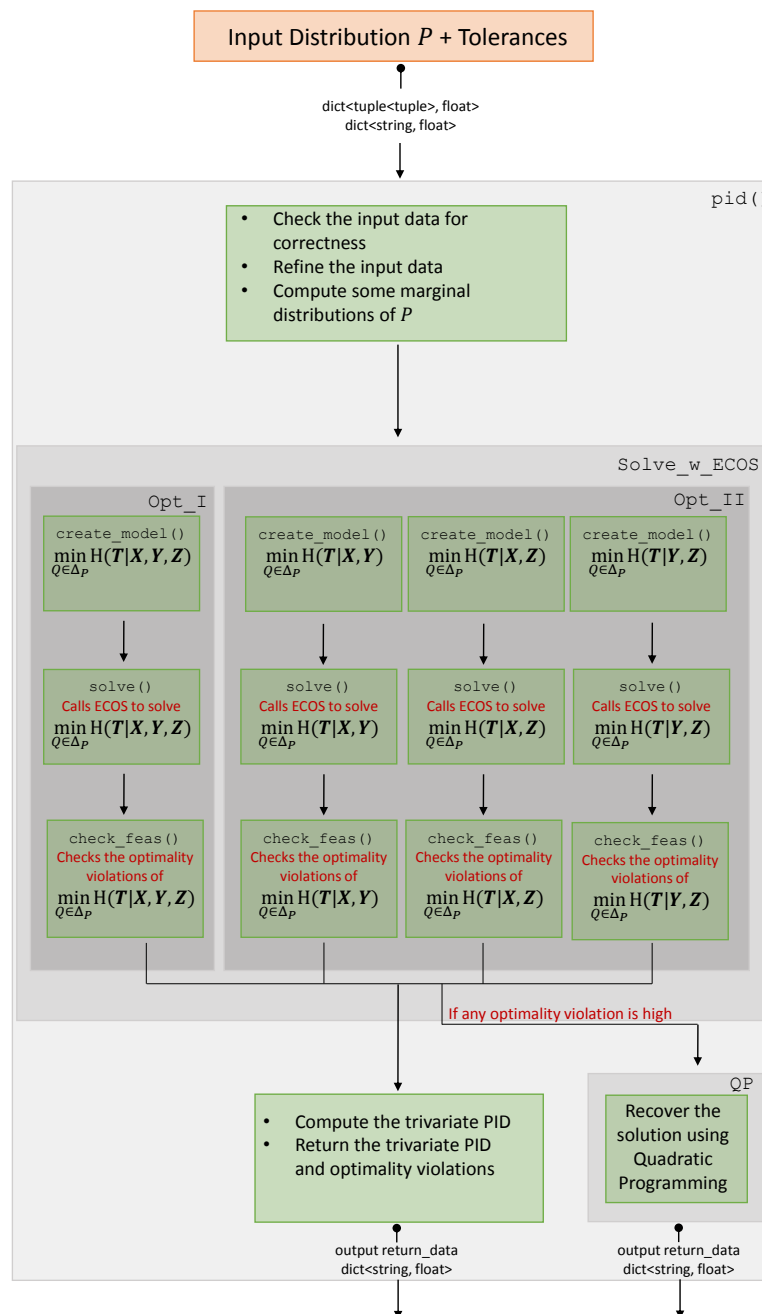


Figure 1. A flowchart describing the process of computing the trivariate PID via MAXENT3D_PID. It gives an overview of how pid() utilizes the classes Solve_w_ECOS and QP in the aim of computing the trivariate PID.

2.2.2. The Class QP

Class QP acts if Solve_w_ECOS returns values of a subset of Equation (1a,b) with high optimality violations. It improves the errant values by best fitting them using quadratic programming, where the PID identities (A12) are respected.

2.3. Using MAXENT3D_PID

The process of computing the PID is packed in the function pid(). This function takes as input the distribution P of (T, X, Y, Z) via a Python dictionary where the tuples (t, x, y, z) are keys and their

associated probability $P(t, x, y, z)$ is the value of the key; see Figure 2. The function formulates and solves the problems of (1a,b) using Solve_w_ECOS and, if needed, uses QP to improve the solution. This function pid() returns a Python dictionary, explained in Tables 1 and 2, containing the PID of (T, X, Y, Z) in addition to the optimality violations.

```
# The function pid() is imported from the module MAXENT3D_PID
from MAXENT3D_PID import ~pid

# The input distribution is defined as a Python dictionary
andDgate = dict()
andDgate[ (0,0,0,0) ] = .25
andDgate[ (0,0,1,0) ] = .25
andDgate[ (0,1,0,1) ] = .25
andDgate[ (1,1,1,1) ] = .25

# pid() is called
sol = pid(andDgate)
# printing the obtained PID
msg="""Synergistic information: {CI}
Unique information in X: {UIX}
Unique information in Y: {UIY}
Unique information in Z: {UIZ}
Unique information in X,Y: {UIXY}
Unique information in X,Z: {UIXZ}
Unique information in Y,Z: {UIYZ}
Shared information: {SI}"""
print(msg.format(**sol))
```

Figure 2. Using MAXENT3D_PID to compute the PID of the distribution obtained from the ANDDUPLICATE gate (andDgate). The ANDDUPLICATE gate evaluates T as the logical and of X and Y ($X \wedge Y$) such that Z copies X.

Table 1. The keys of the trivariate PID quantities in the returned dictionary. Note that $UI(T; X_i \setminus X_j, X_k)$ and $UI(T; X_i, X_k \setminus X_j)$ refer to unique and unique redundant information for $X_i, X_k, X_j \in \{X, Y, Z\}$, $CI(T; X, Y, Z)$ refers to synergistic information, and $SI(T; X, Y, Z)$ refers to redundant or shared information.

Keys	Values	Keys	Values
'UIX'	$UI(T; X \setminus Y, Z)$	'UIYZ'	$UI(T; Y, Z \setminus X)$
'UIY'	$UI(T; Y \setminus X, Z)$	'UIXZ'	$UI(T; X, Z \setminus Y)$
'UIZ'	$UI(T; Z \setminus X, Y)$	'UIXY'	$UI(T; X, Y \setminus Z)$
'CI'	$CI(T; X, Y, Z)$	'SI'	$SI(T; X, Y, Z)$

Table 2. The keys of optimality violations for each problem (1a,b) in the returned dictionary.

Key	Value
'Num_Err_I'	Optimality violations of $\min_{\Delta_P} MI(T; X, Y, Z)$
'Num_Err_12'	Optimality violations of $\min_{\Delta_P} MI(T; X, Y)$
'Num_Err_13'	Optimality violations of $\min_{\Delta_P} MI(T; X, Z)$
'Num_Err_23'	Optimality violations of $\min_{\Delta_P} MI(T; Y, Z)$

The function pid() has three other optional inputs. The first optional input is called parallel (the default value is parallel='off'), which determines whether the process will be parallelized. If parallel='off', then the process is going to be done sequentially, i.e., the four problems of Equation (1a,b) are going to be formulated and solved one after the other. Their optimality violations

are also computed consecutively, and then, final results are obtained; whereas, when `parallel='on'`, the formulation of the four problems Equation (1a,b) is done in parallel. The four problems are solved simultaneously, and finally, the optimality violations along with the final results are computed in parallel. Thus, when `parallel='on'`, there will be three sequential steps: formulating the problems, solving them, and obtaining the final results, as opposed to `parallel='off'`, which requires at least twelve sequential steps.

The second optional input is a dictionary that allows the user to tune the tolerances controlling the optimization routines of ECOS listed in Table 3.

Table 3. Parameters (tolerances) that govern the optimization in ECOS.

Parameter	Description	Default Value
<code>feastol</code>	primal/dual feasibility tolerance	10^{-7}
<code>abstol</code>	absolute tolerance on the duality gap	10^{-6}
<code>reltol</code>	relative tolerance on the duality gap	10^{-6}
<code>feastol_inacc</code>	primal/dual infeasibility <i>relaxed</i> tolerance	10^{-3}
<code>abstol_inacc</code>	absolute <i>relaxed</i> tolerance on the duality gap	10^{-4}
<code>reltol_inacc</code>	<i>relaxed</i> relative duality gap	10^{-4}
<code>max_iter</code>	maximum number of iterations that ECOS does	100

In this dictionary, the user only sets the parameters that will be tuned. For example, if the user wants to achieve high accuracy, then the parameters `abstol` and `reltol` should be small (e.g., 10^{-12}) and the parameter `max_iter` should be high (e.g., 1000). In Figure 3, it is shown how to modify the parameters. In this case, the solver will take longer to return the solution. For further details about the parameter's tuning, check [41].

```
# The function pid() is imported from the module MAXENT3D_PID
from MAXENT3D_PID import pid
# The input distribution is defined as a Python dictionary
andDgate = dict()
andDgate[ (0,0,0,0) ] = .25
andDgate[ (0,0,1,0) ] = .25
andDgate[ (0,1,0,1) ] = .25
andDgate[ (1,1,1,1) ] = .25
# The dictionary is defined to tune ECOS parameters
parms = dict()
parms['abstol'] = 1.e-12 # set abstol
parms['reltol'] = 1.e-12 # set reltol
parms['max_iters'] = 100 # set maximum iteration
# pid is called
pid(andDgate, parallel='on', **parms)
```

Figure 3. Tuning the parameters of ECOS.

The third optional input is called `output`, and it controls what `pid()` will print on the user's screen. This optional input is explained in Table 4.

Table 4. Description of the printing modes in the function pid().

Value	Description
0 (default)	Simple Mode: pid() prints its output (Python dictionary).
1	Time Mode: In addition to what is printed when output=0, pid() prints a flag when it starts preparing the optimization problems in Equation (1a,b), the total time to create each problem, a flag when it calls ECOS, brief stats from ECOS of each problem after solving it (Figure 4), the total time for retrieving the results, the total time for computing the optimality violations, and the total time to store the results.
2	Detailed Time Mode: In addition to what is printed when output=0, pid() prints for each problem the time of each major step of creating the model, brief stats from ECOS of each problem after solving it, the total time of each function used for retrieving the results, the time of each major step used to compute the optimality violations, the time of each function used to obtain the final results, and the total time to store the results.
3	Detailed Optimization Mode: In addition to what is printed when output=1, pid() prints ECOS detailed stats of each problem after solving it (Figure 5).

```
MAXENT3D_PID.pid(): Stats for optimizing H(S|X,Y,Z):
{'exitFlag': 0, 'pcost': -0.3465735936653011, 'dcost': -0.3465735930504127,
'pres': 2.8092696654527348e-09, 'dres': 2.7131747078266765e-10, 'pinf': 0.0,
'dinf': 0.0, 'pinfres': nan, 'dinfres': 0.43220650722288695, 'gap':
6.4243546975262245e-09, 'relgap': 1.8536769144998566e-08, 'r0': 1e-08,
'iter': 19, 'mi_iter': -1, 'infostring': 'Optimal solution found', 'timing':
{'runtime': 0.000498888, 'tsetup': 9.6448e-05, 'tsolve': 0.00040244},
'numerr': 0}
```

Figure 4. Brief stats from ECOS after solving Problem (1a).

```
ECOS 2.0.4 \hl{-} (C) embotech GmbH, Zurich Switzerland, 2012--2015. Web: \hl{www.embotech.com/ECOS}
It pcost dcost gap pres dres k/t mu step sigma IR | BT
0 +0.000e+00 -0.000e+00 +4e+02 1e+00 3e-01 1e+00 1e+00 --- --- 0 0 - | - -
1 -3.608e+00 -2.696e+00 +1e+01 9e-01 2e-02 1e+00 3e-02 0.9791 9e-03 1 1 1 | 1 0
2 -2.806e+00 -2.611e+00 +3e+00 3e-01 5e-03 2e-01 7e-03 0.7833 1e-02 1 1 1 | 1 1
3 -1.863e+00 -1.822e+00 +5e-01 5e-02 9e-04 4e-02 1e-03 0.7833 5e-03 1 1 1 | 0 1
4 -7.797e-01 -7.704e-01 +8e-02 1e-02 2e-04 1e-02 2e-04 0.7833 9e-03 1 1 1 | 1 1
5 -7.319e-01 -7.299e-01 +2e-02 3e-03 5e-05 2e-03 5e-05 0.7833 9e-04 1 1 1 | 0 1
6 -6.969e-01 -6.964e-01 +4e-03 7e-04 1e-05 6e-04 1e-05 0.7833 9e-03 1 1 1 | 1 1
7 -6.950e-01 -6.949e-01 +9e-04 1e-04 3e-06 1e-04 2e-06 0.7833 1e-03 1 1 1 | 0 1
8 -6.933e-01 -6.933e-01 +2e-04 3e-05 6e-07 3e-05 5e-07 0.7833 9e-03 1 0 0 | 1 1
9 -6.932e-01 -6.932e-01 +4e-05 6e-06 1e-07 6e-06 1e-07 0.7833 8e-04 1 0 0 | 0 1
10 -6.932e-01 -6.932e-01 +8e-06 1e-06 3e-08 1e-06 2e-08 0.7833 9e-03 1 0 0 | 1 1
11 -6.932e-01 -6.932e-01 +2e-06 3e-07 6e-09 3e-07 6e-09 0.7833 1e-04 1 0 0 | 0 1
12 -6.931e-01 -6.931e-01 +4e-07 7e-08 1e-09 6e-08 1e-09 0.7833 9e-03 2 0 0 | 1 1
13 -6.931e-01 -6.931e-01 +3e-08 4e-09 4e-11 1e-09 8e-11 0.9791 9e-03 2 0 0 | 1 0
14 -6.931e-01 -6.931e-01 +7e-09 1e-09 3e-11 7e-10 2e-11 0.5013 2e-01 2 0 0 | 4 3
OPTIMAL (within feastol=1.3e-09, reltol=1.1e-08, abstol=7.4e-09).
Runtime: 0.002952 seconds.
```

Figure 5. Detailed stats from ECOS after solving Problem (1a).

3. Illustrations

This section shows some performance tests of MAXENT3D_PID on three types of instances. We will describe each type of instance and show the results of testing MAXENT3D_PID for each one of them. The first two types, paradigmatic and COPY gates, are used as validation and memory tests. The last type, random probability distributions, is used to evaluate the accuracy and efficiency of MAXENT3D_PID in computing the trivariate partial information decomposition. More precisely, accuracy is evaluated as how close the values of $UI(T; X \setminus Y, Z)$ and $UI(T; Y \setminus X, Z)$ are to zero when Z has a considerably higher dimension, which is expected theoretically. The efficiency will be depicted in how fast MAXENT3D_PID is able to produce the results. The machine used comes with an Intel(R) Core(TM) i7-4790K CPU (four cores) and 16 GB of RAM. Only the computations of the last type were done using parallelization.

3.1. Paradigmatic Gates

As a first test, we used some trivariate PIDs that are known and have been studied previously [25]. These examples are the logic gates collected in Table 5. For these examples, the decomposition can be derived analytically, and thus, they serve to check the numerical estimations.

Table 5. Paradigmatic gates with a brief explanation of their operation, where \oplus is the logical XOR and \wedge is the logical AND.

Instance	Operation
XORDUPLICATE	$\mathbf{T} = \mathbf{X} \oplus \mathbf{Y}; \mathbf{Z} = \mathbf{X}; \mathbf{X}, \mathbf{Y}$ i.i.d.
XORLOSES	$\mathbf{T} = \mathbf{X} \oplus \mathbf{Y}; \mathbf{Z} = \mathbf{X} \oplus \mathbf{Y}; \mathbf{X}, \mathbf{Y}$ i.i.d.
XORMULTICOAL	$\mathbf{T} = \mathbf{U} \oplus \mathbf{V} \oplus \mathbf{W}; \mathbf{X} = (\mathbf{U}, \mathbf{V}),$ $\mathbf{Y} = (\mathbf{U}, \mathbf{W}), \mathbf{Z} = (\mathbf{V}, \mathbf{W}); \mathbf{U}, \mathbf{V}, \mathbf{W}$ i.i.d.
ANDDUPLICATE	$\mathbf{T} = \mathbf{X} \wedge \mathbf{Y}; \mathbf{Z} = \mathbf{X}; \mathbf{X}, \mathbf{Y}$ i.i.d.

Testing

The test was implemented in `test_gates.py`. `MAXENT3D_PID` returns, for all gates, the same values as ([25], Table 1) up to a precision error of order 10^{-9} . The slowest solving time (not in parallel) was one millisecond.

3.2. Copy Gate

As a second test, we used the COPY gate example to examine the simulation of large systems. We simulated how the solver handled large systems in terms of speed and reliability. Reliability, in this context, is meant as the consistency of the measure on large systems and the degree to which the results can be trusted to be accurate enough.

The COPY gate is the mapping of (x, y, z) , chosen uniformly at random, to (t, x, y, z) , where $t = (x, y, z)$. The size of the joint distribution of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ scales as $|\mathbf{X}|^2 \cdot |\mathbf{Y}|^2 \cdot |\mathbf{Z}|^2$, where $x, y, z \in X \times Y \times Z$. In our test, $|\mathbf{X}| = \ell, |\mathbf{Y}| = m$ and $|\mathbf{Z}| = n$, where $\ell, m, n \in \{10, 20, \dots, 50\}$.

Since \mathbf{X}, \mathbf{Y} and \mathbf{Z} are independent, it is easy to see that the only nonzero quantities are $\text{UI}(\mathbf{T}; \mathbf{X}_1 \setminus \mathbf{X}_2, \mathbf{X}_3) = H(\mathbf{X}_1)$ for $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$.

Testing

The test was implemented in `test_copy_gate.py`. The slowest solving time was less than 100 s, and the worst deviation from the actual values was 0.0001%. For more details, see Table 6.

Table 6. Copy gate results. The results are divided into three sets ordered increasingly w.r.t. the size of the joint distributions. Dimensions capture the unordered triplet $(|X|, |Y|, |Z|)$, and the deviation is computed as the maximum over all PID quantities of $100|\tilde{r} - r|$ where \tilde{r} is the obtained PID quantity and r is the analytical PID quantity. Note that the theoretical results are either zero or $\log_2(|S|)$, where $S \in X, Y, Z$.

Set 1			Set 2			Set 3		
Dimensions	Time (s)	Deviation (%)	Dimensions	Time (s)	Deviation (%)	Dimensions	Time (s)	Deviation (%)
(10,10,10)	0.82	10^{-7}	(20,20,30)	7.53	10^{-6}	(30,30,40)	25.72	10^{-6}
(10,10,20)	1.06	10^{-7}	(10,30,50)	8.67	10^{-6}	(20,40,50)	24.32	10^{-5}
(10,10,30)	1.62	10^{-7}	(10,40,40)	8.68	10^{-7}	(30,30,50)	27.90	10^{-6}
(10,10,40)	2.08	10^{-7}	(20,20,40)	8.85	10^{-7}	(30,40,40)	29.85	10^{-6}
(10,20,20)	2.21	10^{-7}	(20,30,30)	11.41	10^{-6}	(20,50,50)	34.94	10^{-6}
(10,10,50)	2.61	10^{-6}	(10,40,50)	11.44	10^{-6}	(30,40,50)	47.40	10^{-5}
(10,20,30)	2.99	10^{-6}	(20,20,50)	11.34	10^{-6}	(40,40,40)	42.21	10^{-5}
(10,20,40)	4.11	10^{-6}	(20,30,40)	13.00	10^{-6}	(30,50,50)	55.60	10^{-4}
(20,20,20)	4.96	10^{-6}	(10,50,50)	16.37	10^{-7}	(40,40,50)	55.18	10^{-5}
(10,30,30)	4.43	10^{-7}	(30,30,30)	16.28	10^{-7}	(40,50,50)	89.58	10^{-6}
(10,20,50)	5.51	10^{-7}	(20,30,50)	17.24	10^{-7}	(50,50,50)	97.74	10^{-5}
(10,30,40)	6.51	10^{-6}	(20,40,40)	18.34	10^{-5}			

3.3. Random Probability Distributions

As a last example, we used joint distributions of (T, X, Y, Z) sampled uniformly at random over the probability space, to test the accuracy of the solver. The size of $T, X,$ and Y was fixed to two, whereas $|Z|$ varied in $\{2, \dots, 14\}$. For each $|Z|$, 500 joint distributions of (T, X, Y, Z) were sampled.

Testing

As $|Z|$ increased, the average value of $UI(T; X \setminus Y, Z)$ and of $UI(T; Y \setminus X, Z)$ decreased, while that of $UI(T; Z \setminus X, Y)$ increased. In Figure 6, the accuracy of the optimization is reflected in the low divergence from zero obtained for the unique information $UI(T; X \setminus Y, Z)$ and $UI(T; Y \setminus X, Z)$. In Figure 7, the time has a constant trend, and the highest time value recorded was 0.8 s.

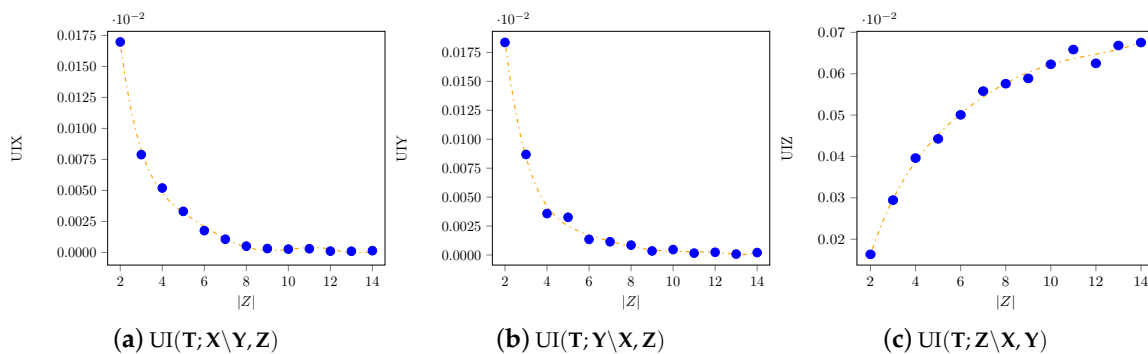


Figure 6. The variation of the unique information, as the size of Z increases, for the random probability distributions described in Section 3.3. It shows that the value of the unique information of Z increases as the dimension of Z increases.

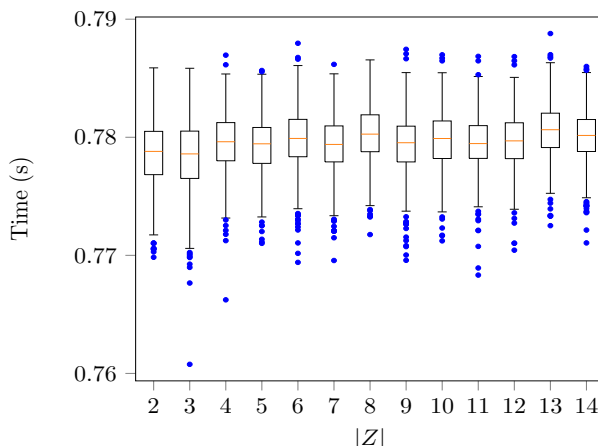


Figure 7. Box plotting of the time for MAXENT3D_PID to compute the PID of random joint probability distributions of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ for $|\mathbf{T}| = |\mathbf{X}| = |\mathbf{Y}| = 2$ and different sizes of \mathbf{Z} . For the size of the sets explored, the computational time shows a flat trend, and its variance is small.

3.4. Challenging Distributions

We tested MAXENT3D_PID on randomly uniformly-sampled distributions, but with large sizes of $T, X, Y,$ and Z . For each m , 500 joint distributions of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ were sampled where $|\mathbf{T}| = |\mathbf{X}| = |\mathbf{Y}| = |\mathbf{Z}| = m$ and $2 \leq m \leq 19$. The idea was to check with random and huge distributions (not structured as in the case of the COPY gate) how stable the estimator was.

3.4.1. Testing

For $m \geq 5$, some of the optimization problems (1a,b) did not converge due to numerical instabilities. This issue started to be frequent and significant when $m \geq 14$, for example 5% of the distributions had numerical problems in some of their optimization problems. We noticed that the returned solution from the non-convergent problem was feasible and far from optimal by a factor of 100 at most. The feasibility of the returned solution suggested fitting it along with the returned (optimal) solutions from the other convergent problems into the system of PID identities (A12), which will reduce the optimality gap.

3.4.2. Recommendation

These challenging distributions have mainly two features, namely the high dimensionality of the quadruple $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ and a significant number of relatively small (almost null) probability masses along with few concentrated probability masses. We suspect that these two features combined were the main reason for the convergence problems. Our approach was to use a quadratic programming (Class QP), which focuses on reducing the optimality gap and thus returns a close PID to the optimal PID (in case of no convergence problems).

Furthermore, we advise users to mitigate such distributions by dropping some of the points with almost null probability masses. Since the objective functions in (1a,b) are continuous and smooth (full support distributions) on Δ_P , then the PID of the mitigated distribution is considered a good approximation of that of the original distribution. Although we did not test this ad hoc on MAXENT3D_PID, the same technique was applied to such instances for BROJA_2PID ([51], Chapter 5).

We speculated that when $m \geq 50$, the solver will suffer dire numerical instabilities. It is recommended for the user to avoid large discrete binning resulting in humongous distributions.

3.4.3. Time Complexity

Theoretically, Makkeh et al. [39,51] showed that the worst running time complexity for solving (1a) (the hardest problem computationally) was $O(N^{3/2} \log N)$ where $N = |T \times X \times Y \times Z|$. Note that this time complexity bound was for the so-called barrier method, whereas ECOS uses the primal-dual Mehrotra predictor-corrector method [54], which does not have a theoretical complexity bound [55].

4. Summary and Discussion

In this work, we presented MAXENT3D_PID, a Python module that computes a trivariate decomposition based on the partial information decomposition (PID) framework of Williams and Beer [6], in particular following the maximum entropy PID of [38] and exploiting the connection with the bivariate decompositions associated with the trivariate one [28]. This is, to our knowledge, the first available implementation extending the maximum-entropy PID framework beyond the bivariate case [39–42].

The PID framework allows decomposing the information that a group of input variables has about a target variable into redundant, unique, and synergistic components. For the bivariate case, this results in decomposition with four components, quantifying the redundancy, synergy, and unique information of each of the two inputs. In the multivariate case, finer parts appear, which do not correspond to purely redundant or unique components. For example, the redundancy components of the multivariate decomposition can be interpreted based on local unfoldings when a new input is added, with each redundancy component unfolding into a component also redundant with the new variable and a component of unique redundancy with respect to it [38]. The PID analysis can qualitatively characterize the distribution of information beyond the standard mutual information measures [56] and has already been proven useful to study information in multivariate systems (e.g., [14,17,37,56–62]).

However, the definition of suited measures to quantify synergy and redundancy is still a subject of debate. From all the proposed PID measures, the maximum entropy measures by Bertschinger et al. [24] have a preeminent role in the bivariate case because they provide bounds to any other alternative measures that share fundamental properties related to the notions of redundancy and unique information. Chicharro [38] generalized the maximum entropy approach, proposing multivariate definitions of redundant information and showing that these measures implement the local unfolding of redundancy via hierarchically-related maximum entropy constraints. The package MAXENT3D_PID efficiently implemented the constrained information minimization operations involved in the calculation of the trivariate maximum-entropy PID decomposition. In Section 2, we described the architecture of the software, presented in detail the main function of the software that computes the PID along with its optional inputs, and described how to use it. In Section 3, we provided examples that verified that the software produced correct results on paradigmatic gates, simulated how the software scaled with large systems, and hinted to the accuracy of the software in estimating PID. In this section, we also presented challenging examples where the MAXENT3D_PID core optimizer had convergence problems and discussed our technique to retrieve an approximate PID and some suggestions to avoid such anomalies.

The possibility to calculate a trivariate decomposition of the mutual information represents a qualitative extension of the PID framework that goes beyond an incremental extension of the bivariate case, both regarding its theoretical development and its applicability. From a theoretical point of view, regarding the maximum-entropy approach, the multivariate case requires the introduction of new types of constraints in the information minimization that do not appear in the bivariate case (Section 2 and [38]). More generally, the trivariate decomposition allows further studying one of the key unsolved issues in the PID formulation, namely the requirement of the nonnegativity of the PID measures in the multivariate case.

In particular, Harder et al. [23] indicated that the original measure proposed by [6] only quantified common amounts of information and required new properties for the PID measures, to quantify qualitatively and not quantitatively how information is distributed. However, for the multivariate

case, these properties have been proven to be incompatible with guaranteeing nonnegativity, by using some counterexamples [30,32,43]. This led some subsequent proposals to define PID measures that either focus on the bivariate case [23,24] or do not require nonnegativity [26,29]. A multivariate formulation was desirable because the notions of synergy and redundancy are not restrained to the bivariate case, while nonnegativity is required for an interpretation of the measures in terms of information communication [34] and not only as a statistical description of the probability distributions. MAXENT3D_PID will allow systematically exploring when negative terms appear, beyond the currently-studied isolated counterexamples. Furthermore, it has been shown that in those counterexamples, the negative terms result from the criterion used to assign the information identity to different pieces of information when deterministic relations exist [32]. Therefore, a systematic analysis of the appearance of negative terms will provide a better understanding of how information identity is assigned when quantifying redundancy, which is fundamental to assess how the PID measures conform to the corresponding underlying concepts.

From a practical point of view, the trivariate decomposition allows studying qualitatively new types of distributed information, identifying finer parts of the information that the inputs have about the target, such as information that is redundant for two inputs and unique with respect to a third [6]. This is particularly useful when examining multivariate representations, such as the interactions between several genes [8,63] or characterizing the nature of coding in neural populations [64,65]. Furthermore, exploiting the connection between the bivariate and the trivariate decomposition due to the invariance of redundancy to context [28], MAXENT3D_PID also allows estimating the finer parts of the synergy component (Appendix D). This also offers a substantial extension in the applicability of the PID framework, in particular for the study of dynamical systems [66,67]. In particular, a question that requires a trivariate decomposition is how information transfer is distributed among multivariate dynamic processes. Information transfer is commonly quantified with the measure called transfer entropy [68–72], which calculates the conditional mutual information between the current state of a certain process Y and the past of another process X , given the past of Y and of any other processes Z that may also influence those two. In this case, by construction, the PID analysis should operate with three inputs corresponding to the pasts of X , Y , and Z . Transfer entropy is widely applied to study information flows between brain areas to characterize dynamic functional connectivity [73–75], and characterizing the synergy, redundancy, and unique information of these flows can provide further information about the degree of integration or segregation across brain areas [76].

More generally, the availability of software implementing the maximum entropy PID framework beyond the bivariate case promises to be useful in a wide range of fields in which interactions in multivariate systems are relevant, spanning the domain of social [12,77] and biological sciences [3,10,17,63]. Furthermore, the PID measures can also be used as a tool for data analysis and to characterize computational models. This comprises dimensionality reduction via synergy or redundancy minimization [19,22], the study of generative networks that emerge from information maximization constraints [78,79], or explaining the representations in deep networks [50].

The MAXENT3D_PID package presents several differences and advantages with respect to other software packages currently available to implement the PID framework. Regarding the maximum entropy approach, other packages only compute bivariate decompositions [39–42]. The dit package [42] also implements several other PID measures, including bivariate implementations for the measure of [23,27]. Among the multivariate decompositions, the ones using the measures I_{min} [6] or I_{MMI} [80] can readily be calculated with standard estimators of the mutual information. However, the former, as discussed above, only quantifies common amounts of information, while the latter is only valid for a certain type of data, namely multivariate Gaussian distributed. Software to estimate multivariate pointwise PIDs is also available [26,29,81]. However, as mentioned above, these measures by construction allow negative components, which may not be desirable for the interpretation of the decomposition, for example in the context of communication theory, and limits their applicability for data analysis in such regimes [22]. Altogether, MAXENT3D_PID is the first software that implements

the mutual information PID framework via hierarchically-related maximum entropy constraints, extending the bivariate case by efficiently computing the trivariate PID measures.

Author Contributions: Conceptualization, A.M., D.C., D.O.T., and R.V.; formal analysis, A.M. and D.C.; funding acquisition, D.C., D.O.T., and R.V.; investigation, A.M. and D.O.T.; methodology, A.M.; project administration, R.V.; software, A.M.; supervision, D.O.T.; validation, R.V.; visualization, A.M.; writing, original draft, A.M. and D.C.; writing, review and editing, A.M., D.C., D.O.T., and R.V.

Funding: This research was supported by the Estonian Research Council, ETAG (Eesti Teadusagentuur), through PUTExploratory Grant #620. D.C. was supported by the Fondation Bertarelli. R.V. also thanks the financial support from ETAG through the personal research grant PUT1476. We also gratefully acknowledge funding by the European Regional Development Fund through the Estonian Center of Excellence in IT, EXCITE.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

Appendix A. Williams–Beer PID Framework

In order to decompose $MI(\mathbf{T}, \mathbf{S})$ where \mathbf{T} is the target and \mathbf{S} are the sources, Williams and Beer [6] defined a set of axioms leading to what is known as the redundancy lattice (Figure A1). These axioms and lattice form the framework for partial information decomposition (PID) upon which all the exiting definitions of PID are formulated.

Appendix A.1. Williams–Beer Axioms

Suppose that a source A is a subset of \mathbf{S} and a collection α is a set of sources. A shorthand notation inspired by [38] will be used to represent the collection of sources; for example, if the system is $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$, then the collection of sources $\{\{X, Y\}, \{X, Z\}\}$ will be denoted as $XY.XZ$. [6] defined the following axioms that redundancy should comply with:

- Symmetry (**S**): $MI(\mathbf{T}; \alpha)$ is invariant with respect to the order of the sources in the collection.
- Self-redundancy (**SR**): The redundancy of a collection formed by a single source is equal to the mutual information of that source.
- Monotonicity (**M**): Adding sources to a collection can only decrease the redundancy of the resulting collection, and redundancy is kept constant when adding a superset of any of the existing sources.

Appendix A.2. The Redundancy Lattice

Williams and Beer [6] defined a lattice formed from the collections of sources. They used (**M**) to define the partial ordering between the collections. The axiom (**S**) reflects the fact that each atom of the lattice will represent a partial information decomposition quantity. More importantly, not all the collections of sources will be considered as atoms since adding a superset of any source to the examined system does not change redundancy, i.e., (**M**). The set of collections of sources included in the lattice which will form its atoms is defined as:

$$\mathcal{A}(\mathbf{S}) = \{\alpha \in \mathcal{P}(\mathbf{S}) - \{\emptyset\} : \forall A_i, A_j \in \alpha, A_i \not\subseteq A_j\}, \tag{A1}$$

where $\mathcal{P}(\mathbf{S})$ is the power set of \mathbf{S} . For this set of collections (atoms), the partial ordering relation that constructs the redundancy lattice is:

$$\forall \alpha, \beta \in \mathcal{A}(\mathbf{S}), (\alpha \preceq \beta \Leftrightarrow \forall B \in \beta, \exists A \in \alpha, A \subseteq B), \tag{A2}$$

i.e., for two collections α and β , $\alpha \preceq \beta$, if for each source in β , there is a source in α that is a subset of that source. In Figure A1, the bivariate and trivariate redundancy lattices are shown.

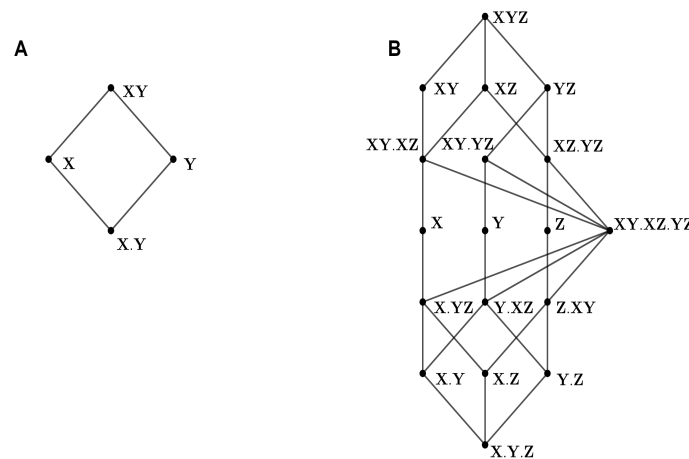


Figure A1. (A) Bivariate and (B) trivariate redundancy lattices. Letters indicate the mapping of terms between the lattices.

Appendix A.3. Defining PID over the Redundancy Lattice

The mutual information decomposition was constructed in [6] by implicitly defining partial information measures $\delta_C(\mathbf{T}; \alpha)$ associated with each node α of the redundancy lattice \mathcal{C} (Figure A1), such that the redundancy measures are obtained as:

$$MI(\mathbf{T}, \alpha) = \sum_{\beta \in \downarrow \alpha} \delta_C(\mathbf{T}; \beta), \tag{A3}$$

where $\downarrow \alpha$ refers to the set of collections lower than or equal to α in the partial ordering, and hence reachable descending from α in the lattice \mathcal{C} .

Appendix B. Bivariate Partial Information Decomposition

Let \mathbf{T} be the target random variable, \mathbf{X} and \mathbf{Y} be the two source random variables, and P be the joint probability distribution of $(\mathbf{T}, \mathbf{X}, \mathbf{Y})$. The PID captures the synergistic, unique, and redundant information as follows:

- The synergistic information between \mathbf{X} and \mathbf{Y} about \mathbf{T} , namely $CI(\mathbf{T}; \mathbf{X} : \mathbf{Y})$.
- The redundant information of \mathbf{X} and \mathbf{Y} about \mathbf{T} , namely $SI(\mathbf{T}; \mathbf{X}, \mathbf{Y})$.
- The unique information of \mathbf{X} about \mathbf{T} , namely $UI(\mathbf{T}; \mathbf{X} \setminus \mathbf{Y})$.
- The unique information of \mathbf{Y} about \mathbf{T} , namely $UI(\mathbf{T}; \mathbf{Y} \setminus \mathbf{X})$.

This decomposition, using Beer–Williams axioms, yields these identities:

$$\begin{aligned}
 MI(\mathbf{T}; \mathbf{X}, \mathbf{Y}) &= CI(\mathbf{T}; \mathbf{X} : \mathbf{Y}) + SI(\mathbf{T}; \mathbf{X}, \mathbf{Y}) + UI(\mathbf{T}; \mathbf{X} \setminus \mathbf{Y}) + UI(\mathbf{T}; \mathbf{Y} \setminus \mathbf{X}) \\
 MI(\mathbf{T}; \mathbf{X}_i) &= SI(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j) + UI(\mathbf{T}; \mathbf{X}_i \setminus \mathbf{X}_j) \quad \text{for all } \mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}\}.
 \end{aligned}
 \tag{A4}$$

Given the generic structure of the PID framework, the work in [24] (BROJA) defined PID measures considering the following polytope:

$$\Delta_P = \{Q \in \Delta : Q(\mathbf{T}, \mathbf{X}) = P(\mathbf{T}, \mathbf{X}), Q(\mathbf{T}, \mathbf{Y}) = P(\mathbf{T}, \mathbf{Y})\}, \tag{A5}$$

where Δ is the set of all joint distributions of $(\mathbf{T}, \mathbf{X}, \mathbf{Y})$. The work in [24] (BROJA) used the maximum entropy decomposition over Δ_P in order to quantify the above quantities. Moreover, BROJA assumed that the following assumptions hold.

Assumption A1 (Lemma 3 [24]). *On the bivariate redundancy lattice (Figure A1), the following assumptions must hold to quantify the PID*

1. All partial information measures of the redundancy lattice are nonnegative.
2. The terms $\delta(\mathbf{T}; \mathbf{X}, \mathbf{Y})$, $\delta(\mathbf{T}; \mathbf{X})$, and $\delta(\mathbf{T}; \mathbf{Y})$, are constant on Δ_P .
3. The synergistic term, namely $\delta(\mathbf{T}, \mathbf{X}, \mathbf{Y})$, vanishes on Δ_P upon minimizing the mutual information $\text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y})$.

Under the above assumptions and using maximal entropy decomposition, BROJA defined the following optimization problems that compute the PID quantities.

$$\widetilde{\text{CI}}(\mathbf{T}; \mathbf{X} : \mathbf{Y}) = \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{Y}, \mathbf{X}) \quad (\text{A6a})$$

$$\widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i \setminus \mathbf{X}_j) = \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}_j) \quad \text{for all } \mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}\} \quad (\text{A6b})$$

$$\widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) = \max_{Q \in \Delta_P} \text{CoI}(\mathbf{T}; \mathbf{X}; \mathbf{Y}) \quad (\text{A6c})$$

where $\text{CoI}(\mathbf{T}; \mathbf{X}; \mathbf{Y})$ is the co-information of \mathbf{T} , \mathbf{X} , and \mathbf{Y} defined as $\text{MI}(\mathbf{T}, \mathbf{X}) - \text{MI}(\mathbf{T}, \mathbf{X} | \mathbf{Y})$. Note that [38] proved that (A6c) is equivalent to:

$$\widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) = \min_{\substack{Q \in \Delta_P, \\ \text{CoI}(\mathbf{T}; \mathbf{X}; \mathbf{Y})=0}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}).$$

Appendix B.1. Mutual Information over the Bivariate Redundancy Lattice

This subsection writes down some mutual information quantities in terms of redundancy lattice partial information measures using (A3). These formulas will be used in the following subsection to verify that the measures defined in (A6a–c) to quantify the desired partial information quantities. $\text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y})$ will be the sum of partial information measure on every node of the redundancy lattice \mathcal{C} as follows:

$$\text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) = \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}) + \delta(\mathbf{T}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}). \quad (\text{A7})$$

The mutual information of one source and the target is expressed as:

$$\text{MI}(\mathbf{T}; \mathbf{X}_i) = \delta(\mathbf{T}, \mathbf{X}_i) + \delta(\mathbf{T}, \mathbf{X}_i, \mathbf{X}_j) \quad \text{for } \mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}\}. \quad (\text{A8})$$

The mutual information of one source and the target conditioned on knowing the other source is expressed as:

$$\text{MI}(\mathbf{T}; \mathbf{X}_i | \mathbf{X}_j) = \delta(\mathbf{T}, \mathbf{X}_i, \mathbf{X}_j) + \delta(\mathbf{T}, \mathbf{X}_i) \quad \text{for all } \mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}\}. \quad (\text{A9})$$

The co-information $\text{CoI}(\mathbf{T}; \mathbf{X}; \mathbf{Y})$ is expressed as:

$$\text{CoI}(\mathbf{T}; \mathbf{X}; \mathbf{Y}) = \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}) - \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}). \quad (\text{A10})$$

Appendix B.2. Verification of BROJA Optimization

This subsection will verify that the measures defined in (A6a–c) quantify the desired partial information quantities under the maximum decomposition principle. Assumption A1 implies that

$\min_{Q \in \Delta_P} \delta(\mathbf{T}, \mathbf{XY}) = 0$, $\min_{Q \in \Delta_P} \delta(\mathbf{T}, \mathbf{X.Y}) = \delta(\mathbf{T}, 1.2)$, $\min_{Q \in \Delta_P} \delta(\mathbf{T}, \mathbf{X}) = \delta(\mathbf{T}, \mathbf{X})$, and $\min_{Q \in \Delta_P} \delta(\mathbf{T}, \mathbf{Y}) = \delta(\mathbf{T}, \mathbf{Y})$.
Therefore, it is easy to see that:

$$\begin{aligned} \widetilde{\text{CI}}(\mathbf{T}; \mathbf{X} : \mathbf{Y}) &= \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{Y}, \mathbf{X}) = \delta(\mathbf{T}, \mathbf{XY}) \\ \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X} \setminus \mathbf{Y}) &= \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{Y}) = \delta(\mathbf{T}, \mathbf{X}) \\ \widetilde{\text{UI}}(\mathbf{T}; \mathbf{Y} \setminus \mathbf{X}) &= \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}) = \delta(\mathbf{T}, \mathbf{Y}). \end{aligned}$$

Now, $\text{Col}(\mathbf{T}; \mathbf{X}; \mathbf{Y}) = 0$ implies that $\delta(\mathbf{T}, \mathbf{XY}) = \delta(\mathbf{T}, \mathbf{X.Y})$; thus:

$$\widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) = \min_{\substack{Q \in \Delta_P \\ \text{Col}(\mathbf{T}; \mathbf{X}; \mathbf{Y})=0}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) = \delta(\mathbf{T}, \mathbf{X.Y}).$$

Hence, under Assumption A1,

$$\begin{aligned} \widetilde{\text{CI}}(\mathbf{T}; \mathbf{X} : \mathbf{Y}) &= \text{CI}(\mathbf{T}; \mathbf{X} : \mathbf{Y}), \quad \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X} \setminus \mathbf{Y}) = \text{UI}(\mathbf{T}; \mathbf{X} \setminus \mathbf{Y}) \\ \widetilde{\text{UI}}(\mathbf{T}; \mathbf{Y} \setminus \mathbf{X}) &= \text{UI}(\mathbf{T}; \mathbf{Y} \setminus \mathbf{X}), \quad \widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \text{SI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}). \end{aligned} \tag{A11}$$

Appendix C. Maximum Entropy Decomposition of Trivariate PID

Let \mathbf{T} be the target random variable, $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ be the source random variables, and P be the joint probability distribution of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. Chicharro [38] using maximum entropy decomposed mutual information $\text{MI}(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ into: synergistic, unique, unique redundant, and redundant information. In this decomposition,

- the synergistic quantity, $\widetilde{\text{CI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z})$, captures the sum of all individual synergistic terms, namely $\delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XY}) + \delta(\mathbf{T}; \mathbf{XZ}) + \delta(\mathbf{T}; \mathbf{YZ}) + \delta(\mathbf{T}; \mathbf{XY.XZ}) + \delta(\mathbf{T}; \mathbf{XY.YZ}) + \delta(\mathbf{T}; \mathbf{XZ.YZ}) + \delta(\mathbf{T}; \mathbf{XY.XZ.YZ})$,
- the unique information, $\widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i \setminus \mathbf{X}_j, \mathbf{X}_k)$, captures the sum of the information that \mathbf{X}_i has about \mathbf{T} solely, $\delta(\mathbf{T}; \mathbf{X}_i)$, and the information \mathbf{X}_i knows redundantly with the synergy of $(\mathbf{X}_j, \mathbf{X}_k)$, $\delta(\mathbf{T}; \mathbf{X}_i.\mathbf{X}_j.\mathbf{X}_k)$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$,
- the unique redundant information, $\widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j \setminus \mathbf{X}_k)$, captures the actual unique information that \mathbf{X}_i and \mathbf{X}_j have redundantly about \mathbf{T} , $\delta(\mathbf{T}; \mathbf{X}_i.\mathbf{X}_j)$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$,
- and the redundant information, $\widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ captures the actual redundant information of \mathbf{X}, \mathbf{Y} , and \mathbf{Z} about \mathbf{T} , i.e., $\delta(\mathbf{T}; \mathbf{X.Y.Z})$.

Using Beer–Williams axioms. the decomposition yields these identities:

$$\begin{aligned} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \widetilde{\text{CI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X} \setminus \mathbf{Y}, \mathbf{Z}) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{Y} \setminus \mathbf{X}, \mathbf{Z}) \\ &\quad + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{Z} \setminus \mathbf{X}, \mathbf{Y}) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y} \setminus \mathbf{Z}) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}, \mathbf{Z} \setminus \mathbf{Y}) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{Y}, \mathbf{Z} \setminus \mathbf{X}) \\ \text{MI}(\mathbf{T}; \mathbf{X}_i) &= \widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}_i; \mathbf{X}_j; \mathbf{X}_k) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i \setminus \mathbf{X}_j, \mathbf{X}_k) + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j \setminus \mathbf{X}_k) \\ &\quad + \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_k \setminus \mathbf{X}_j) \quad \text{for all } \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}. \end{aligned} \tag{A12}$$

and Δ is the set of all joint distributions of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$. The measure uses the maximum entropy decomposition over Δ_P in order to compute the above quantities. Moreover, the work in [38] made some assumptions over the partial information measures of the redundancy lattice.

Assumption A2 (Assumptions a.1 and a.2 in [38]). *On the trivariate redundancy lattice (Figure A1), the following assumptions are made to quantify the PID*

1. All partial information measures of the redundancy lattice are nonnegative.
2. The terms $\delta(\mathbf{T}; \mathbf{X.Y.Z})$ and $\delta(\mathbf{T}; \mathbf{X}_i.\mathbf{X}_j)$ for all $\mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ are invariant on Δ_P .

3. The summands $\delta(\mathbf{T}; \mathbf{X}_i) + \delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k)$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ are invariant on Δ_P .
4. The terms $\delta(\mathbf{T}; \mathbf{X}\mathbf{Y}\mathbf{Z}), \delta(\mathbf{T}; \mathbf{X}\mathbf{Y}.\mathbf{X}\mathbf{Z}.\mathbf{Y}\mathbf{Z}), \delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j), \delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j \mathbf{X}_i \mathbf{X}_k), \delta(\mathbf{T}; \mathbf{X}_i),$ and $\delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k)$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ are not constant on Δ_P .
5. All synergistic terms, $\delta(\mathbf{T}; \mathbf{X}\mathbf{Y}\mathbf{Z}), \delta(\mathbf{T}; \mathbf{X}\mathbf{Y}.\mathbf{X}\mathbf{Z}.\mathbf{Y}\mathbf{Z}), \delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j),$ and $\delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j \mathbf{X}_i \mathbf{X}_k)$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ vanish at the minimum over Δ_P .
6. The partial information measures $\delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k)$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ vanish at the minimum over Δ_P .

Under the above assumptions and using maximal entropy decomposition, the work in [38] defined the following optimization problems that compute the PID quantities.

$$\widetilde{\text{CI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{Y}, \mathbf{X}, \mathbf{Z}) \tag{A13a}$$

$$\begin{aligned} \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i \setminus \mathbf{X}_j, \mathbf{X}_k) &= \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}_j, \mathbf{X}_k) \\ &\text{for all } \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} \end{aligned} \tag{A13b}$$

$$\begin{aligned} \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j \setminus \mathbf{X}_k) &= \min_{\substack{Q \in \Delta_P, \\ \text{Col}(\mathbf{T}; \mathbf{X}_i; \mathbf{X}_j | \mathbf{X}_k) = 0}} \text{MI}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) - \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) \\ &\text{for all } \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\} \end{aligned} \tag{A13c}$$

$$\widetilde{\text{SI}}(\mathbf{T}; \mathbf{Z}, \mathbf{Y}, \mathbf{Z}) = \min_{\substack{Q \in \Delta_P, \text{Col}(\mathbf{T}; \mathbf{X}; \mathbf{Y}) = 0, \\ \text{Col}(\mathbf{T}; \mathbf{X}; \mathbf{Y} | \mathbf{Z}) = 0, w(Q)}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{\substack{Q \in \Delta_P, w(Q), \\ \text{Col}(\mathbf{T}; \mathbf{X}; \mathbf{Y} | \mathbf{Z}) = 0}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}), \tag{A13d}$$

where:

$$\begin{aligned} w(Q) := \{Q \in \Delta : \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) &= \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}), \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Z}) = \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Z}), \\ \text{MI}(\mathbf{T}; \mathbf{Y}, \mathbf{Z}) &= \min_{Q \in \Delta_P} \text{MI}(\mathbf{T}; \mathbf{Y}, \mathbf{Z})\}. \end{aligned}$$

Mutual Information over the Trivariate Redundancy Lattice

This subsection writes down some mutual information quantities in terms of the trivariate redundancy lattice’s partial information measures using (A3). The verification that the optimization defined in (A13a–d) quantifies the desired partial information quantities was discussed in detail by [38] and so will be skipped. However, these formulas are needed later when discussing how to compute the individual PID terms using a hierarchy of BROJA and [38] PID decompositions. The mutual information quantities are in terms of redundancy lattice partial information measures.

$\text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ will be the sum of the partial information measure on every node of the redundancy lattice \mathcal{C} as follows.

$$\begin{aligned} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}.\mathbf{X}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}.\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z}.\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}.\mathbf{X}\mathbf{Z}.\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}) \\ &\quad + \delta(\mathbf{T}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}.\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}.\mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Z}.\mathbf{X}\mathbf{Y}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}.\mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}.\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}.\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}.\mathbf{Y}.\mathbf{Z}). \end{aligned} \tag{A14}$$

For all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$, the mutual information of two sources (jointly) and the target is expressed as:

$$\begin{aligned} \text{MI}(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j) &= \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j) + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \mathbf{X}_i \mathbf{X}_k) + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \mathbf{X}_j \mathbf{X}_k) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \mathbf{X}_i \mathbf{X}_k \mathbf{X}_j \mathbf{X}_k) + \delta(\mathbf{T}, \mathbf{X}_i) + \delta(\mathbf{T}, \mathbf{X}_j) + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}_j \mathbf{X}_i \mathbf{X}_k) + \delta(\mathbf{T}, \mathbf{X}_k \mathbf{X}_i \mathbf{X}_j) + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j) + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_k) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}_j \mathbf{X}_k) + \delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k). \end{aligned} \tag{A15}$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the mutual information of one source and the target is as follows:

$$MI(T; X_i) = \delta(T, X_i) + \delta(T, X_i X_j X_k) + \delta(T, X_i X_j) + \delta(T, X_i X_k) + \delta(T, X_i X_j X_k). \quad (A16)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the mutual information of two sources (jointly) and the target conditioned on knowing the other source is evaluated as:

$$\begin{aligned} MI(T; X_i, X_j | X_k) &= \delta(T, X_i X_j X_k) + \delta(T, X_i X_j) + \delta(T, X_i X_k) + \delta(T, X_j X_k) \\ &+ \delta(T, X_i X_j X_k) + \delta(T, X_i X_j X_k) + \delta(T, X_i X_k X_j X_k) \\ &+ \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_i) \\ &+ \delta(T, X_j) + \delta(T, X_i X_j X_k) + \delta(T, X_j X_i X_k) + \delta(T, X_i X_j). \end{aligned} \quad (A17)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the mutual information of one source and the target conditioned on knowing only one of the other sources is written as:

$$\begin{aligned} MI(T; X_i | X_j) &= \delta(T, X_i X_j) + \delta(T, X_i X_j X_k) + \delta(T, X_i X_j X_k) \\ &+ \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_i) + \delta(T, X_i X_j X_k) \\ &+ \delta(T, X_k X_i X_j) + \delta(T, X_i X_k). \end{aligned} \quad (A18)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the mutual information of one source and the target conditioned on knowing the other sources is:

$$\begin{aligned} MI(T; X_i | X_j, X_k) &= \delta(T, X_i X_j X_k) + \delta(T, X_i X_j) + \delta(T, X_i X_k) \\ &+ \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_i). \end{aligned} \quad (A19)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the co-information of two sources and the target is expressed as:

$$\begin{aligned} Col(T; X_i; X_j) &= \delta(T, X_i X_j) + \delta(T, X_i X_j X_k) - (\delta(T, X_i X_j) + \delta(T, X_i X_j X_k) \\ &+ \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_k X_i X_j)). \end{aligned} \quad (A20)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the co-information of one source, two sources (jointly), and the target is as follows:

$$\begin{aligned} Col(T; X_i; X_j, X_k) &= \delta(T, X_i X_j X_k) + \delta(T, X_i X_j) + \delta(T, X_i X_k) + \delta(T, X_i X_j X_k) \\ &- \left(\delta(T, X_i X_j X_k) + \delta(T, X_i X_j) + \delta(T, X_i X_k) + \delta(T, X_i X_j X_k X_j X_k) \right). \end{aligned} \quad (A21)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the co-information of two sources (jointly), two sources (jointly), and the target is evaluated as:

$$\begin{aligned} Col(T; X_i, X_j; X_i, X_k) &= \delta(T, X_i X_j X_k) + \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_i) \\ &+ \delta(T, X_k X_i X_j) + \delta(T, X_i X_j) + \delta(T, X_i X_j X_k) + \delta(T, X_j X_i X_k) \\ &+ \delta(T, X_i X_k) + \delta(T, X_j X_k) + \delta(T, X_i X_j X_k) - \delta(T, X_i X_j X_k) \\ &- \delta(T, X_j X_k). \end{aligned} \quad (A22)$$

For all $X_i, X_j, X_k \in \{X, Y, Z\}$, the co-information of two sources and the target conditioning on knowing the other source can be written as:

$$\begin{aligned} Col(T; X_i; X_j | X_k) &= \delta(T, X_i X_k X_j X_k) + \delta(T, X_i X_j X_k X_j X_k) + \delta(T, X_i X_j X_k) \\ &+ \delta(T, X_j X_i X_k) + \delta(T, X_i X_j) - \delta(T, X_i X_j X_k) - \delta(T, X_i X_k). \end{aligned} \quad (A23)$$

Appendix D. The Finer Quantities of Trivariate Maximum Entropy PID

In Appendix C, the maximum entropy decomposition for trivariate PID returns a synergistic term, which is the sum of all individual synergy quantities, and a unique term, which is the sum of unique and unique redundancy quantities. This section aims to show how to use maximum entropy decomposition for bivariate PID in order to obtain each individual synergy quantity, as well as each individual unique and unique redundancy quantity.

Let T be the target random variable, X, Y, Z be the source random variables, and P be the joint probability distribution of (T, X, Y, Z) . Now, BROJA will be applied to some subsystems of (T, X, Y, Z) , namely, $(T, (X_i, X_j), X_k)$, (one single source) and $(T, (X_i, X_j), (X_i, X_k))$ (two double sources) for all $X_i, X_j, X_k \in \{X, Y, Z\}$. Note that the pairs (X_i, X_j) and (X_i, X_k) are ordered alphabetically. Consider the following probability polytopes upon which the optimization will be carried out:

$$\begin{aligned} \Delta_P &= \{Q \in \Delta; Q(T, X) = P(T, X), Q(T, Y) = P(T, Y), Q(T, Z) = P(T, Z)\} \\ \Delta_P^{X_i, X_j, X_k} &= \{Q \in \Delta; Q(T, X_i, X_j) = P(T, X_i, X_j), Q(T, X_k) = P(T, X_k)\} \\ &\quad \text{where } X_i \neq X_j, X_i \neq X_k, X_j \neq X_k \text{ for all } X_i, X_j, X_k \in \{X, Y, Z\} \\ \Delta_P^{X_i, X_j, X_i, X_k} &= \{Q \in \Delta; Q(T, X_i, X_j) = P(T, X_i, X_j), Q(T, X_i, X_k) = P(T, X_i, X_k)\} \\ &\quad \text{where } X_i \neq X_j, X_i \neq X_k, X_j \neq X_k \text{ for all } X_i, X_j, X_k \in \{X, Y, Z\}. \end{aligned} \tag{A24}$$

Note that $\Delta_P \subsetneq \Delta_P^{X_i, X_j, X_k} \subsetneq \Delta_P^{X_i, X_j, X_i, X_k}$ for all $X_i, X_j, X_k \in \{X, Y, Z\}$.

Appendix D.1. One Single Source Subsystems

These subsystems have the form $(T, (X_i, X_j), X_k)$ where $X_i, X_j, X_k \in \{X, Y, Z\}$, $X_i \neq X_j \neq X_k$, and $X_i \neq X_k$. Now, apply the BROJA decomposition to the subsystem $(T, (X, Y), Z)$. Therefore, its four PID quantities are defined as follows:

$$\begin{aligned} \widetilde{CI}(T; X, Y) &= MI(T; X, Y, Z) - \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; X, Y, Z) \\ \widetilde{UI}(T; X, Y \setminus Z) &= \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; X, Y, Z) - \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; Z) \\ \widetilde{UI}(T; Z \setminus X, Y) &= \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; X, Y, Z) - \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; X, Y) \\ \widetilde{SI}(T; X, Y, Z) &= \min_{\substack{Q \in \Delta_P^{X, Y, Z}, \\ Col(T; X, Y, Z) = 0}} MI(T; X, Y, Z) - \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; X, Y, Z). \end{aligned}$$

Note that the (X, Y) marginal distribution is fixed. This implies that the mutual information $MI(T; X, Y)$, $MI(T; X | Y)$, $MI(T; Y | X)$, and $Col(T; X, Y)$ is invariant over $\Delta_P^{X, Y, Z}$. Therefore, the summands $\delta(T, XY) + \delta(T, XY.XZ) + \delta(T, XY.YZ) + \delta(T, XY.XZ.YZ) + \delta(T, Z.XY)$ are fixed. However, from Assumption A2 and the fact that the X, Y marginal is fixed, the redundancy $\delta(T; Z.XY)$ is invariant over $\Delta_P^{X, Y, Z}$. Thus, in addition to 2 in Assumption A2, the following partial information measures are invariant over $\Delta_P^{X, Y, Z}$.

1. $\delta(T; Z.XY)$ since the (X, Y) marginal is fixed.
2. $\delta(T; Z)$ since $MI(T; Z)$ and $\delta(T; Z.XY)$ are invariant over $\Delta_P^{X, Y, Z}$.
3. $\delta(T, XY) + \delta(T, XY.XZ) + \delta(T, XY.YZ) + \delta(T, XY.XZ.YZ)$ since $Col(T; X, Y)$ and $\delta(T; Z.XY)$ are invariant over $\Delta_P^{X, Y, Z}$.

Thus, using Assumption A2 and the definition of $MI(T; X, Y, Z)$ over the redundancy lattice,

$$\begin{aligned} \min_{Q \in \Delta_P^{X, Y, Z}} MI(T; X, Y, Z) &= \delta(T, XY) + \delta(T, XY.XZ) + \delta(T, XY.YZ) + \delta(T, XY.XZ.YZ) \\ &\quad + \delta(T, X) + \delta(T, Y) + \delta(T, Z) + \delta(T, X.YZ) + \delta(T, Y.XZ) \\ &\quad + \delta(T, Z.XY) + \delta(T, X.Y) + \delta(T, X.Z) + \delta(T, Y.Z) + \delta(T, X.Y.Z). \end{aligned}$$

The synergy is evaluated as:

$$\begin{aligned} \widetilde{CI}(T; X, Y) &= MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,Z}} MI(T; X, Y, X) \\ &= \delta(T, XYZ) + \delta(T, XZ) + \delta(T, YZ) + \delta(T, XZ.YZ). \end{aligned}$$

The unique information of (X, Y) in terms of the redundancy lattice atoms is:

$$\begin{aligned} \widetilde{UI}(T; XY \setminus Z) &= \min_{Q \in \Delta_p^{XY,Z}} MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,Z}} MI(T; Z) \\ &= \delta(T, XY) + \delta(T, XY.XZ) + \delta(T, XY.YZ) + \delta(T, XY.XZ.YZ) \\ &\quad + \delta(T, X) + \delta(T, Y) + \delta(T, X.YZ) + \delta(T, Y.XZ) + \delta(T, X.Y). \end{aligned}$$

The unique information of Z is written as:

$$\widetilde{UI}(T; Z \setminus XY) = \min_{Q \in \Delta_p^{XY,Z}} MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,Z}} MI(T; X, Y) = \delta(T, Z)$$

When $Col(T; X, Y, Z) = 0$, then $\delta(T, Z.XY) + \delta(T, X.Z) + \delta(T, Y.Z) + \delta(T, X.Y.Z)$ is equal to $\delta(T, XYZ) + \delta(T, XZ) + \delta(T, YZ) + \delta(T, XZ.YZ)$.

Then, in terms of redundancy lattice atoms, the shared information of (X, Y) and Z is:

$$\begin{aligned} \widetilde{SI}(T; X, Y, Z) &= \min_{\substack{Q \in \Delta_p^{XY,Z}, \\ Col(T; X, Y, Z)=0}} MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,Z}} MI(T; X, Y, Z) \\ &= \delta(T, Z.XY) + \delta(T, X.Z) + \delta(T, Y.Z) + \delta(T, X.Y.Z). \end{aligned}$$

Hence, for all $X_i, X_j, X_k \in \{X, Y, Z\}$, the BROJA decomposition of $(T, (X_i, X_j), X_k)$ is:

$$\begin{aligned} \widetilde{CI}(T; (X_i, X_j), X_k) &= \delta(T; X_i X_j X_k) + \delta(T; X_i X_k) + \delta(T; X_j X_k) + \delta(T; X_i X_k X_j X_k) \\ \widetilde{UI}(T; (X_i, X_j) \setminus X_k) &= \delta(T; X_i) + \delta(T; X_i X_j X_k) + \delta(T; X_j) + \delta(T; X_j X_i X_k) + \delta(T; X_i X_j) \\ &\quad + \delta(T; X_i X_j) + \delta(T; X_i X_j X_i X_k) + \delta(T; X_i X_j X_j X_k) \\ &\quad + \delta(T; X_i X_j X_i X_k X_j X_k) \\ \widetilde{UI}(T; X_k \setminus X_i, X_j) &= \delta(T; X_k) \\ \widetilde{SI}(T; X_i, X_j, X_k) &= \delta(T; X_k X_i X_j) + \delta(T; X_i X_k) + \delta(T; X_j X_k) + \delta(T; X_i X_j X_k). \end{aligned}$$

Appendix D.2. Two Double Source Subsystems

These subsystems have the form $(T, (X_i, X_j), (X_i, X_k))$ where $X_i, X_j, X_k \in \{X, Y, Z\}$, $X_i \neq X_j$, $X_i \neq X_k$, and $X_j \neq X_k$. Now, apply the BROJA decomposition to the subsystem $(T, (X, Y), (X, Z))$. Therefore, its four PID quantities are defined as follows:

$$\begin{aligned} \widetilde{CI}(T; X, Y) &= MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,XZ}} MI(T; X, Y, Z) \\ \widetilde{UI}(T; (X, Y) \setminus (X, Z)) &= \min_{Q \in \Delta_p^{XY,XZ}} MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,XZ}} MI(T; X, Z) \\ \widetilde{UI}(T; (X, Z) \setminus (X, Y)) &= \min_{Q \in \Delta_p^{XY,XZ}} MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,XZ}} MI(T; X, Y) \\ \widetilde{SI}(T; (X, Y), (X, Z)) &= \min_{\substack{Q \in \Delta_p^{XY,XZ}, \\ Col(T; X, Y, X, Z)=0}} MI(T; X, Y, Z) - \min_{Q \in \Delta_p^{XY,XZ}} MI(T; X, Y, Z). \end{aligned}$$

Note that the (X, Y) and (X, Z) marginal distributions are fixed. Then, $MI(T; X_1, X_2)$, $MI(T; X_1 | X_2)$, $Col(T; X_1; X_2)$, and $MI(T; X : Y, X : Z)$ are invariant over $\Delta_p^{XY,XZ}$ for $(X_1, X_2) = (X, Y)$ and $(X_1, X_2) = (X, Z)$. Therefore, for $(X_1, X_2, X_3) = (X, Y, Z)$ and $(X_1, X_2, X_3) = (X, Z, Y)$, $\delta(T, X_1 X_2) + \delta(T, X_1 X_2 X_1 X_3) + \delta(T, X_1 X_2 X_2 X_3) + \delta(T, X_1 X_2 X_1 X_3 X_2 X_3) + \delta(T, X_3 X_1 X_2)$ are fixed. However, from Assumption A2 and the two fixed (X, Y) and (X, Z) marginals, then the redundancies $\delta(T; Z.XY)$ and $\delta(T; Y.XZ)$ are invariant over $\Delta_p^{XY,XZ}$. Therefore, in addition to 2 in Assumption A2, the following partial information measures are invariant $\Delta_p^{XY,XZ}$:

1. $\delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}\mathbf{X}_j)$, for all $\mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{Y}, \mathbf{Z}\}$ since the $(\mathbf{X}, \mathbf{X}_j)$ marginal is fixed.
2. $\delta(\mathbf{T}; \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}, \mathbf{Y}\mathbf{Z})$ since $\text{MI}(\mathbf{T}; \mathbf{X} : \mathbf{Y}, \mathbf{X} : \mathbf{Z})$ and $\delta(\mathbf{T}; \mathbf{Z}, \mathbf{X}\mathbf{Y})$ are invariant.
3. $\delta(\mathbf{T}; \mathbf{X}_i)$, for all $\mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{Y}, \mathbf{Z}\}$ since $\text{MI}(\mathbf{T}; \mathbf{X}_i)$ and $\delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}\mathbf{X}_j)$ are invariant over $\Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}$.
4. $\delta(\mathbf{T}, \mathbf{X}\mathbf{X}_i) + \delta(\mathbf{T}, \mathbf{X}\mathbf{X}_i, \mathbf{X}_i, \mathbf{X}_j)$, is invariant for all $\mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{Y}, \mathbf{Z}\}$ since $\delta(\mathbf{T}; \mathbf{X}\mathbf{X}_i, \mathbf{X}\mathbf{X}_j) + \delta(\mathbf{T}, \mathbf{X}\mathbf{X}_i, \mathbf{X}\mathbf{X}_j, \mathbf{X}_i, \mathbf{X}_j)$, and $\text{CoI}(\mathbf{T}; \mathbf{X}; \mathbf{X}_i), \delta(\mathbf{T}; \mathbf{X}_j, \mathbf{X}\mathbf{X}_i)$ are invariant over $\Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}$.

Thus, using Assumption A2 and the definition of $\text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ over the redundancy lattice,

$$\begin{aligned} \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{Y}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}) + \delta(\mathbf{T}, \mathbf{Y}) \\ &\quad + \delta(\mathbf{T}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Z}, \mathbf{X}\mathbf{Y}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}). \end{aligned}$$

The synergy is evaluated as:

$$\begin{aligned} \widetilde{\text{CI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) &= \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \\ &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}\mathbf{Z}). \end{aligned}$$

The unique information of (\mathbf{X}, \mathbf{Y}) is expressed as:

$$\begin{aligned} \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y} \setminus \mathbf{X}, \mathbf{Z}) &= \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Z}) \\ &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}). \end{aligned}$$

The unique information of (\mathbf{X}, \mathbf{Z}) is written as:

$$\begin{aligned} \widetilde{\text{UI}}(\mathbf{T}; \mathbf{X}, \mathbf{Z} \setminus \mathbf{X}, \mathbf{Y}) &= \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}) \\ &= \delta(\mathbf{T}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Z}). \end{aligned}$$

When $\text{CoI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}; \mathbf{X}, \mathbf{Z}) = 0$, then:

$$\begin{aligned} \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}\mathbf{Z}) &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}, \mathbf{Y}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Z}, \mathbf{X}\mathbf{Y}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}). \end{aligned}$$

The shared information of \mathbf{X}, \mathbf{Y} and \mathbf{X}, \mathbf{Z} is evaluated as:

$$\begin{aligned} \widetilde{\text{SI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \min_{\substack{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}, \\ \text{CoI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}; \mathbf{X}, \mathbf{Z})=0}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{Q \in \Delta_p^{\mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) \\ &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}, \mathbf{X}\mathbf{Z}, \mathbf{Y}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Z}, \mathbf{X}\mathbf{Y}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}). \end{aligned}$$

Hence, then the BROJA decomposition of the subsystem $(\mathbf{T}, (\mathbf{X}_i, \mathbf{X}_j), (\mathbf{X}_i, \mathbf{X}_k))$ is:

$$\begin{aligned} \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}_i, \mathbf{X}_j), (\mathbf{X}_i, \mathbf{X}_k)) &= \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_j, \mathbf{X}_k), \\ \widetilde{\text{UI}}(\mathbf{T}; (\mathbf{X}_i, \mathbf{X}_j) \setminus (\mathbf{X}_i, \mathbf{X}_k)) &= \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j) + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_j, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_j), \\ \widetilde{\text{UI}}(\mathbf{T}; (\mathbf{X}_i, \mathbf{X}_j) \setminus (\mathbf{X}_i, \mathbf{X}_k)) &= \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_k, \mathbf{X}_j, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_k), \\ \widetilde{\text{SI}}(\mathbf{T}; (\mathbf{X}_i, \mathbf{X}_j), (\mathbf{X}_i, \mathbf{X}_k)) &= \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_j, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_i) + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_i, \mathbf{X}_k, \mathbf{Y}_j, \mathbf{X}_k) \\ &\quad + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_j, \mathbf{X}_i, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_k, \mathbf{X}_i, \mathbf{X}_j) \\ &\quad + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j) + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_j, \mathbf{X}_k) + \delta(\mathbf{T}; \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k). \end{aligned}$$

Appendix D.3. Synergy of Three Double Source Systems

Consider the system of the form $(\mathbf{T}, (\mathbf{X}, \mathbf{Y}), (\mathbf{X}, \mathbf{Z}), (\mathbf{Y}, \mathbf{Z}))$. The sources here are called composite, as they are compositions of the primary sources \mathbf{X}, \mathbf{Y} and \mathbf{Z} . Applying the PID measure [38] based on maximum entropy decomposition (A13a–d) captures the synergy of composite sources only and cannot capture other contributions such as those involving unique or redundant composite sources; meaning that the optimization (A13a) is the only useful one for a system of composite sources. Therefore, using (A13a), the optimization is taken over the polytope:

$$\Delta_P^{XY.XZ.YZ} = \{Q \in \Delta; Q(\mathbf{T}, \mathbf{X}_i, \mathbf{X}_j) = P(\mathbf{T}, \mathbf{X}_i, \mathbf{X}_j) \text{ for all } \mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}\}. \quad (\text{A25})$$

In this polytope, $\text{MI}(\mathbf{X}_i, \mathbf{X}_j)$, $\text{CoI}(\mathbf{X}_i, \mathbf{X}_j)$, and $\text{MI}(\mathbf{X}_i | \mathbf{X}_j)$ are invariant for all $\mathbf{X}_i, \mathbf{X}_j \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$. Therefore, in addition to Assumption 2, the following partial information measures are invariant $\Delta_P^{XY.XZ.YZ}$.

1. $\delta(\mathbf{T}; \mathbf{X}_k \cdot \mathbf{X}_i \mathbf{X}_j)$, for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ since the $(\mathbf{X}_i, \mathbf{X}_j)$ marginal is fixed.
2. $\delta(\mathbf{T}; \mathbf{X}_i \mathbf{X}_j \cdot \mathbf{X}_i \mathbf{X}_k)$, for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ since $(\mathbf{X}_i, \mathbf{X}_j)$, $(\mathbf{X}_i, \mathbf{X}_k)$, and $(\mathbf{X}_j, \mathbf{X}_k)$ marginals are fixed.
3. $\delta(\mathbf{T}; \mathbf{X}\mathbf{Y} \cdot \mathbf{X}\mathbf{Z} \cdot \mathbf{Y}\mathbf{Z})$ since the (\mathbf{X}, \mathbf{Y}) , (\mathbf{X}, \mathbf{Z}) , and (\mathbf{Y}, \mathbf{Z}) marginals are fixed.
4. $\delta(\mathbf{T}; \mathbf{X}_i)$, for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ since $\text{MI}(\mathbf{T}; \mathbf{X}_i)$ and $\delta(\mathbf{T}; \mathbf{X}_i \cdot \mathbf{X}_j \mathbf{X}_k)$ are invariant over $\Delta_P^{XY.XZ.YZ}$.
5. $\delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j)$, for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$ since $\text{CoI}(\mathbf{T}; \mathbf{X}_i; \mathbf{X}_j)$, $\delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \cdot \mathbf{X}_i \mathbf{X}_k)$, $\delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \cdot \mathbf{X}_j \mathbf{X}_k)$, $\delta(\mathbf{T}, \mathbf{X}_i \mathbf{X}_j \cdot \mathbf{X}_i \mathbf{X}_k \cdot \mathbf{X}_j \mathbf{X}_k)$, and $\Delta(\mathbf{T}; \mathbf{X}_k \cdot \mathbf{X}_i \mathbf{X}_j)$ are invariant over $\Delta_P^{XY.XZ.YZ}$.

Hence, the only partial information measure that is not fixed is $\delta(\mathbf{T}; \mathbf{X}\mathbf{Y}\mathbf{Z})$ and:

$$\begin{aligned} \min_{Q \in \Delta_P^{XY.XZ.YZ}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y} \cdot \mathbf{X}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y} \cdot \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Z} \cdot \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X}\mathbf{Y} \cdot \mathbf{X}\mathbf{Z} \cdot \mathbf{Y}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X}) + \delta(\mathbf{T}, \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{X} \cdot \mathbf{Y}\mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y} \cdot \mathbf{X}\mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{Z} \cdot \mathbf{X}\mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X} \cdot \mathbf{Y}) + \delta(\mathbf{T}, \mathbf{X} \cdot \mathbf{Z}) + \delta(\mathbf{T}, \mathbf{Y} \cdot \mathbf{Z}) \\ &\quad + \delta(\mathbf{T}, \mathbf{X} \cdot \mathbf{Y} \cdot \mathbf{Z}). \end{aligned} \quad (\text{A26})$$

The synergy is evaluated as:

$$\widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Y}), (\mathbf{X}, \mathbf{Z}), (\mathbf{Y}, \mathbf{Z})) = \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) - \min_{Q \in \Delta_P^{XY.XZ.YZ}} \text{MI}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{X}) = \delta(\mathbf{T}, \mathbf{X}\mathbf{Y}\mathbf{Z}).$$

Appendix D.4. Computing the Finest Parts of the Trivariate PID

The values of $\delta(\mathbf{T}; \mathbf{X})$, $\delta(\mathbf{T}; \mathbf{Y})$, $\delta(\mathbf{T}; \mathbf{Z})$, $\delta(\mathbf{T}; \mathbf{X} \cdot \mathbf{Y}\mathbf{Z})$, $\delta(\mathbf{T}; \mathbf{Y} \cdot \mathbf{X}\mathbf{Z})$, and $\delta(\mathbf{T}; \mathbf{Z} \cdot \mathbf{X}\mathbf{Y})$ are recovered from $\widetilde{\text{UI}}(\mathbf{X}_k \setminus \mathbf{X}_i, \mathbf{X}_j)$ of $(\mathbf{T}, (\mathbf{X}_i, \mathbf{X}_j), \mathbf{X}_k)$ and $\widetilde{\text{UI}}(\mathbf{X}_k \setminus \mathbf{X}_i, \mathbf{X}_j)$ of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$, for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$.

To recover the individual synergistic quantities, construct the following system of equations from the synergy of $(\mathbf{T}, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$, $(\mathbf{T}, (\mathbf{X}, \mathbf{Y}), (\mathbf{X}, \mathbf{Z}), (\mathbf{Y}, \mathbf{Z}))$, $(\mathbf{T}, (\mathbf{X}_i, \mathbf{X}_j), \mathbf{X}_k)$, and $(\mathbf{T}, (\mathbf{X}_i, \mathbf{X}_j), (\mathbf{X}_i, \mathbf{X}_k))$ for all $\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k \in \{\mathbf{X}, \mathbf{Y}, \mathbf{Z}\}$.

$$\begin{aligned}\widetilde{\text{CI}}(\mathbf{T}; \mathbf{X}, \mathbf{Y}, \mathbf{Z}) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XY.XZ.YZ}) + \delta(\mathbf{T}; \mathbf{XY}) + \delta(\mathbf{T}; \mathbf{XZ}) \\ &\quad + \delta(\mathbf{T}; \mathbf{YZ}) + \delta(\mathbf{T}; \mathbf{XY.XZ}) + \delta(\mathbf{T}; \mathbf{XY.YZ}) + \delta(\mathbf{T}; \mathbf{XZ.YZ}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Y}), \mathbf{Z}) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XZ}) + \delta(\mathbf{T}; \mathbf{YZ}) + \delta(\mathbf{T}; \mathbf{XZ.YZ}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Z}), \mathbf{Y}) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XY}) + \delta(\mathbf{T}; \mathbf{YZ}) + \delta(\mathbf{T}; \mathbf{XY.YZ}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{Y}, \mathbf{Z}), \mathbf{X}) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XY}) + \delta(\mathbf{T}; \mathbf{XZ}) + \delta(\mathbf{T}; \mathbf{XY.XZ}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Y}), (\mathbf{X}, \mathbf{Z})) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{YZ}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Y}), (\mathbf{Y}, \mathbf{Z})) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XZ}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Z}), (\mathbf{Y}, \mathbf{Z})) &= \delta(\mathbf{T}; \mathbf{XYZ}) + \delta(\mathbf{T}; \mathbf{XY}) \\ \widetilde{\text{CI}}(\mathbf{T}; (\mathbf{X}, \mathbf{Y}), (\mathbf{X}, \mathbf{Z}), (\mathbf{Y}, \mathbf{Z})) &= \delta(\mathbf{T}; \mathbf{XYZ}).\end{aligned}$$

This hierarchy that is needed to compute the trivariate PID quantities is implemented in script file `test_trivariate_finer_parts.py`.

References

- Amari, S. Information Geometry on Hierarchy of Probability Distributions. *IEEE Trans. Inf. Theory* **2001**, *47*, 1701–1711. [[CrossRef](#)]
- Schneidman, E.; Still, S.; Berry, M.J.; Bialek, W. Network Information and Connected Correlations. *Phys. Rev. Lett.* **2003**, *91*, 238701. [[CrossRef](#)] [[PubMed](#)]
- Timme, N.; Alford, W.; Flecker, B.; Beggs, J.M. Synergy, Redundancy, and Multivariate Information Measures: An Experimentalist's Perspective. *J. Comput. Neurosci.* **2014**, *36*, 119–140. [[CrossRef](#)] [[PubMed](#)]
- Olbrich, E.; Bertschinger, N.; Rauh, J. Information Decomposition and Synergy. *Entropy* **2015**, *17*, 3501–3517. [[CrossRef](#)]
- Perrone, P.; Ay, N. Hierarchical Quantification of Synergy in Channels. *Front. Robot. AI* **2016**, *2*, 35. [[CrossRef](#)]
- Williams, P.L.; Beer, R.D. Nonnegative Decomposition of Multivariate Information. *arXiv* **2010**, arXiv:1004.2515.
- Williams, P.L. Information Dynamics: Its Theory and Application to Embodied Cognitive Systems. Ph.D. Thesis, Indiana University, Bloomington, IN, USA, 2011.
- Anastassiou, D. Computational Analysis of the Synergy among Multiple Interacting Genes. *Mol. Syst. Biol.* **2007**, *3*, 83. [[CrossRef](#)]
- Watkinson, J.; Liang, K.C.; Wang, X.; Zheng, T.; Anastassiou, D. Inference of Regulatory Gene Interactions from Expression Data Using Three-way Mutual Information. *Ann. N. Y. Acad. Sci.* **2009**, *1158*, 302–313. [[CrossRef](#)]
- Chatterjee, P.; Pal, N.R. Construction of Synergy Networks from Gene Expression Data Related to Disease. *Gene* **2016**, *590*, 250–262. [[CrossRef](#)]
- Katz, Y.; Tunström, K.; Ioannou, C.C.; Huepe, C.; Couzin, I.D. Inferring the Structure and Dynamics of Interactions in Schooling Fish. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18720–18725. [[CrossRef](#)]
- Flack, J.C. Multiple Time-scales and the Developmental Dynamics of Social Systems. *Philos. Trans. R. Soc. B Biol. Sci.* **2012**, *367*, 1802–1810. [[CrossRef](#)] [[PubMed](#)]
- Ay, N.; Der, R.; Prokopenko, M. Information-Driven Self-Organization: The Dynamical System Approach to Autonomous Robot Behavior. *Theory Biosci.* **2012**, *131*, 125–127. [[CrossRef](#)] [[PubMed](#)]
- Frey, S.; Albino, D.K.; Williams, P.L. Synergistic Information Processing Encrypts Strategic Reasoning in Poker. *Cogn. Sci.* **2018**, *42*, 1457–1476. [[CrossRef](#)]
- Marre, O.; El Boustani, S.; Frégnac, Y.; Destexhe, A. Prediction of Spatiotemporal Patterns of Neural Activity from Pairwise Correlations. *Phys. Rev. Lett.* **2009**, *102*, 138101. [[CrossRef](#)] [[PubMed](#)]
- Faes, L.; Marinazzo, D.; Nollo, G.; Porta, A. An Information-Theoretic Framework to Map the Spatiotemporal Dynamics of the Scalp Electroencephalogram. *IEEE. Trans. Biomed. Eng.* **2016**, *63*, 2488–2496. [[CrossRef](#)]

17. Pica, G.; Piasini, E.; Safaai, H.; Runyan, C.A.; Diamond, M.E.; Fellin, T.; Kayser, C.; Harvey, C.D.; Panzeri, S. Quantifying How Much Sensory Information in a Neural Code is Relevant for Behavior. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
18. Latham, P.E.; Nirenberg, S. Synergy, Redundancy, and Independence in Population Codes, Revisited. *J. Neurosci.* **2005**, *25*, 5195–5206. [[CrossRef](#)]
19. Ver Steeg, G.; Brekelmans, R.; Harutyunyan, H.; Galstyan, A. Disentangled Representations via Synergy Minimization. *arXiv* **2017**, arXiv:1710.03839.
20. Rauh, J.; Ay, N. Robustness, Canalizing Functions and Systems Design. *Theory Biosci.* **2014**, *133*, 63–78. [[CrossRef](#)] [[PubMed](#)]
21. Tishby, N.; Pereira, F.C.; Bialek, W. The Information Bottleneck Method. *arXiv* **2000**, arXiv:physics/0004057.
22. Banerjee, P.K.R.; Montúfar, G. The Variational Deficiency Bottleneck. *arXiv* **2018**, arXiv:1810.11677.
23. Harder, M.; Salge, C.; Polani, D. Bivariate Measure of Redundant Information. *Phys. Rev. E* **2013**, *87*, 012130. [[CrossRef](#)] [[PubMed](#)]
24. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying Unique Information. *Entropy* **2014**, *16*, 2161–2183. [[CrossRef](#)]
25. Griffith, V.; Koch, C. Quantifying Synergistic Mutual Information. In *Guided Self-Organization: Inception*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 159–190.
26. Ince, R.A.A. Measuring Multivariate Redundant Information with Pointwise Common Change in Surprisal. *Entropy* **2017**, *19*, 318. [[CrossRef](#)]
27. James, R.G.; Emenheiser, J.; Crutchfield, J.P. Unique Information via Dependency Constraints. *arXiv* **2017**, arXiv:1709.06653v1.
28. Chicharro, D.; Panzeri, S. Synergy and Redundancy in Dual Decompositions of Mutual Information Gain and Information Loss. *Entropy* **2017**, *19*, 71. [[CrossRef](#)]
29. Finn, C.; Lizier, J.T. Pointwise Information Decomposition Using the Specificity and Ambiguity Lattices. *arXiv* **2018**, arXiv:1801.09010v1.
30. Rauh, J. Secret Sharing and Shared Information. *Entropy* **2017**, *19*, 601. [[CrossRef](#)]
31. James, R.G.; Emenheiser, J.; Crutchfield, J.P. A Perspective on Unique Information: Directionality, Intuitions, and Secret Key Agreement. *arXiv* **2017**, arXiv:1808.08606.
32. Chicharro, D.; Pica, G.; Panzeri, S. The Identity of Information: How Deterministic Dependencies Constrain Information Synergy and Redundancy. *Entropy* **2018**, *20*, 169. [[CrossRef](#)]
33. Rauh, J.; Banerjee, P.K.; Olbrich, E.; Jost, J.; Bertschinger, N. On Extractable Shared Information. *Entropy* **2017**, *19*, 328. [[CrossRef](#)]
34. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.
35. Wibral, M.; Lizier, J.T.; Priesemann, V. Bits from Brains for Biologically Inspired Computing. *Front. Robot. AI* **2015**, *2*, 5. [[CrossRef](#)]
36. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
37. Pica, G.; Piasini, E.; Chicharro, D.; Panzeri, S. Invariant Components of Synergy, Redundancy, and Unique Information among Three Variables. *Entropy* **2017**, *19*, 451. [[CrossRef](#)]
38. Chicharro, D. Quantifying Multivariate Redundancy with Maximum Entropy Decompositions of Mutual Information. *arXiv* **2017**, arXiv:1708.03845v2.
39. Makkeh, A.; Theis, D.O.; Vicente, R. Bivariate Partial Information Decomposition: The Optimization Perspective. *Entropy* **2017**, *19*, 530. [[CrossRef](#)]
40. Banerjee, P.K.; Rauh, J.; Montúfar, G. Computing the Unique Information. *arXiv* **2018**, arXiv:1709.07487v2.
41. Makkeh, A.; Theis, D.O.; Vicente, R. BROJA_2PID: A Robust Estimator for Bivariate Partial Information Decomposition. *Entropy* **2018**, *20*, 271. [[CrossRef](#)]
42. 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](#)]
43. 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*; Springer: Cham, Switzerland, 2012; pp. 251–269.
44. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed.; Springer: New York, NY, USA, 2002.

45. Stramaglia, S.; Angelini, L.; Wu, G.; Cortes, J.M.; Faes, L.; Marinazzo, D. Synergetic and Redundant Information Flow Detected by Unnormalized Granger Causality: Application to Resting State fMRI. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 2518–2524.
46. 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](#)]
47. Ghazi-Zahedi, K.; Langer, C.; Ay, N. Morphological Computation: Synergy of Body and Brain. *Entropy* **2017**, *19*, 456. [[CrossRef](#)]
48. Faes, L.; Marinazzo, D.; Stramaglia, S. Multiscale Information Decomposition: Exact Computation for Multivariate Gaussian Processes. *Entropy* **2017**, *19*, 408. [[CrossRef](#)]
49. Tax, T.M.S.; Mediano, P.A.M.; Shanahan, M. The Partial Information Decomposition of Generative Neural Network Models. *Entropy* **2017**, *19*, 474. [[CrossRef](#)]
50. Schwartz-Ziv, R.; Tishby, N. Opening the Black Box of Deep Neural Networks via Information. *arXiv* **2017**, arXiv:1703.00810.
51. Makkeh, A. Applications of Optimization in Some Complex Systems. Ph.D. Thesis, University of Tartu, Tartu, Estonia, 2018.
52. Domahidi, A.; Chu, E.; Boyd, S. ECOS: An SOCP Solver for Embedded Systems. In Proceedings of the European Control Conference, Zurich, Switzerland, 17–19 July 2013; pp. 3071–3076.
53. Makkeh, A.; Theis, D.O.; Vicente, R.; Chicharro, D. A Trivariate PID Estimator. 2018. Available online: https://github.com/Abzinger/MAXENT3D_PID (accessed on 21 June 2018).
54. Mehrotra, S. On the implementation of a primal-dual interior point method. *SIAM J. Optim.* **1992**, *2*, 575–601. [[CrossRef](#)]
55. Potra, F.A.; Wright, S.J. Interior-point methods. *J. Comput. Appl. Math.* **2000**, *124*, 281–302. [[CrossRef](#)]
56. James, R.G.; Crutchfield, J.P. Multivariate Dependence Beyond Shannon Information. *Entropy* **2017**, *19*, 531. [[CrossRef](#)]
57. Lizier, J.T.; Flecker, B.; Williams, P.L. Towards a Synergy-based Approach to Measuring Information Modification. In Proceedings of the 2013 IEEE Symposium on Artificial Life (ALife), Singapore, 16–19 April 2013; pp. 43–51.
58. Wibral, M.; Priesemann, V.; Kay, J.W.; Lizier, J.T.; Phillips, W.A. Partial Information Decomposition as a Unified Approach to the Specification of Neural Goal Functions. *Brain Cogn.* **2015**, *112*, 25–38. [[CrossRef](#)]
59. Banerjee, P.K.; Griffith, V. Synergy, Redundancy, and Common Information. *arXiv* **2015**, arXiv:1509.03706v1.
60. Kay, J.W.; Ince, R.A.A. Exact Partial Information Decompositions for Gaussian Systems Based on Dependency Constraints. *arXiv* **2018**, arXiv:1803.02030v1.
61. Crosato, E.; Jiang, L.; Lecheval, V.; Lizier, J.T.; Wang, X.R.; Tichit, P.; Theraulaz, G.; Prokopenko, M. Informative and Misinformative Interactions in a School of Fish. *Swarm Intell.* **2018**, *12*, 283–305. [[CrossRef](#)]
62. Sootla, S.; Theis, D.O.; Vicente, R. Analyzing Information Distribution in Complex Systems. *Entropy* **2017**, *19*, 636. [[CrossRef](#)]
63. Erwin, D.H.; Davidson, E.H. The Evolution of Hierarchical Gene Regulatory Networks. *Nat. Rev. Genet.* **2009**, *10*, 141–148. [[CrossRef](#)] [[PubMed](#)]
64. Olshausen, B.A.; Field, D.J. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? *Vis. Res.* **1997**, *37*, 3311–3325. [[CrossRef](#)]
65. Palmer, S.E.; Marre, O.; Berry, M.J.; Bialek, W. Predictive Information in a Sensory Population. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6908–6913. [[CrossRef](#)]
66. Faes, L.; Kugiumtzis, D.; Nollo, G.; Jurysta, F.; Marinazzo, D. Estimating the Decomposition of Predictive Information in Multivariate Systems. *Phys. Rev. E* **2015**, *91*, 032904. [[CrossRef](#)]
67. Chicharro, D.; Ledberg, A. Framework to Study Dynamic Dependencies in Networks of Interacting Processes. *Phys. Rev. E* **2012**, *86*, 041901. [[CrossRef](#)]
68. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [[CrossRef](#)]
69. Sun, J.; Cafaro, C.; Bollt, E. Identifying the coupling structure in complex systems through the optimal causation entropy principle. *Entropy* **2014**, *16*, 3416–3433. [[CrossRef](#)]
70. Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer Entropy: A Model-free Measure of Effective Connectivity for the Neurosciences. *J. Comput. Neurosci.* **2011**, *30*, 45–67. [[CrossRef](#)]
71. Hlaváčková-Schindler, K.; Paluš, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **2007**, *441*, 1–46. [[CrossRef](#)]

72. Vicente, R.; Wibral, M. Efficient estimation of information transfer. In *Directed Information Measures in Neuroscience*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 37–58.
73. Valdes-Sosa, P.A.; Roebroeck, A.; Daunizeau, J.; Friston, K. Effective Connectivity: Influence, Causality and Biophysical Modeling. *NeuroImage* **2011**, *58*, 339–361. [[CrossRef](#)] [[PubMed](#)]
74. Wibral, M.; Vicente, R.; Lizier, J.T. *Directed Information Measures in Neuroscience*; Springer: Berlin/Heidelberg, Germany, 2014.
75. Wibral, M.; Vicente, R.; Lindner, M. Transfer entropy in neuroscience. In *Directed Information Measures in Neuroscience*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 3–36.
76. Deco, G.; Tononi, G.; Boly, M.; Kringelbach, M.L. Rethinking Segregation and Integration: Contributions of Whole-brain Modelling. *Nat. Rev. Neurosci.* **2015**, *16*, 430–439. [[CrossRef](#)] [[PubMed](#)]
77. Daniels, B.C.; Ellison, C.J.; Krakauer, D.C.; Flack, J.C. Quantifying Collectivity. *Curr. Opin. Neurobiol.* **2016**, *37*, 106–113. [[CrossRef](#)] [[PubMed](#)]
78. Linsker, R. Local Synaptic Learning Rules Suffice to Maximize Mutual Information in a Linear Network. *Neural Comput.* **1992**, *4*, 691–702. [[CrossRef](#)]
79. Bell, J.A.; Sejnowski, T.J. An Information Maximisation Approach to Blind Separation and Blind Deconvolution. *Neural Comput.* **1995**, *7*, 1129–1159. [[CrossRef](#)] [[PubMed](#)]
80. Barrett, A.B. Exploration of Synergistic and Redundant Information Sharing in Static and Dynamical Gaussian Systems. *Phys. Rev. E* **2015**, *91*, 052802. [[CrossRef](#)]
81. Lizier, J.T. JIDT: An information-theoretic toolkit for studying the dynamics of complex systems. *Front. Robot. AI* **2014**, *1*, 11. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).