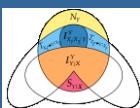
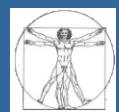




Information-Theoretic Analysis of Brain and Physiological Networks: *Methods*

Luca Faes

*Department of Engineering
University of Palermo, Italy*



OUTLINE

- **INTRODUCTION**

- Network Physiology
- Issues and challenges

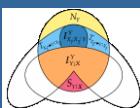
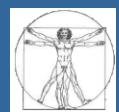
- **INFORMATION DYNAMICS: THEORY**

- Information Measures
- Information-theoretic analysis of dynamical systems
 - *New Information, Predictive Information*
 - *Information Storage, Internal Information*
 - *Information Transfer*
 - *Information Modification*
- Information Decomposition →
 - *Information Storage, Internal Information*
 - *Information Transfer*
 - *Information Modification*

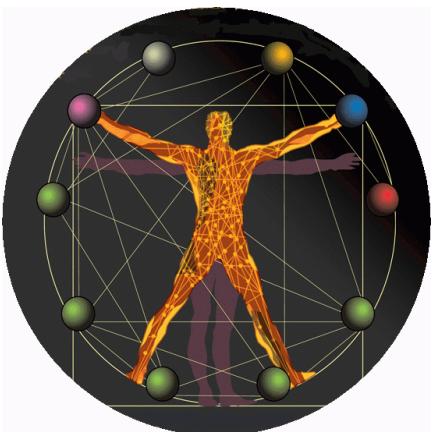
- **INFORMATION DYNAMICS: ESTIMATION**

- Linear model-based estimator
 - *Binning*
- Nonlinear model-free estimators →
 - *Kernel*
 - *Nearest neighbor*
- Challenges of model-free estimation
- Model free estimation: embedding

- **APPLICATIONS TO BRAIN AND PHYSIOLOGICAL NETWORKS
TOMORROW!**



NETWORK PHYSIOLOGY: A NEW FIELD IN SYSTEM MEDICINE AND BIOLOGY



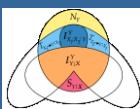
"The human organism is an integrated network where complex physiologic systems, each with its own regulatory mechanisms, continuously interact, and where failure of one system can trigger a breakdown of the entire network"

[A. Bashan et al., Nature Communications 2012]

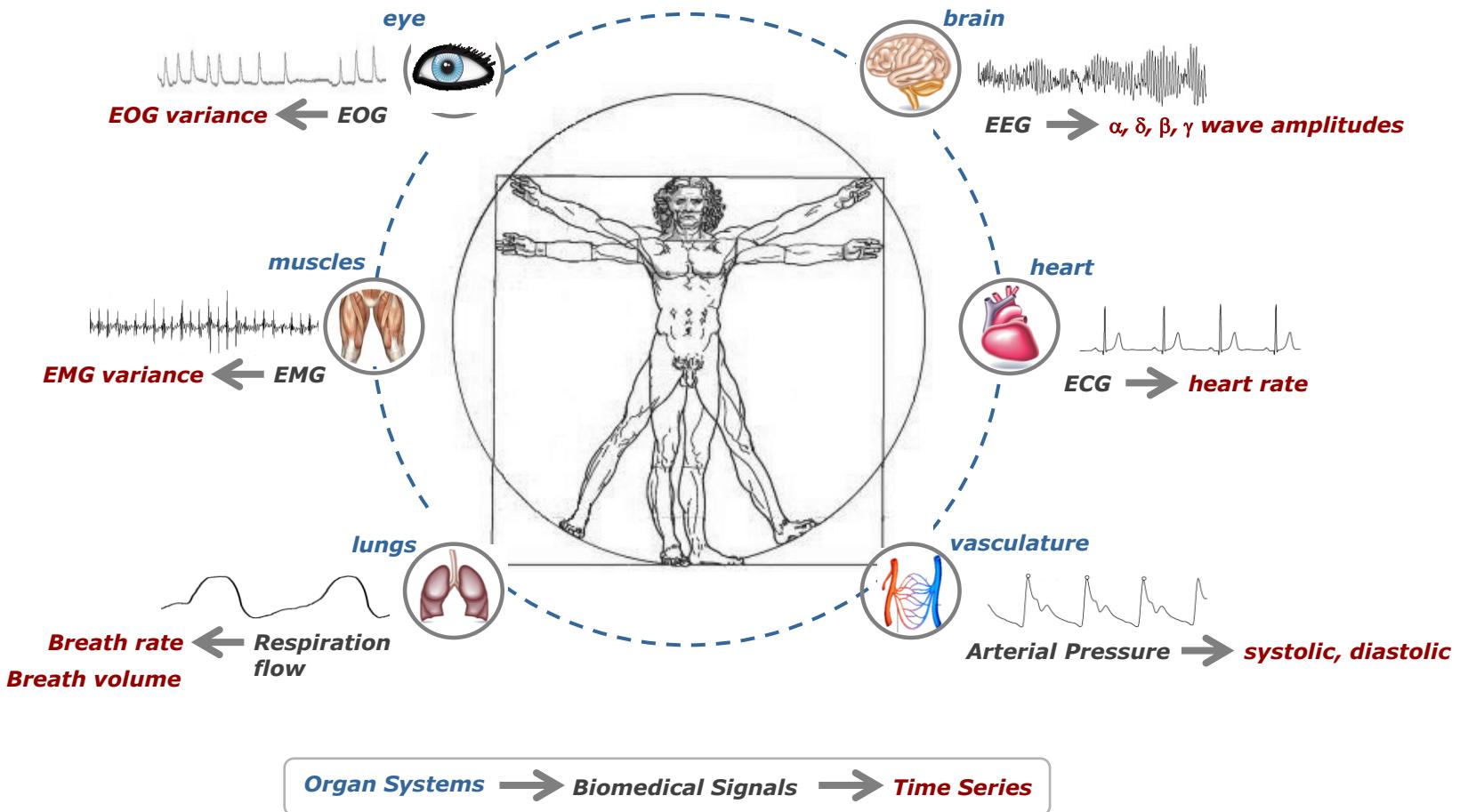
A new field, Network Physiology, is needed to probe the interactions among diverse physiologic systems

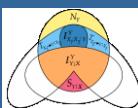
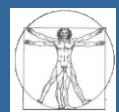
Different organ systems dynamically interact to accomplish vital functions

- **Traditional, “reductionist” approach**
To study the function of single organ in isolation
- **New approach, fostered by Network Physiology**
To look simultaneously at multiple organs
Each organ system is seen as a node of a complex network of physiological interactions



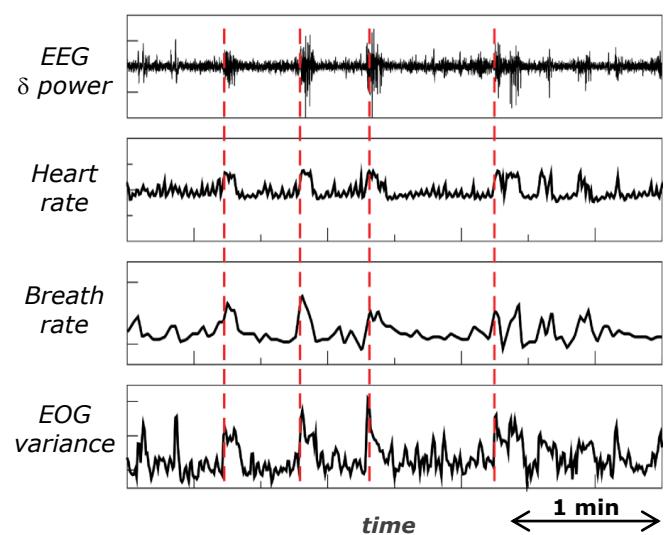
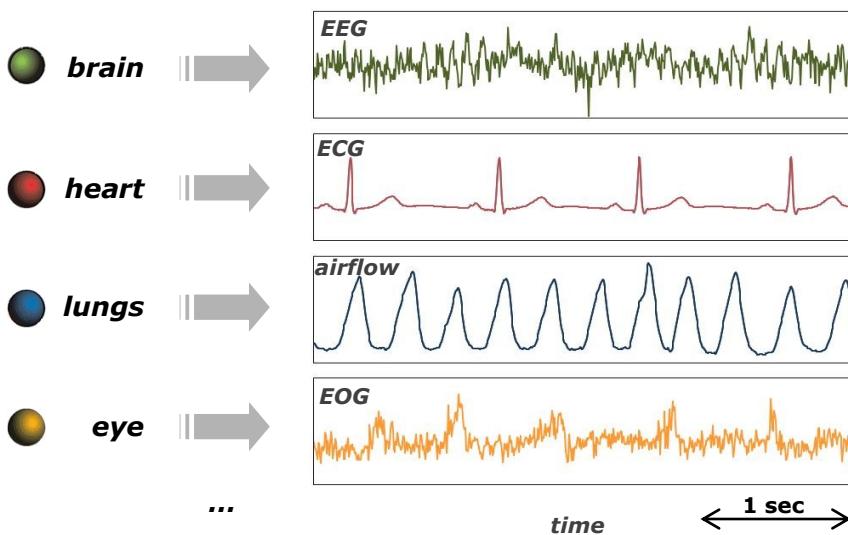
How to extract valuable information from physiological signals?





Network Physiology has high potential but is still in large part unexplored

SYSTEMS → SIGNALS → TIME SERIES

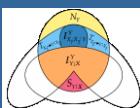


- ISSUES:**
- ✓ **Multiple network nodes**
 - ✓ **Multiple time scales**
 - ✓ **non-linear dynamics**



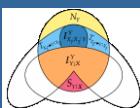
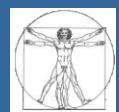
- ✓ **Multivariate measures**
- ✓ **Multi-scale measures**
- ✓ **Nonlinear measures**

- Challenging problems for physicists, engineers and physiologists
- We face these issues with the unifying framework of **INFORMATION DYNAMICS**



INFORMATION DYNAMICS: THEORY

- **Information-theoretic analysis of dynamical systems**
- **Information Measures**
 - *New Information, Predictive Information*
 - *Information Storage, Internal Information*
 - *Information Transfer*
 - *Information Modification*
- **Information Decomposition** →
 - *Information Storage, Internal Information*
 - *Information Transfer*
 - *Information Modification*

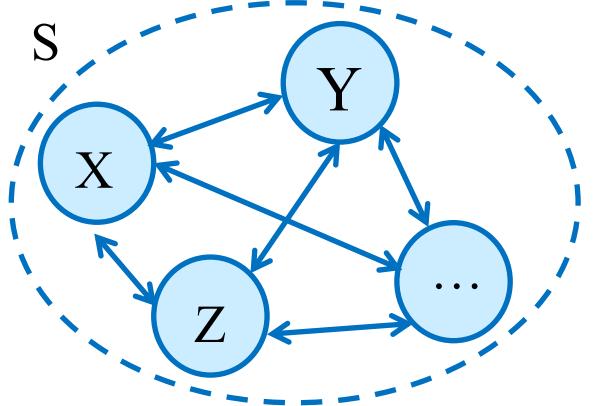


Physiological Networks

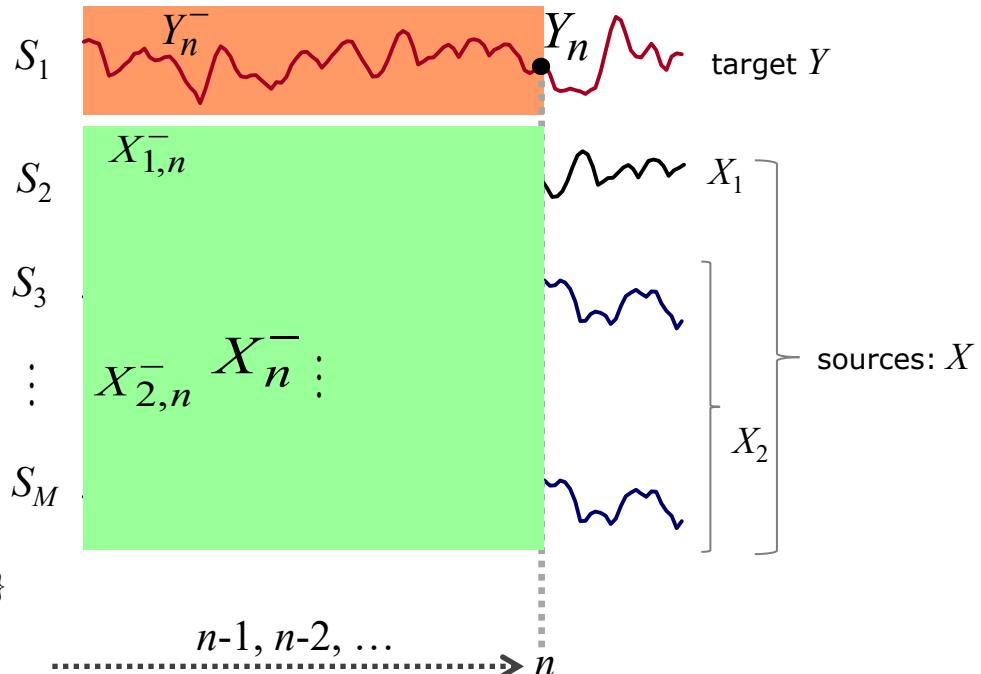


Networks of Dynamical systems

- Dynamic System $S = \{S_1, \dots, S_M\}$



- Dynamic Process S



- With reference to a target system Y :

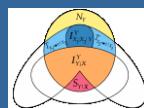
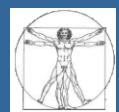
$$X = \{X_1, \dots, X_{M-1}\} \rightarrow S = \{X_1, \dots, X_{M-1}, Y\} = \{X, Y\}$$

- **Investigation of Statistical dependencies:**

✓ **SELF effects:** $Y_n^- \rightarrow Y_n$

✓ **CAUSAL effects:** $X_n^- \rightarrow Y_n$

✓ **INTERACTION effects:** $(X_{1,n}^- \leftrightarrow X_{2,n}^-) \rightarrow Y_n$

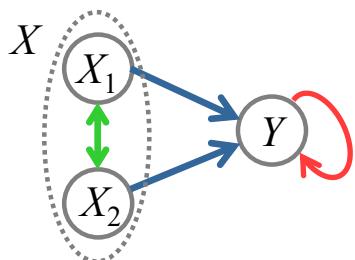


THE FRAMEWORK OF INFORMATION DYNAMICS

- **Information dynamics:**

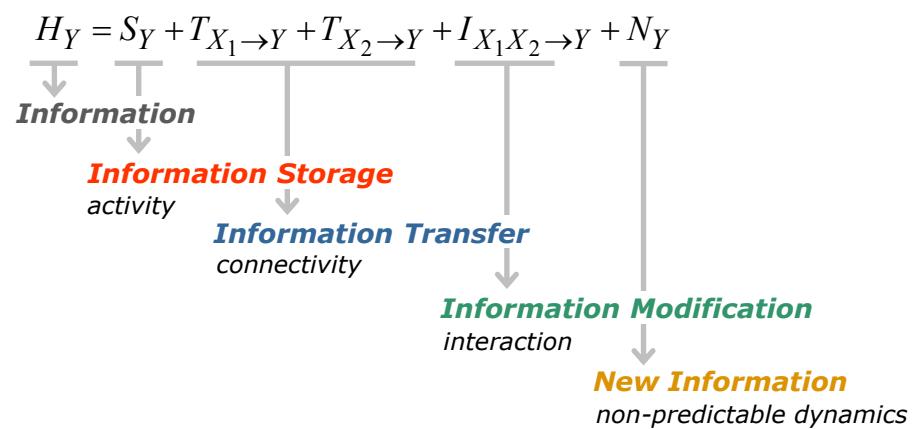
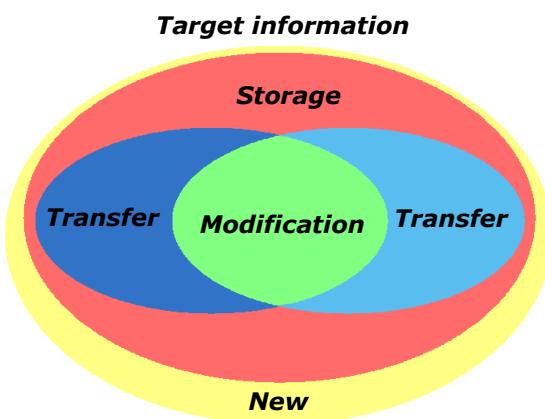
a new time series analysis approach to quantify how information is processed in a network of dynamical systems

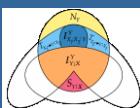
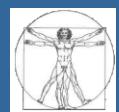
$$S = \{X, Y\} = \{X_1, X_2, Y\}$$



- ✓ SELF effects: $Y_n^- \rightarrow Y_n$ → **Information storage
Internal Information**
- ✓ CAUSAL effects: $X_n^- \rightarrow Y_n$ → **Information transfer**
- ✓ INTERACTION effects: $(X_{1,n}^- \leftrightarrow X_{2,n}^-) \rightarrow Y_n$ → **Information modification**

- **Decomposition of the “information” contained in the target process**





INFORMATION MEASURES

- **ENTROPY:** $H(X) = -E[\log p(x)]$

$$\begin{array}{ll} \text{Discrete variable} \rightarrow H(X) = - \sum_{x \in \Omega_X} p(x) \log p(x) & \text{Continuous variable} \rightarrow H(X) = - \int_{D_X} p(x) \log p(x) dx \end{array}$$

Measures the information contained in X as the average uncertainty about its outcomes

- **JOINT ENTROPY:** $H(X, Y) = -E[\log p(x, y)]$

Information contained in X and Y considered as a vector variable (X, Y)

- **CONDITIONAL ENTROPY:** $H(X | Y) = -E[\log p(x | y)] \leftarrow p(x | y) = \frac{p(x, y)}{p(y)}$

Residual information contained in X when Y is known

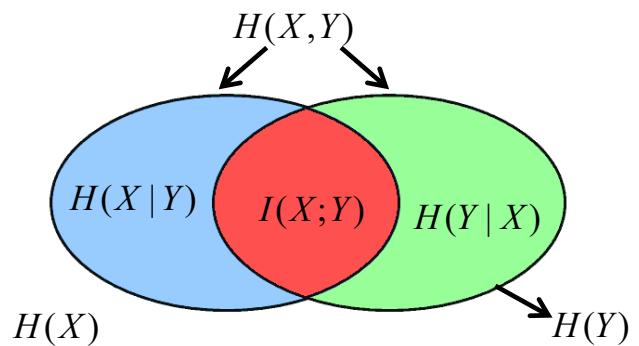
- **MUTUAL INFORMATION (MI):** $I(X; Y) = E\left[\log \frac{p(x, y)}{p(x)p(y)}\right]$

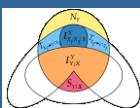
Uncertainty about X resolved by knowing Y

Uncertainty about Y resolved by knowing X

$$H(X, Y) = H(X) + H(Y | X) = H(Y) + H(X | Y)$$

$$\begin{aligned} I(X; Y) &= I(Y; X) = H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X | Y) \\ &= H(Y) - H(Y | X) \end{aligned}$$



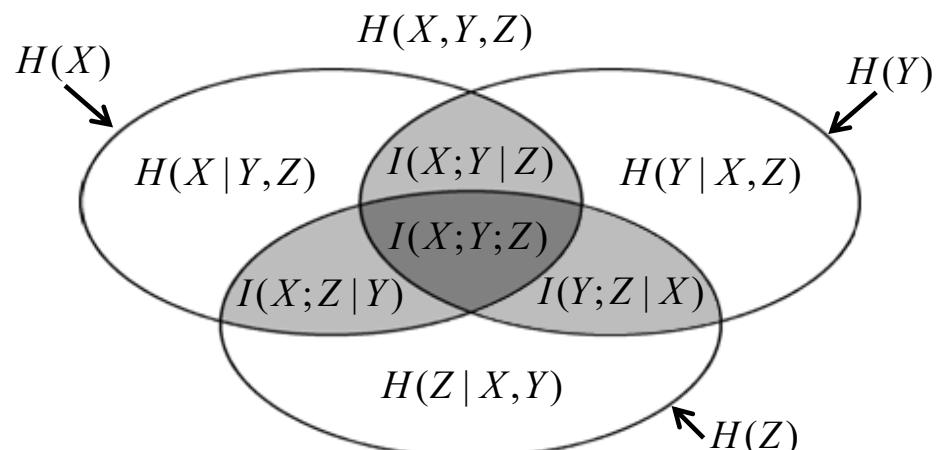


INFORMATION MEASURES

- **CONDITIONAL MUTUAL INFORMATION:** $I(X;Y|Z) = \text{E}\left[\log \frac{p(x,y|z)}{p(x|z)p(y|z)}\right]$
 Information shared between X and Y which is not shared with Z
 Residual mutual information between X and Y when Z is known
- **INTERACTION INFORMATION:** $I(X;Y;Z) = I(X;Y) + I(X;Z) - I(X;Y,Z)$
 Information that X shares with Y and Z when they are taken individually but not when they are taken together

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= H(Y|Z) - H(Y|X,Z) \end{aligned}$$

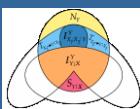
$$\begin{aligned} I(X;Y;Z) &= I(X;Y) - I(X;Y|Z) \\ &= I(X;Z) - I(X;Z|Y) \\ &= I(Y;Z) - I(Y;Z|X) \end{aligned}$$



- **The interaction information can be negative!**

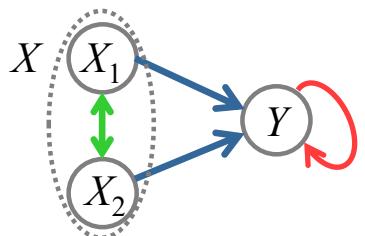
$$I(X;Y) + I(X;Z) > I(X;Y,Z) \Rightarrow I(X;Y;Z) > 0 \quad \text{REDUNDANCY}$$

$$I(X;Y,Z) > I(X;Y) + I(X;Z) \Rightarrow I(X;Y;Z) < 0 \quad \text{SYNTERGY}$$



INFORMATION DYNAMICS TOTAL INFORMATION ABOUT THE TARGET

- Observed multivariate system S :

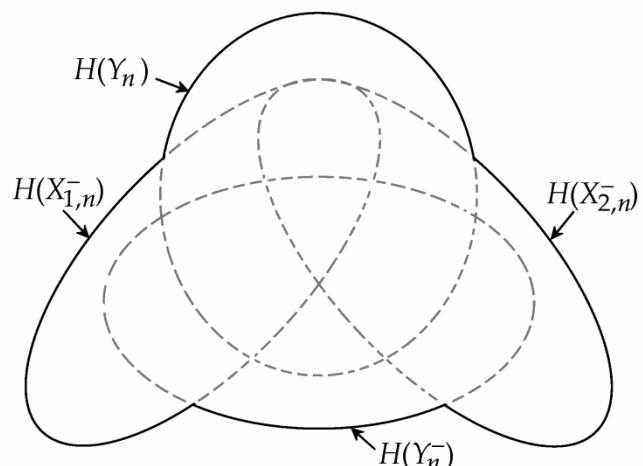


- Process dynamics:

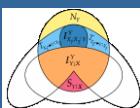
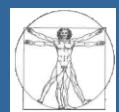


- Total process information:

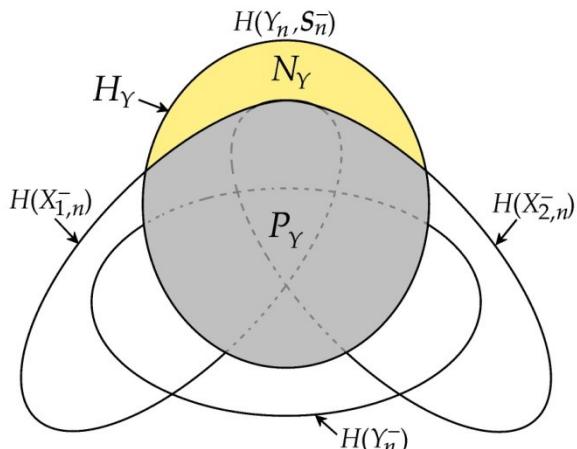
$$H(Y_n, S_n^-) = H(Y_n, Y_n^-, X_{1,n}^-, X_{2,n}^-)$$



Information contained in the network which is relevant to the target process Y



TARGET INFORMATION DECOMPOSITION



$$H(Y_n) = I(Y_n; \mathbf{X}_n^-, Y_n^-) + H(Y_n | \mathbf{X}_n^-, Y_n^-)$$

↓ H_Y
Information ↓ P_Y
Predictive Information ↓ N_Y
New created Information

- **Present Information** about Y : $H_Y = H(Y_n)$

Information contained in the present of the process Y



Uncertainty about the present state of the target

- **Predictive Information** about Y : $P_Y = I(Y_n; Y_n^-, X_n^-)$

Information contained in the past of $S=(X,Y)$ that can be used to predict the present of the target Y



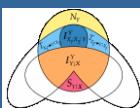
Predictability of the target given the past network states

- **New information** about Y : $N_Y = H(Y_n | Y_n^-, X_n^-)$

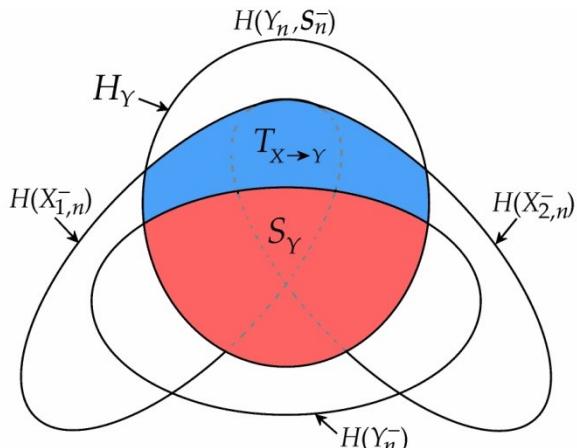
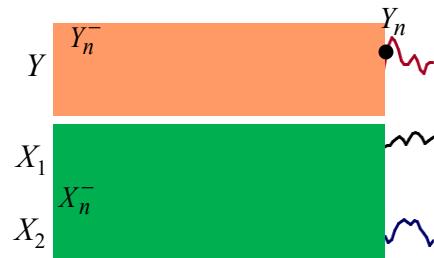
Information contained in the present of Y that cannot be predicted from the past of $S=(X,Y)$



Information generated in the target by the state transition



PREDICTIVE INFORMATION DECOMPOSITION



$$I(Y_n; \mathbf{X}_n^-, Y_n^-) = I(Y_n; Y_n^-) + I(Y_n; \mathbf{X}_n^- | Y_n^-)$$

P_Y
 Information Storage
 Information Transfer

- **Predictive Information** about Y : $P_Y = I(Y_n; Y_n^-, X_n^-)$

Information contained in the past of $S=(X, Y)$ that can be used to predict the present of the target Y

→ **Predictability of the target given the network past states**

- **Information Storage** in Y : $S_Y = I(Y_n; Y_n^-)$

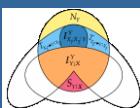
Information contained in the past of Y that can be used to predict its present

→ **Predictability of the target from its own past states**

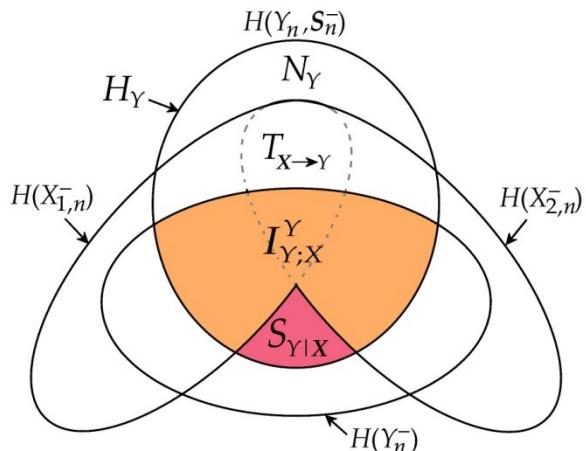
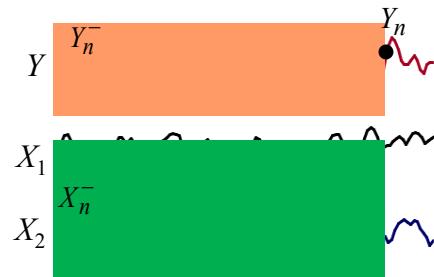
- **Information transfer** from X to Y : $T_{X \rightarrow Y} = I(Y_n; X_n^- | Y_n^-)$

Information contained in the past of X that can be used to predict the present of Y above and beyond the information contained in the past of Y

→ **Causal interactions from all sources to the target**



INFORMATION STORAGE DECOMPOSITION



$$I(Y_n; Y_n^-) = I(Y_n; Y_n^- | \mathbf{X}_n^-) + I(Y_n; Y_n^-; \mathbf{X}_n^-)$$

↓ ↓ ↓
 S_Y $S_{Y|X}$ $I_{Y|X}^Y$
Information Storage **Internal Information** **Interaction Information Storage**

- **Information Storage** in Y : $S_Y = I(Y_n; Y_n^-)$

Information contained in the past of Y that can be used to predict its present



*Predictability of the target
From its own history*

- **Internal information** in Y : $S_{Y|X} = I(Y_n; Y_n^- | \mathbf{X}_n^-)$

Information contained in the past of Y that can be used to predict its present above and beyond the information contained in the past of X



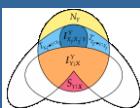
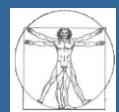
*Causal self-interactions
in the target*

- **Interaction information storage** of Y : $I_{Y|X}^Y = I(Y_n; Y_n^-; \mathbf{X}_n^-)$

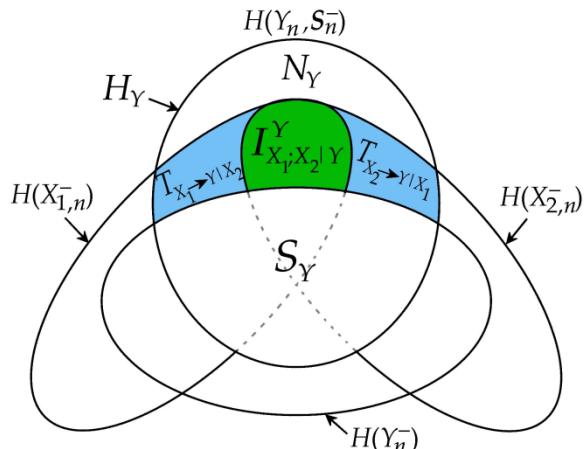
Information contained in the past $S=(X,Y)$ that can be used to predict the present of Y when the past of Y and in the past of X are taken individually but not when they are taken together



*Redundant or synergistic
interactions contributing to
storage*



INFORMATION TRANSFER DECOMPOSITION



$$I(Y_n; \mathbf{X}_n^- | Y_n^-) = I(Y_n; X_{1,n}^- | Y_n^-, X_{2,n}^-) + I(Y_n; X_{2,n}^- | Y_n^-, X_{1,n}^-) + I(Y_n; X_{1,n}^-; X_{2,n}^- | Y_n^-)$$

$T_{X \rightarrow Y}$
Information Transfer

$T_{X_1 \rightarrow Y|X_2}$
Conditional information transfer

$T_{X_2 \rightarrow Y|X_1}$

$I_{X_1;X_2|Y}^Y$
Interaction Information Transfer

- **Information transfer** from X to Y : $T_{X \rightarrow Y} = I(Y_n; X_n^- | Y_n^-)$

Information contained in the past of X that can be used to predict the present of Y above and beyond the information contained in the past of Y

→ *Causal interactions from all sources to the target*

- **Conditional information transfer**: $T_{X_1 \rightarrow Y|X_2} = I(Y_n; X_{1,n}^- | Y_n^-, X_{2,n}^-)$

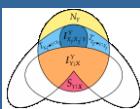
Information contained in the past of X_1 that can be used to predict the present of Y above and beyond the information contained in the past of Y and X_2

→ *Causal interactions from one source to the target*

- **Interaction information transfer**: $I_{X_1;X_2|Y}^Y = I(Y_n; X_{1,n}^-; X_{2,n}^- | Y_n^-)$

Information contained in the past of X_1 and X_2 that can be used to predict the present of Y when X_1 and X_2 are taken individually but not when they are taken together

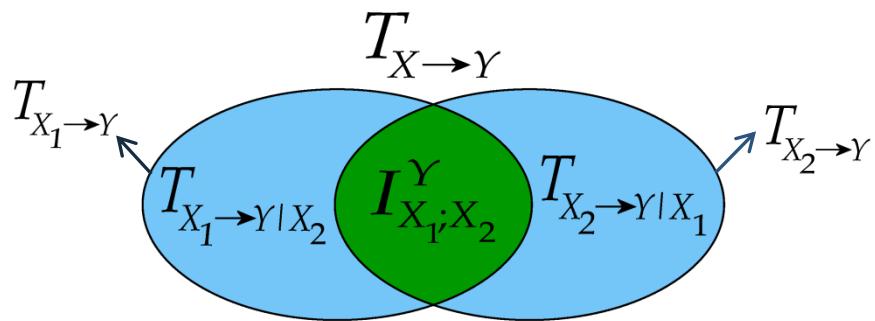
→ *Redundant or synergistic interactions contributing to transfer*



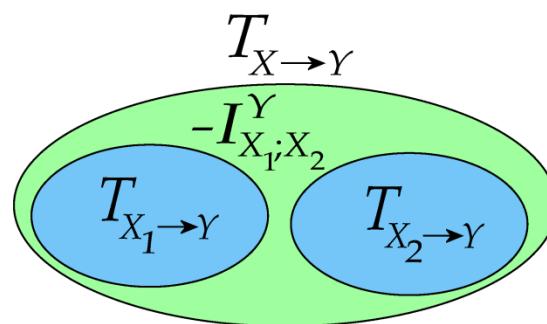
INFORMATION MODIFICATION: REDUNDANCY AND SYNERGY

- **Interpretation of Information Modification:** $I_{X_1;X_2}^Y = T_{X_1 \rightarrow Y} + T_{X_2 \rightarrow Y} - T_{X_1, X_2 \rightarrow Y}$

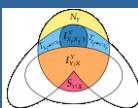
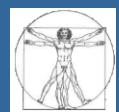
REDUNDANCY: $T_{X_1, X_2 \rightarrow Y} < T_{X_1 \rightarrow Y} + T_{X_2 \rightarrow Y} \rightarrow I_{X_1;X_2}^Y > 0$



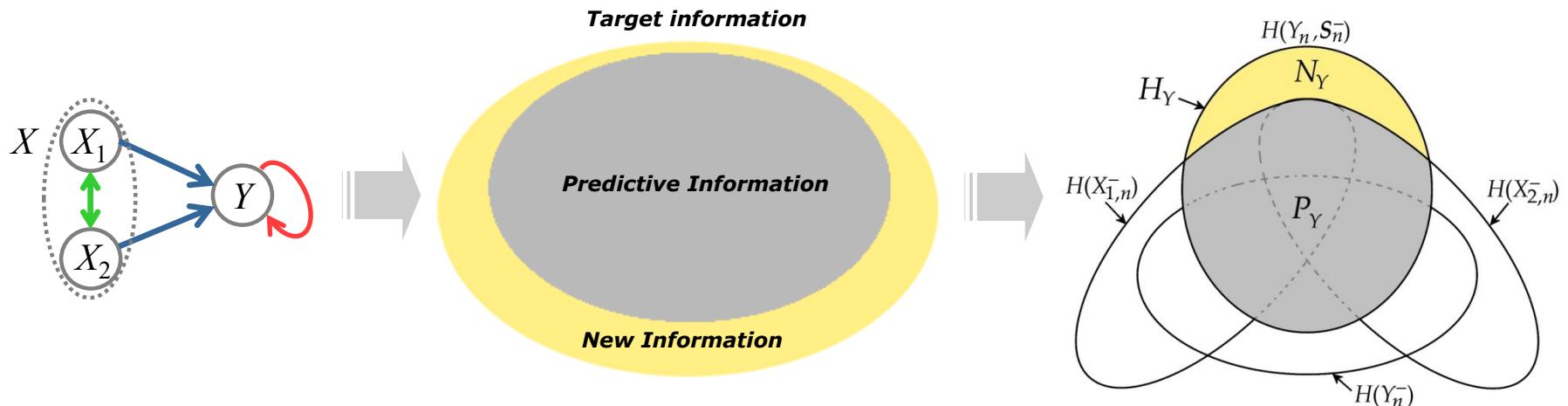
SYNERGY: $T_{X_1, X_2 \rightarrow Y} > T_{X_1 \rightarrow Y} + T_{X_2 \rightarrow Y} \rightarrow I_{X_1;X_2}^Y < 0$



Interaction information can be negative: synergy!



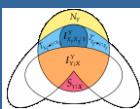
THE FRAMEWORK OF INFORMATION DYNAMICS



$$\begin{aligned}
 H_Y &= N_Y + P_Y = N_Y + S_Y + T_{X \rightarrow Y} = N_Y + S_{Y|X} + I_{Y;X}^Y + T_{X_1 \rightarrow Y|X_2} + T_{X_2 \rightarrow Y|X_1} + I_{X_1;X_2|Y}^Y \\
 &\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \\
 \textbf{Information} &\qquad \textbf{Predictive predictable dynamics} & \textbf{Information Storage} &\qquad \textbf{Internal autonomous dynamics} & \textbf{Interaction Storage} &\qquad \textbf{Conditional Transfer direct causal connectivity} & \textbf{Interaction Transfer} \\
 &\downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \qquad \downarrow \\
 \textbf{New Information} &\qquad \textit{non-predictable dynamics} & \textbf{Information Transfer} &\qquad \textit{causal connectivity} & \textbf{Information Modification} &\qquad \textit{interaction between systems}
 \end{aligned}$$

L Faes, A Porta, G Nollo, 'Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics', *Entropy, special issue on "Entropy and Cardiac Physics"*, 2015, 17:277-303.

L Faes, A Porta, G Nollo, M Javorka, 'Information decomposition in multivariate systems: definitions, implementation and application to cardiovascular networks', *Entropy, special issue on Multivariate Entropy Measures and their applications*, 2017, 19(1), 5



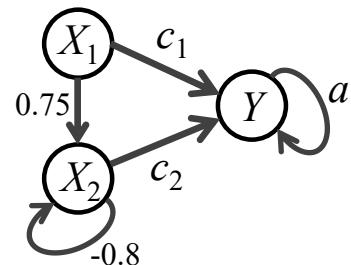
THEORETICAL EXAMPLE

- Simulated VAR process:

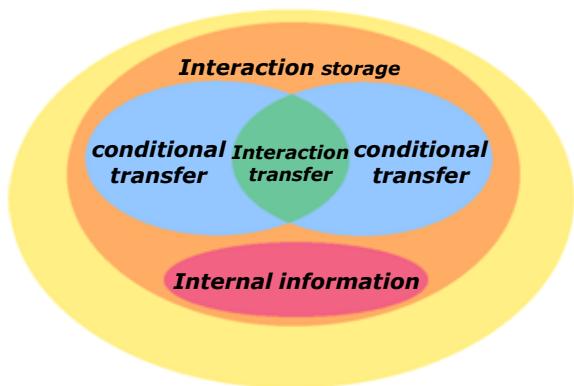
$$X_{1,n} = U_n$$

$$X_{2,n} = -0.8 \cdot X_{2,n-2} + 0.75 \cdot X_{1,n-1} + V_n$$

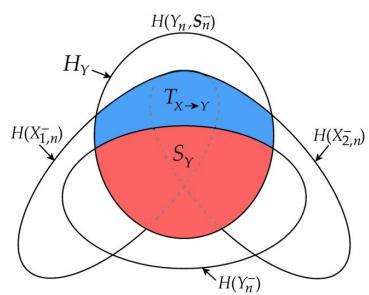
$$Y_n = -a \cdot Y_{n-2} + c_1 X_{1,n-1} + c_2 X_{2,n-1} + W_n$$



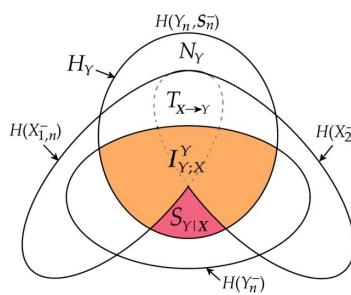
- Computation of all measures of information dynamics (theoretical values)



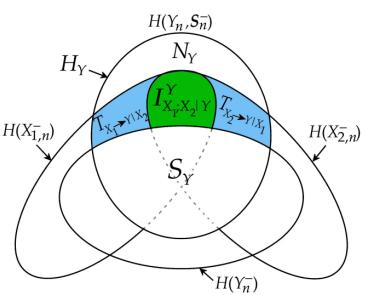
Predictive Information Decomposition (PID)

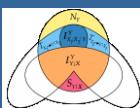
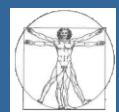


Information Storage Decomposition (ITD)



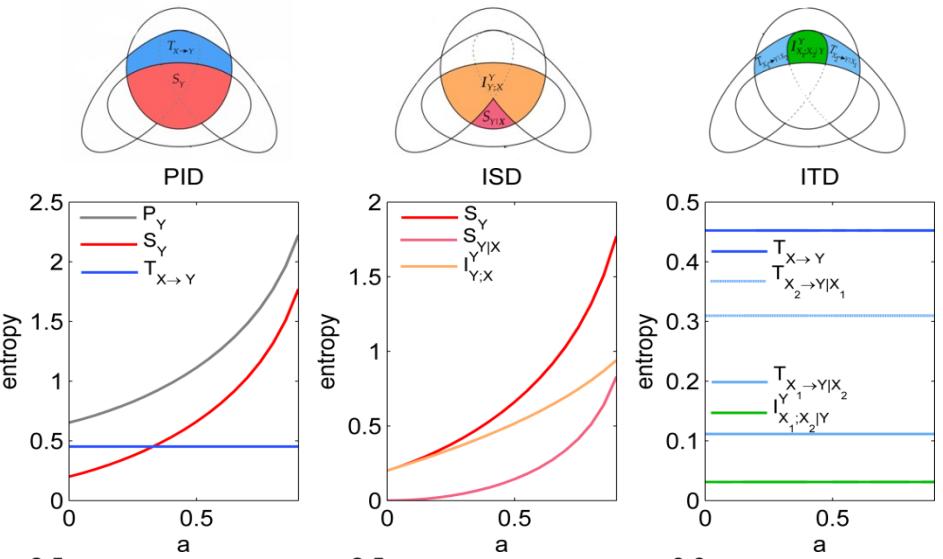
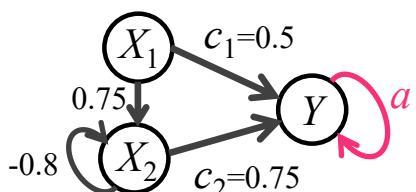
Information Transfer Decomposition (ITD)



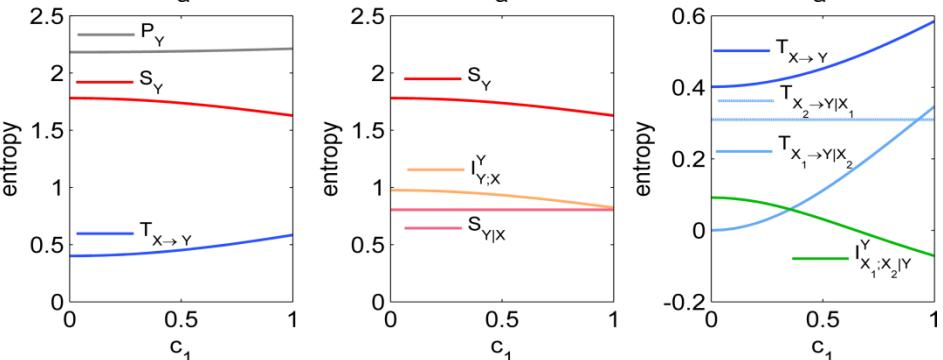
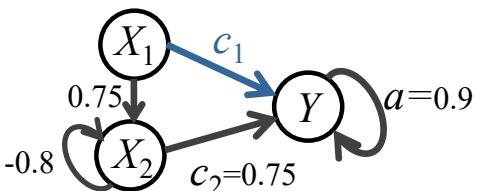


THEORETICAL EXAMPLE: results

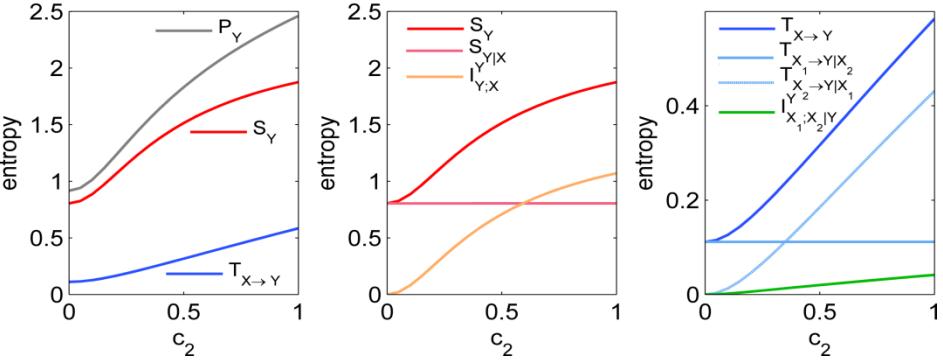
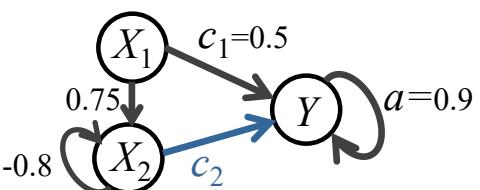
- Vary internal dynamics of Y :**

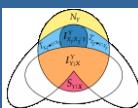
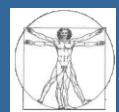


- Vary causal coupling $X_1 \rightarrow Y$:**



- Vary causal coupling $X_2 \rightarrow Y$:**





Information Modification: PARTIAL INFORMATION DECOMPOSITION

$$T_{X_1, X_2 \rightarrow Y} = U_{X_1 \rightarrow Y} + U_{X_2 \rightarrow Y} + R_{X_1; X_2}^Y + S_{X_1; X_2}^Y$$

$$T_{X_1 \rightarrow Y} = U_{X_1 \rightarrow Y} + R_{X_1; X_2}^Y, \quad T_{X_2 \rightarrow Y} = U_{X_2 \rightarrow Y} + R_{X_1; X_2}^Y$$

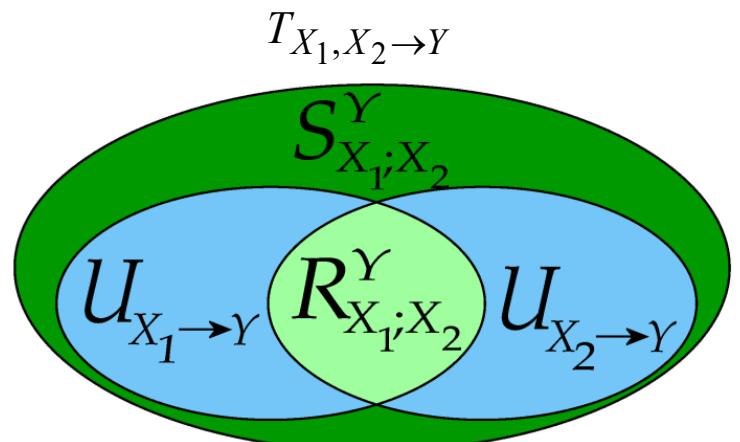
[P.L. Williams & R.D. Beer, ArXiv 1004.2515, 2010]

- **Information Modification:**

$U_{X_1 \rightarrow Y}, U_{X_2 \rightarrow Y}$ **UNIQUE TRANSFER ENTROPY**

$R_{X_1; X_2}^Y$ **REDUNDANT TRANSFER ENTROPY**

$S_{X_1; X_2}^Y$ **SYNERGISTIC TRANSFER ENTROPY**



Relation with interaction information:

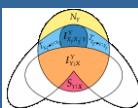
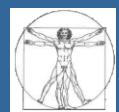
$$I_{X_1; X_2}^Y = R_{X_1; X_2}^Y - S_{X_1; X_2}^Y$$

- **PID cannot be achieved through classic information theory**

❖ **Minimum Mutual Information PID** [A.B. Barrett, Phys. Rev. E 91, 2015]

$$R_{X_1; X_2}^Y = \min\{T_{X_1 \rightarrow Y}, T_{X_2 \rightarrow Y}\} \quad (\text{the other quantities easily follow from the PID})$$

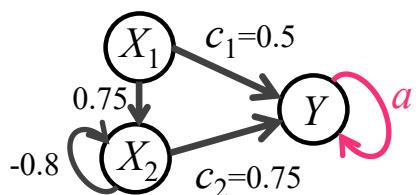
- ✓ Unique and redundant TE terms depend on the marginal distributions of (X_n^-, Y_n) and (Z_n^-, Y_n)
Only the synergistic TE terms depends on the joint distribution of (X_n^-, Z_n^-, Y_n)
- ✓ The minimum mutual information is unique for Gaussian systems



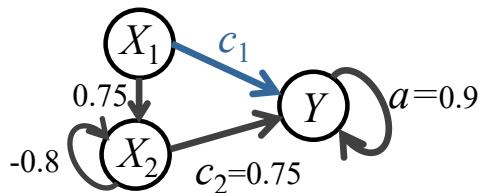
THEORETICAL EXAMPLE:

Interaction Information Decomposition vs. Partial Information Decomposition

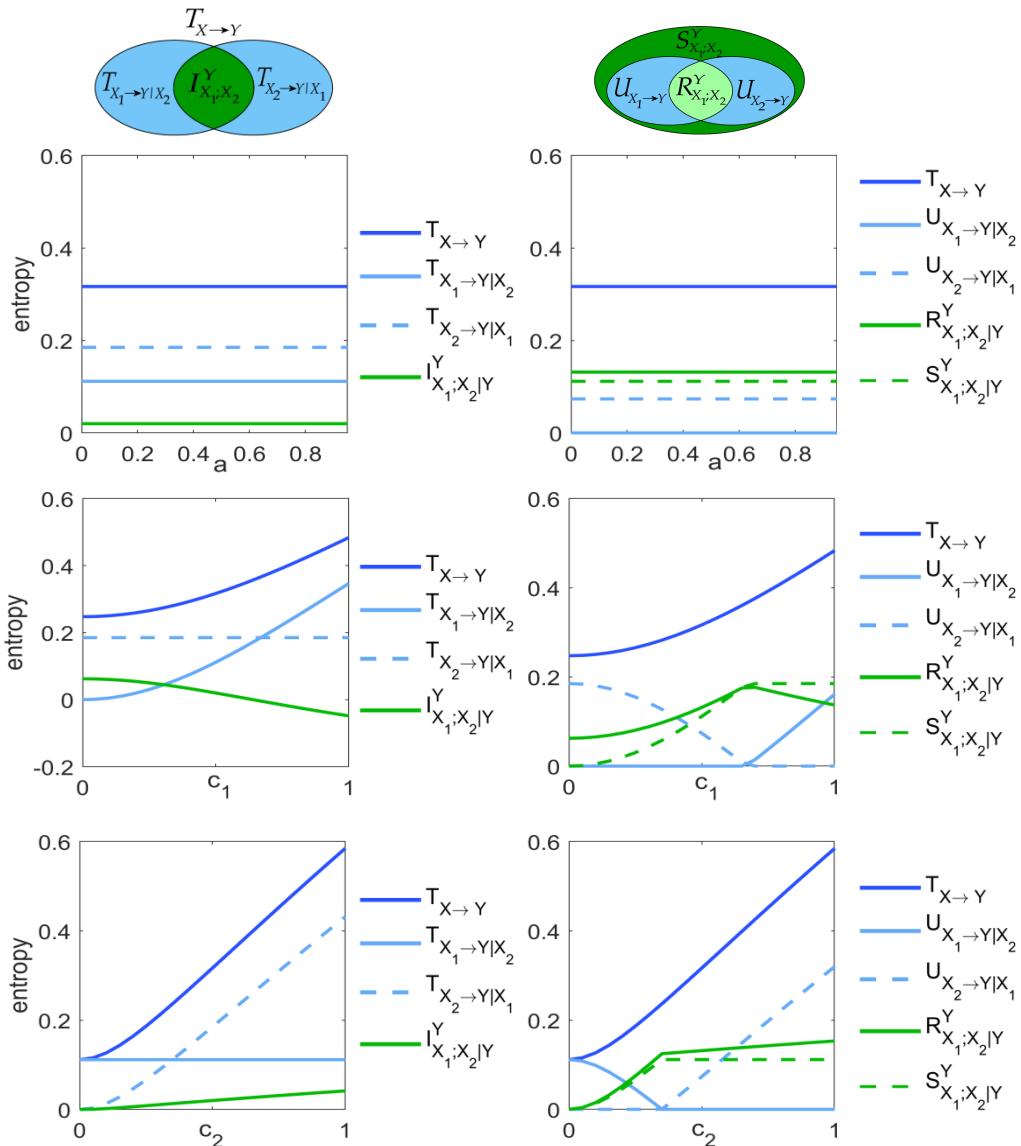
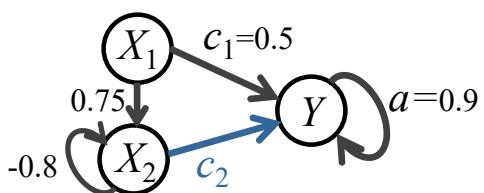
- **Vary internal dynamics of Y :**

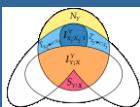
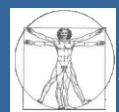


- **Vary causal coupling $X_1 \rightarrow Y$:**



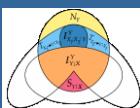
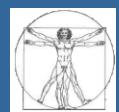
- **Vary causal coupling $X_2 \rightarrow Y$:**





INFORMATION DYNAMICS: ESTIMATION

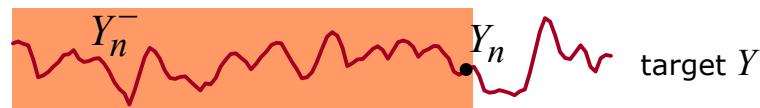
- Linear model-based estimator
- Nonlinear model-free estimators →
 - *Binning*
 - *Kernel*
 - *Nearest neighbor*
- Challenges of model-free estimation



PRACTICAL COMPUTATION OF INFORMATION DYNAMICS

- All measures of Information dynamics are expressed in terms of measures of **(conditional) entropy**, **(conditional) mutual information**, or **interaction information**
- **Estimation of entropy and conditional entropy for variables with different dimension**

✓ Example: Information Storage



$$S_Y = I(Y_n; Y_n^-) = H(Y_n) - H(Y_n | Y_n^-) = H(Y_n) - H(Y_n, Y_n^-) + H(Y_n^-)$$

Approximation of the past history

$$Y_n^- = [Y_{n-1} Y_{n-2} \dots] \quad \rightarrow \quad Y_n^- \cong Y_n^L = [Y_{n-1} Y_{n-2} \dots Y_{n-L}]$$

Computation

Discrete variables

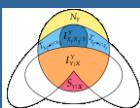
$$H(Y_n) = - \sum_{y_n \in \Omega_Y} p(y_n) \log p(y_n)$$

$$H(Y_n^-) \cong H(Y_n^L) = - \sum_{y_{n-1}} \dots \sum_{y_{n-L}} p(y_n^L) \log p(y_n^L)$$

Continuous variables

$$H(Y_n) = - \int_{D_Y} p(y_n) \log p(y_n) dy_n$$

$$H(Y_n^-) \cong H(Y_n^L) = - \int_{D_Y} \dots \int_{D_Y} \dots p(y_n^L) \log p(y_n^L) dy_{n-1} \dots dy_{n-L}$$



PARAMETRIC ESTIMATION: LINEAR METHOD

- Exact Computation under the **assumption of Gaussianity**

- Generic scalar variable $V \in N(0, \sigma(V)) \Rightarrow H(V) = \frac{1}{2} \ln 2\pi e \cdot \sigma(V)$

Variance of V : $\sigma(V) = E[V^2]$
- Generic d -dimensional vector variable Z : $H(Z) = \frac{1}{2} \ln(2\pi e)^d \cdot |\Sigma(Z)|$

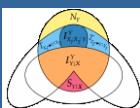
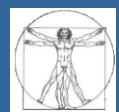
Covariance matrix of Z : $\Sigma(Z) = E[ZZ^T]$
- Conditional entropy:** $H(V | Z) = H(V, Z) - H(Z) = \frac{1}{2} \ln 2\pi e \frac{|\Sigma(V, Z)|}{|\Sigma(Z)|}$

- Main result from [Barnett et al, Phys Rev Lett 2009]:

$$\frac{|\Sigma(V, Z)|}{|\Sigma(Z)|} = \left| \sigma(V) - \Sigma(V; Z) \Sigma(Z)^{-1} \Sigma(V; Z)^T \right|$$

\downarrow
 $\sigma(V | Z)$ **Partial Variance** of V given Z
Variance of the prediction error of a linear regression of V on Z

$$H(V | Z) = \frac{1}{2} \ln 2\pi e \sigma(V | Z)$$



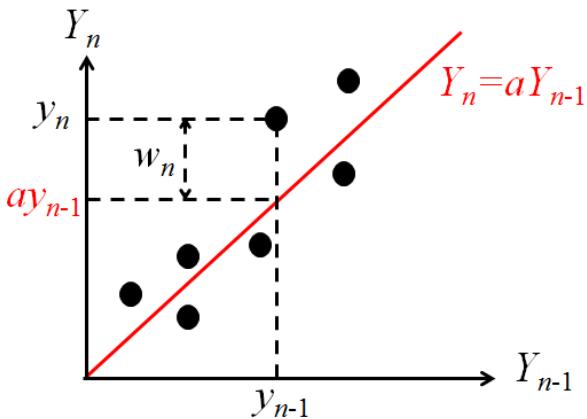
PARAMETRIC ESTIMATION: LINEAR METHOD

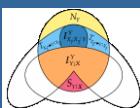
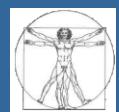
- **Entropy of Y_n :** $H_Y = H(Y_n) = \frac{1}{2} \ln 2\pi e \cdot \sigma(Y_n)$
 - **Conditional Entropy of Y_n given $Y_n^- \cong Y_n^L$:** $H(Y_n | Y_n^-) \cong H(Y_n | Y_n^L) = \frac{1}{2} \ln 2\pi e \cdot \sigma(Y_n | Y_n^L)$
- ↑
linear regression of Y_n on Y_{n-1}, \dots, Y_{n-L} : $Y_n = a_1 Y_{n-1} + \dots + a_L Y_{n-L} + W_n$

- **Information Storage of Y :** $S_Y = \frac{1}{2} \ln \frac{\sigma(Y_n)}{\sigma(Y_n | Y_n^L)} = \frac{1}{2} \ln \frac{\sigma_Y^2}{\sigma_W^2}$
 $S_Y = I(Y_n; Y_n^-) = H(Y_n) - H(Y_n | Y_n^-)$

- Example: $L=1$

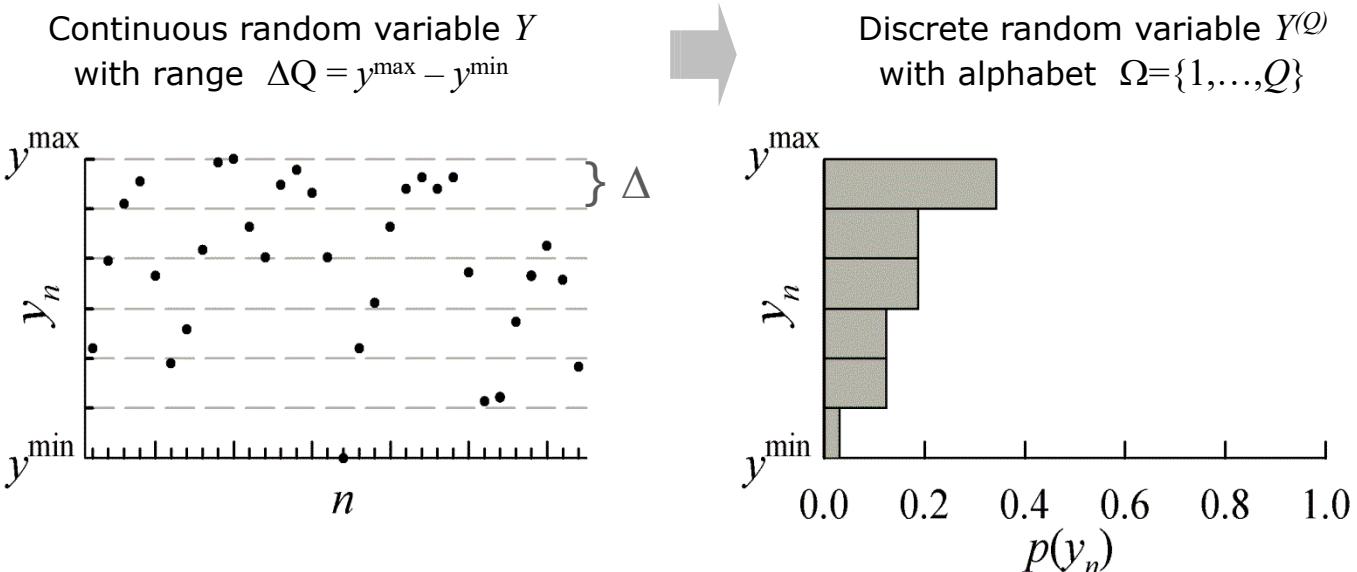
$$Y_n^L \cong Y_n^1 = Y_{n-1}$$





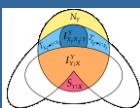
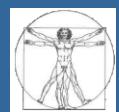
NONPARAMETRIC ESTIMATION: BINNING METHOD

- Histogram method - quantization



- Entropy:**

$$p(q) = p(Y_n = q) \cong \frac{N(q)}{N} \quad \Rightarrow \quad \hat{H}_Y = \hat{H}(Y_n) = - \sum_{q \in \Omega} p(Y_n = q) \log p(Y_n = q)$$



NONPARAMETRIC ESTIMATION: BINNING METHOD

- **System Information of Y :** $H_Y = H(Y_n)$
- **Conditional Entropy of Y_n given $Y_n^- \cong Y_n^L$:** $H(Y_n | Y_n^-) \cong \hat{H}(Y_n | Y_n^L) = \hat{H}(Y_n, Y_n^L) - \hat{H}(Y_n^L)$
- **Information Storage of Y :** $\hat{S}_Y = \hat{H}(Y_n) - \hat{H}(Y_n | Y_n^L) = \hat{H}(Y_n) - \hat{H}(Y_n, Y_n^L) + \hat{H}(Y_n^L)$

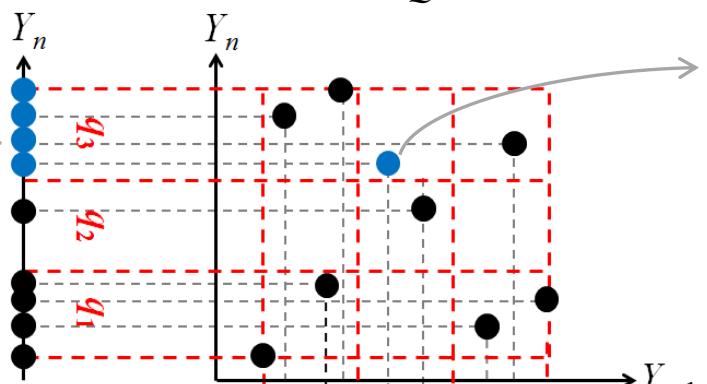
- Example: $L=1$

$$Y_n^L \cong Y_n^1 = Y_{n-1}$$

$$p(Y_n = q_3) \cong \frac{N(q_3)}{N}$$

$$\hat{H}(Y_n)$$

Uniform partition with $Q=3$ bins

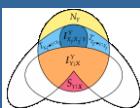
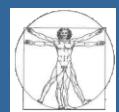


$$p(Y_n = q_3, Y_{n-1} = q_2) \cong \frac{N(q_3, q_2)}{N}$$

$$\hat{H}(Y_n, Y_{n-1})$$

$$p(Y_{n-1} = q_2) \cong \frac{N(q_2)}{N}$$

$$\hat{H}(Y_{n-1})$$



NONPARAMETRIC ESTIMATION: KERNEL METHOD

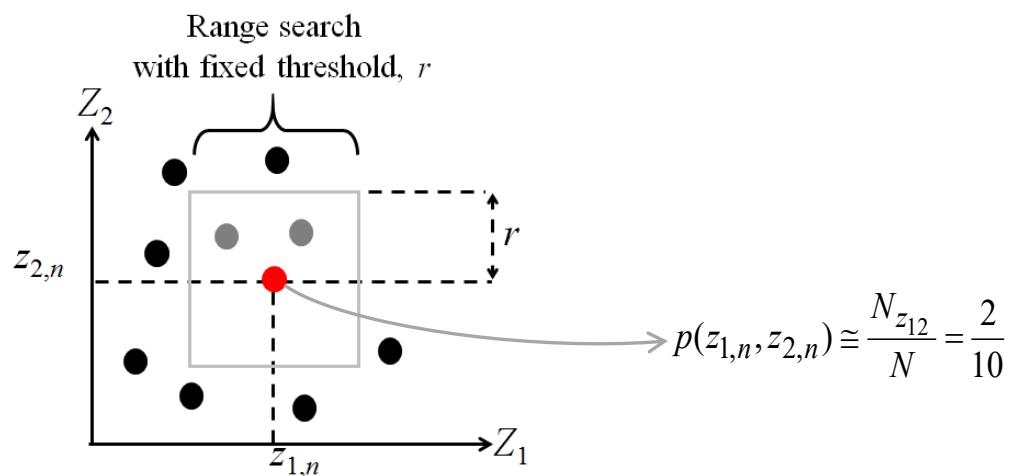
- Kernel estimate of a continuous probability distribution

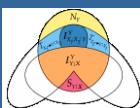
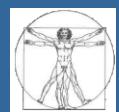
- Generic d -dimensional vector variable Z : $\hat{p}(z_n) = \frac{1}{N} \sum_{i=1}^N K(\|z_n - z_i\|)$

Heaviside kernel function: $K = \Theta(\|z_n - z_i\|) = \begin{cases} 1 & , \|z_n - z_i\| \leq r \\ 0 & , \|z_n - z_i\| > r \end{cases}$

Chebyshev distance: $\|z_n - z_i\| = \max_{1 \leq k \leq d} |z_{n,k} - z_{i,k}|$

- Entropy:** $H(Z) = -E[\log p(z)] \cong -\log \frac{1}{N} \sum_{i=1}^N \hat{p}(z_i) = -\log \langle \hat{p}(z) \rangle$
- Example: $d=2$



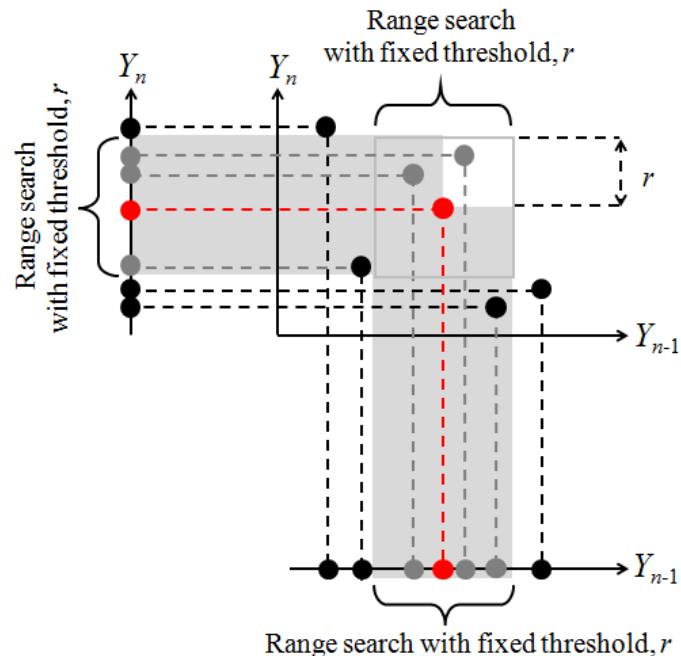


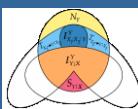
NONPARAMETRIC ESTIMATION: KERNEL METHOD

- **System Information of Y :** $\hat{H}_Y = \hat{H}(Y_n) = -\log \langle p(y_n) \rangle$
- **Conditional Entropy of Y_n given $Y_n^- \cong Y_n^L$:** $H(Y_n | Y_n^-) \cong \hat{H}(Y_n | Y_n^L) = -\log \left\langle \frac{p(y_n, y_n^L)}{p(y_n)} \right\rangle$
 $H(Y_n | Y_n^-) \cong \hat{H}(Y_n | Y_n^L) = \hat{H}(Y_n, Y_n^L) - \hat{H}(Y_n^L)$
- **Information Storage of Y :** $\hat{S}_Y = \log \left\langle \frac{p(y_n, y_n^L)}{p(y_n)p(y_n^L)} \right\rangle$
 $S_Y \cong \hat{H}(Y_n) - \hat{H}(Y_n, Y_n^L) + \hat{H}(Y_n^L)$

- Example: $L=1$

$$Y_n^L \cong Y_n^1 = Y_{n-1}$$





NONPARAMETRIC ESTIMATION: NEAREST NEIGHBOR METHOD

- Nearest neighbor estimate of a continuous probability distribution

- Generic d -dimensional vector variable Z :

*Study the distance between an **observation** z and its **k -th neighbor***

Main result from [Kozachenko L, Leonenko N, Prob Inform Transm 1987]

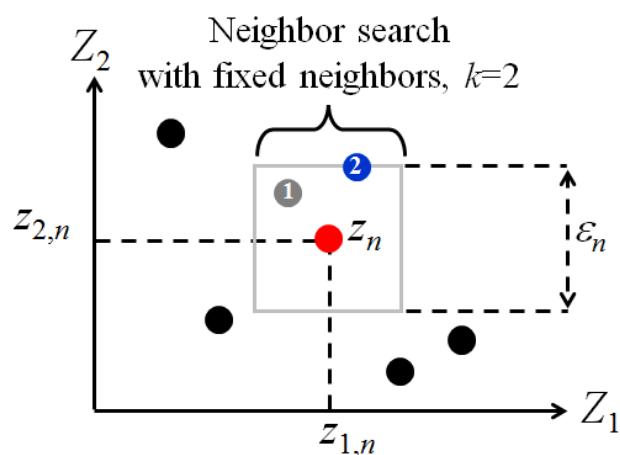
- **Entropy:** $H(Z) = -E[\log p(z)] \approx -\psi(k) + \psi(N) + d \langle \log \varepsilon_n \rangle$

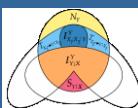
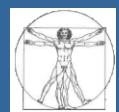
ψ : Digamma function $\psi(x) = \frac{d \log \Gamma(x)}{dx}$

ε : 2-distance from z to its k -th neighbor

N : number of outcomes of X

- Example: $d=2$





NONPARAMETRIC ESTIMATION: NEAREST NEIGHBOR METHOD

- ESTIMATION BIAS: *Conditional Entropy and MI estimates are biased at increasing the dimension*
- Solution: DISTANCE PROJECTION [Kraskov A et al, Phys Rev E 2004]

Compensation of the estimation bias in a sum of entropies by looking for neighbors in the higher dimensional spaces, and counting within ranges in the lower dimensional spaces

- Information Storage:**

$$S_Y = \hat{H}(Y_n) - \hat{H}(Y_n | Y_n^L) = \hat{H}(Y_n) - \hat{H}(Y_n, Y_n^L) + \hat{H}(Y_n^L)$$

$$\hat{H}(Y_n, Y_n^L) = -\psi(k) + \psi(N) + (L+1)\langle \log \varepsilon_n \rangle \rightarrow \text{Neighbor Search}$$

distance from (y_n, y_n^L) to its k -th neighbor in the outcomes of (Y_n, Y_n^L)

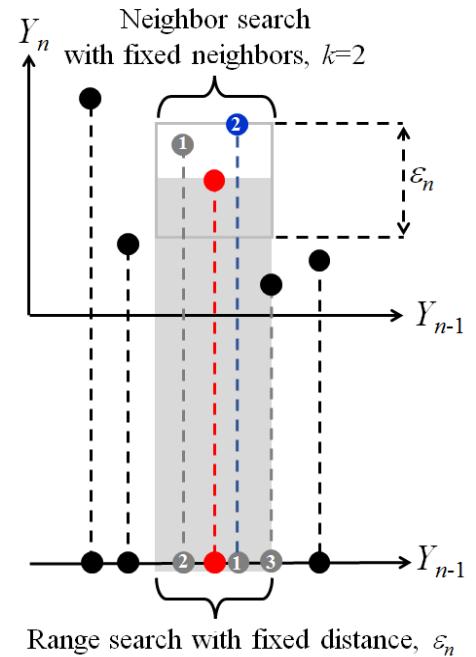
$$\hat{H}(Y_n) = -\langle \psi(N_{Y_n}) \rangle + \psi(N) + \langle \log \varepsilon_n \rangle \rightarrow \text{Range Search}$$

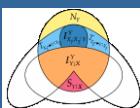
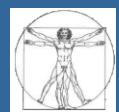
$$\hat{H}(Y_n^L) = -\langle \psi(N_{Y_n^L}) \rangle + \psi(N) + L\langle \log \varepsilon_n \rangle \rightarrow \text{Range Search}$$

number of outcomes of Y_n^L with distance to y_n^L strictly lower than $\varepsilon_n/2$

$$\hat{S}_Y = \psi(N) + \psi(k) - \langle \psi(N_{Y_n^L}) + \psi(N_{Y_n}) \rangle$$

- Example: $L=1$ $Y_n^L \cong Y_n^1 = Y_{n-1}$

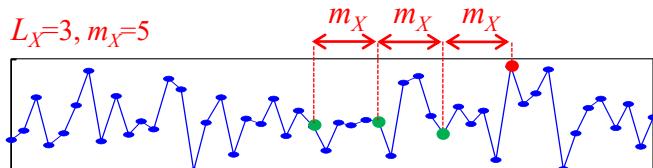




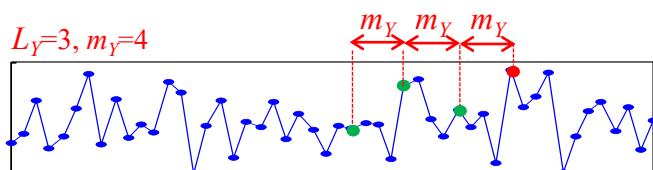
NONPARAMETRIC ESTIMATION: APPROXIMATION OF THE SYSTEM PAST

- Uniform embedding (UE):

$$X_n^- \approx [X_{n-m_X} \dots X_{n-L_X m_X}] \rightarrow X$$



$$Y_n^- \approx [Y_{n-m_Y} \dots Y_{n-L_Y m_Y}] \rightarrow Y$$



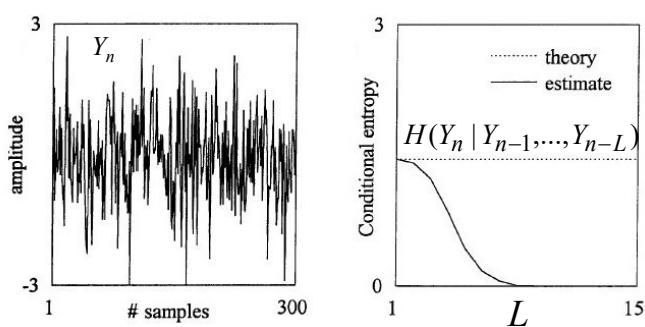
ISSUES:

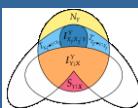
- Sub-optimality of separate reconstruction of X_n^- and Y_n^-
- Selection of the embedding parameters m , L
- **UE introduces irrelevant and redundant components** → **Curse of dimensionality**

Conditional Entropy estimates decrease towards zero at increasing the dimension

Example: white noise

[A Porta et al., Biol. Cyb. 1998]





NONPARAMETRIC ESTIMATION: NON-UNIFORM EMBEDDING

- **Non-uniform embedding (NUE):**

The embedding vector is formed progressively, including at each step the lagged variable better describing the target process

- ***Sequential procedure:***

- (a) **$k=0$: Initialization**

Set of initial candidate components (e.g., $\Omega = \{X_{n-1}, \dots, X_{n-L}, Y_{n-1}, \dots, Y_{n-L}\}$)

Initial embedding vector: $V_n^{(0)} = [\cdot]$

- (b) **$k \geq 1$: Selection – maximum relevance, minimum redundancy**

Select the component $W_n \in \Omega$ that maximizes $I(Y_n, W_n | V_n^{(k-1)}) \rightarrow V_n^{(k)} = [\hat{W}_n, V_n^{(k-1)}]$

- (c) **Termination** when \hat{W}_n does not add significant information to Y_n

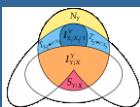
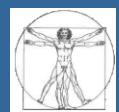
Generate N surrogates of \hat{W}_n by sample shuffling: $\hat{W}_n^{(S_1)}, \dots, \hat{W}_n^{(S_N)}$

Threshold for $I(Y_n, \hat{W}_n | V_n^{(k-1)}) : I_{th}$

Stop if $I(Y_n, \hat{W}_n | V_n^{(k-1)}) < I_{th}$; final set of components: $V_n = V_n^{(k-1)}$

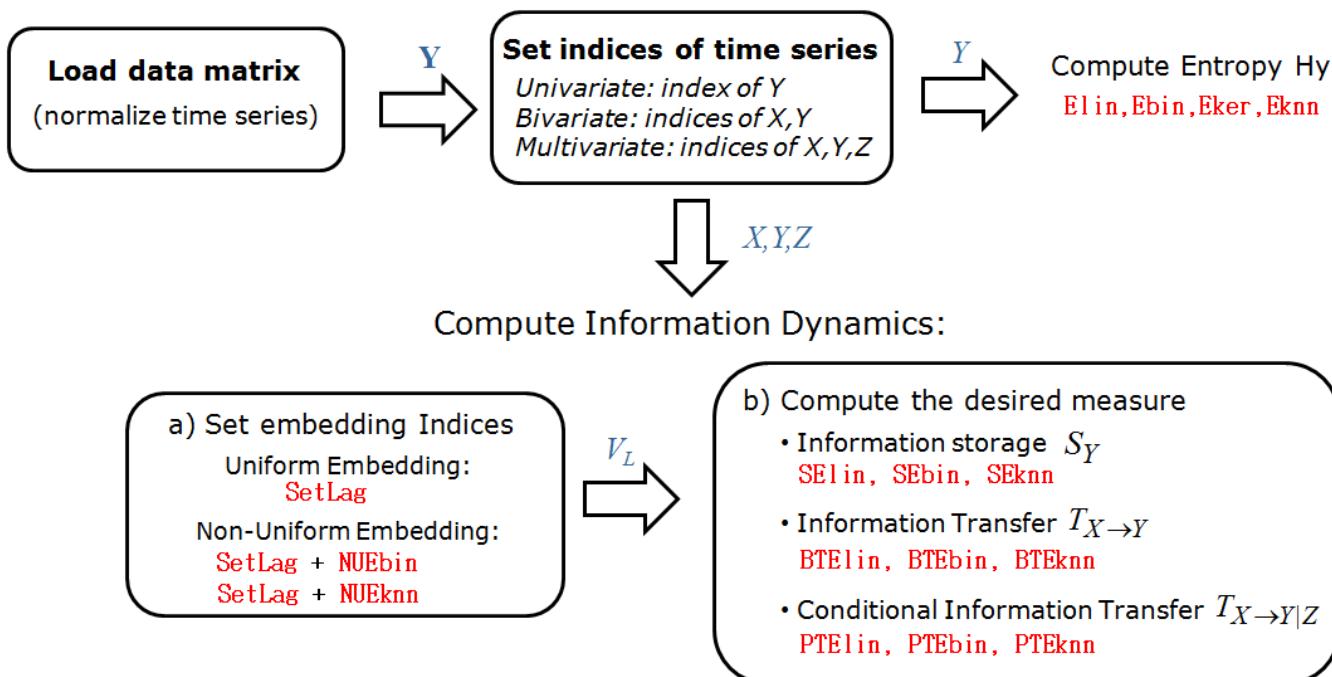
- (d) **After termination – embedding vector** $V_n = [V_n^X, V_n^Y, V_n^Z] \rightarrow X_n^- \approx V_n^X \quad Y_n^- \approx V_n^Y \quad Z_n^- \approx V_n^Z$

Statistical significance results directly from the components acceptance based on randomization



ITS Toolbox:

A Matlab toolbox for the practical computation of Information Dynamics



<http://www.lucafaes.net/its.html>

luca.faes@unipa.it