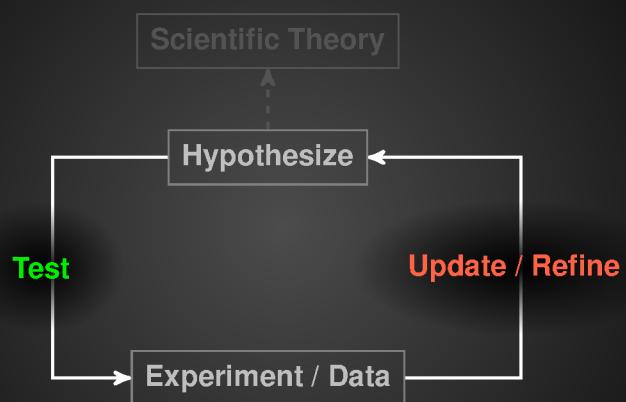




## The Scientific Method

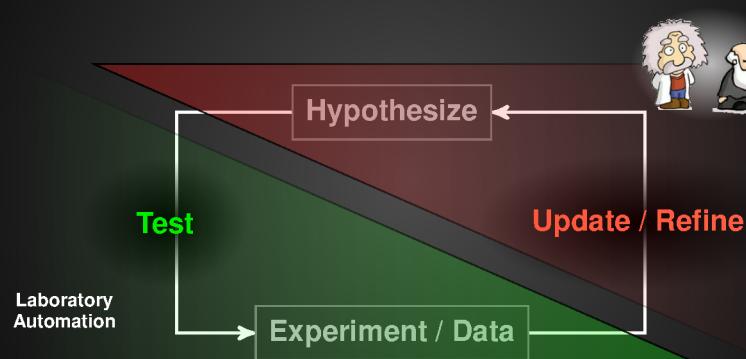
A la Francis Bacon





## The Scientific Method

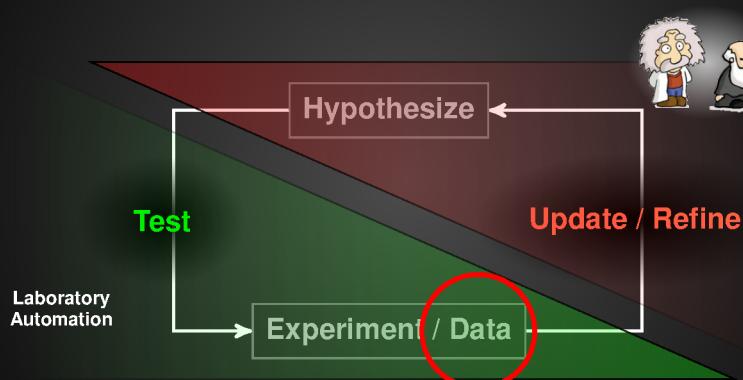
A la Francis Bacon





## The Scientific Method

A la Francis Bacon



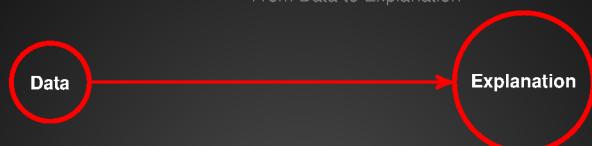
... computational approaches have been more successful in **small, well-defined systems** than in larger, less studied, or more complex ones. The explosion of data from high-throughput experiments, however, increasingly presents researchers with very complicated systems.

— *Machine Science*, James Evans and Andrey Rzhetsky, *Science*, 329 (5990), 399-400 (2010)



## Automating Science

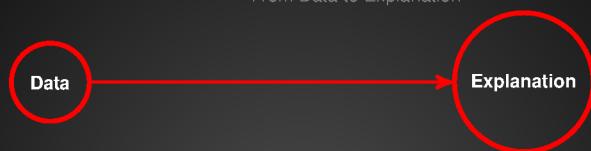
From Data to Explanation





## Automating Science

From Data to Explanation



- Universal Distance Metric
- Features



## Automating Science

From Data to Explanation



- Universal Distance Metric
- Features



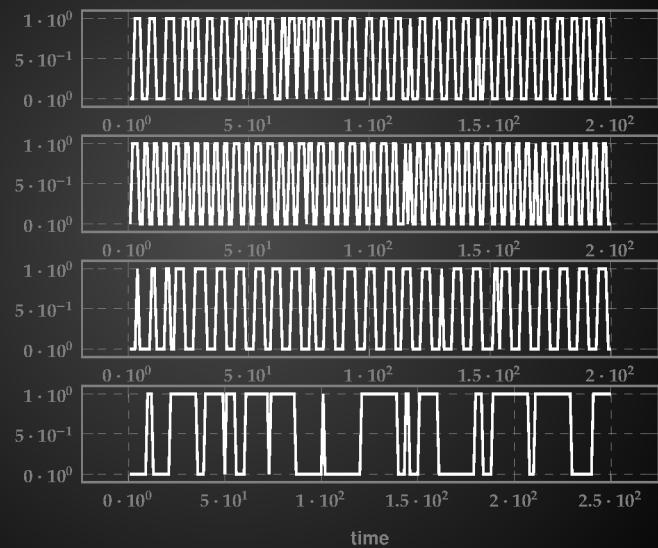
not similar





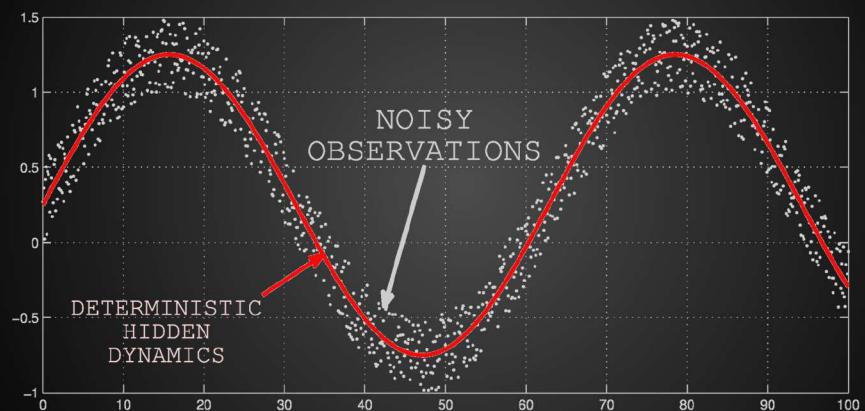
## Similarity

Pick the  
odd one  
out!



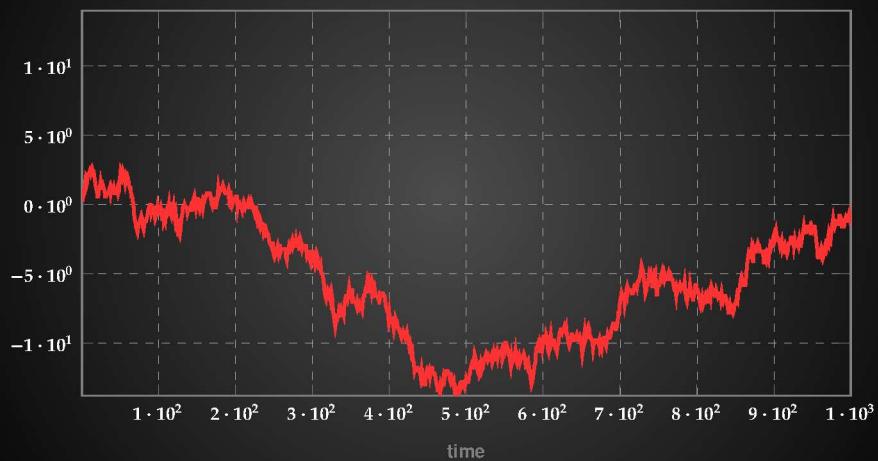


Stochastic Process  $\neq$  Deterministic + Noise



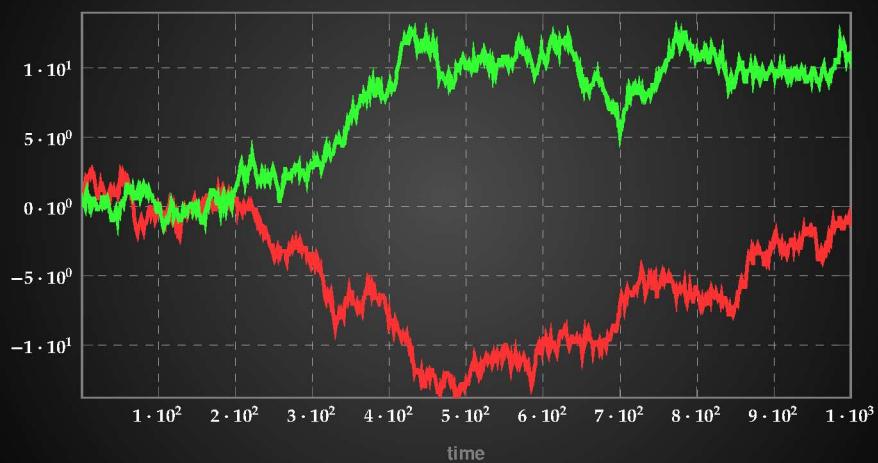


Stochastic Process  $\neq$  Deterministic + Noise





Stochastic Process  $\neq$  Deterministic + Noise





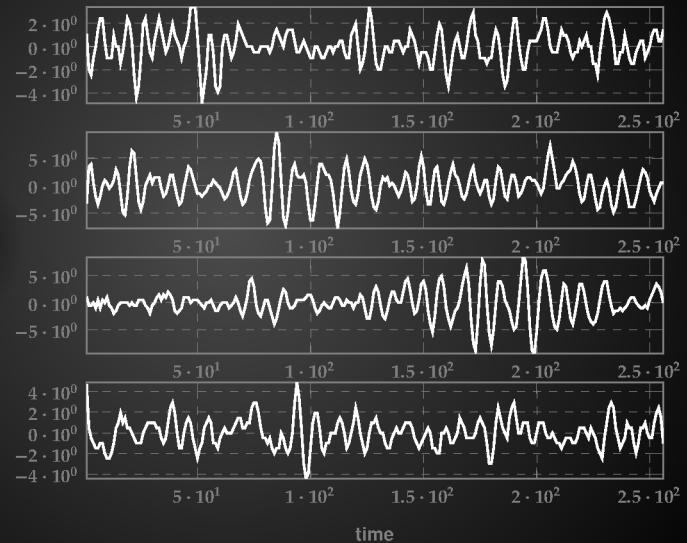
Stochastic Process  $\neq$  Deterministic + Noise





## Brainwaves: Identical Stimuli

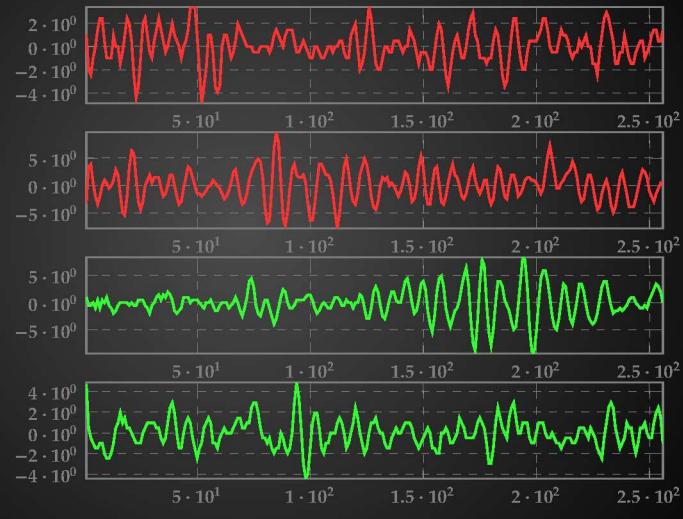
Are they  
from the  
same individual?





## Brainwaves: Identical Stimuli

Subject A



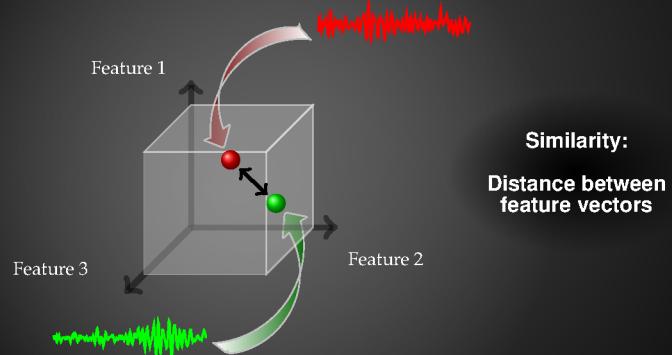
Subject B

time



## State of Art

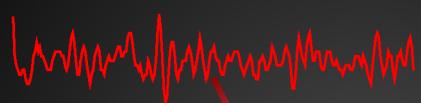
Requires Features





## Data Smashing

Signal 1



**Quantize:** bcbbbbacabbcbbbbabbba...

Signal 2

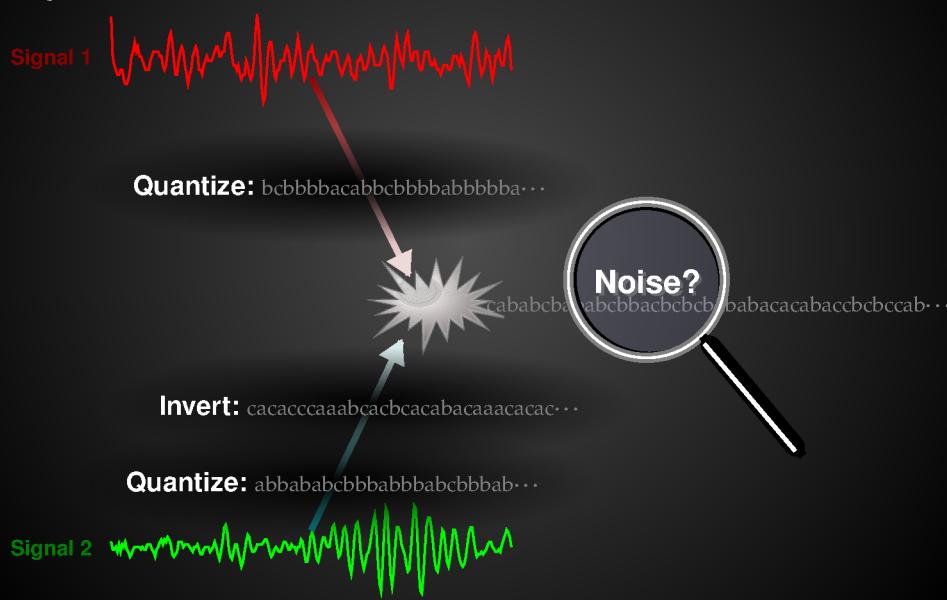
**Invert:** cacacccaaabcacbcacabacaacacac...

**Quantize:** abbababcbbaabbabcbba...

cababcbababcbbacbcbcbababacacabaccbcbccab...



## Data Smashing





## Anti-streams

---

Intuitive Description

Stream  $s \rightarrow$  Anti-stream  $s'$

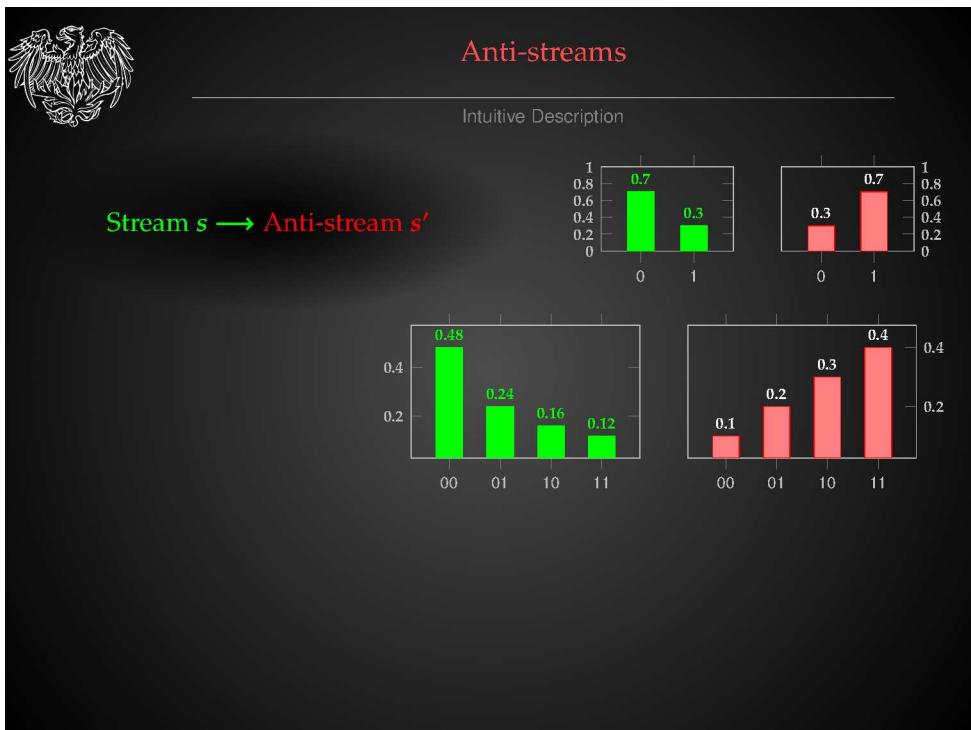


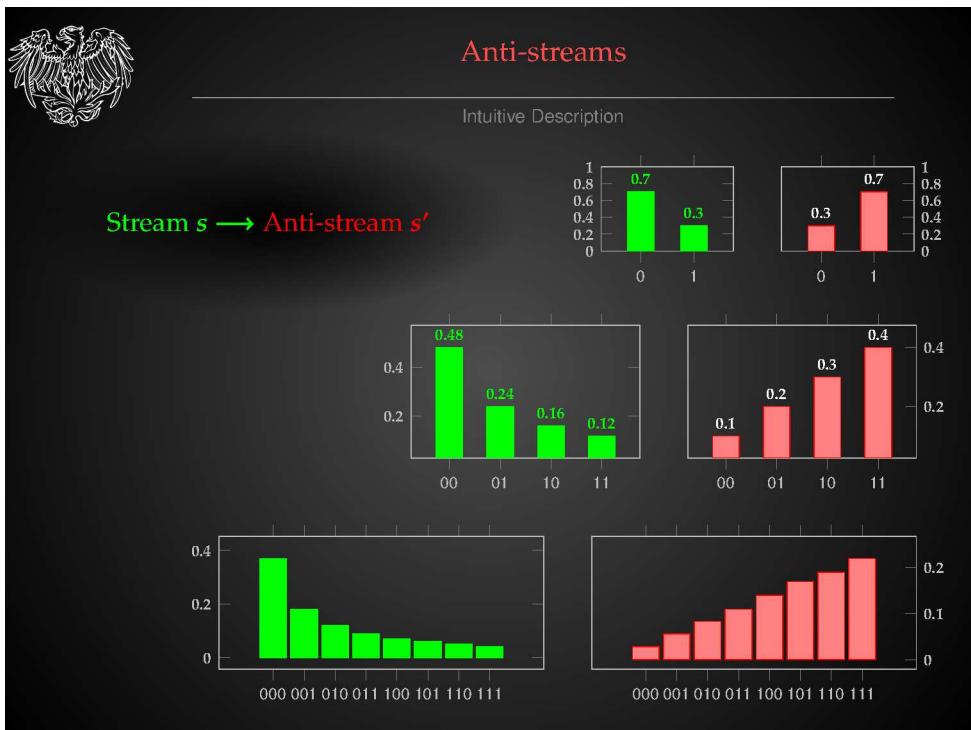
## Anti-streams

Intuitive Description

Stream  $s \rightarrow$  Anti-stream  $s'$



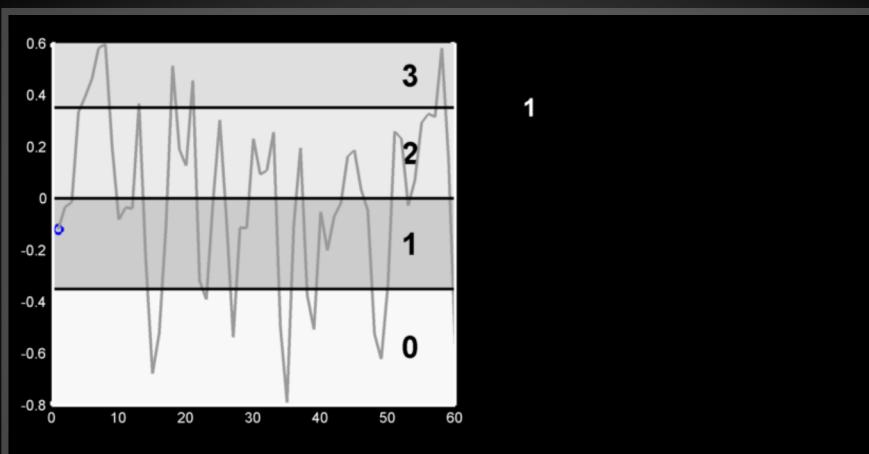






## Quantization

Mapping Continuous Data to Symbol Stream

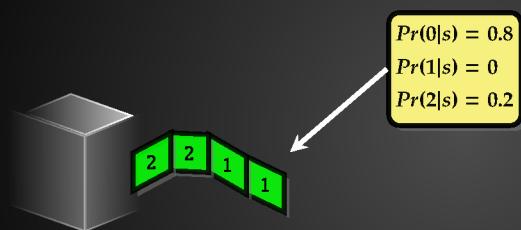


Quantization Alphabet  $\Sigma = \{0, 1, 2, 3\}$



## The Black Box Approach

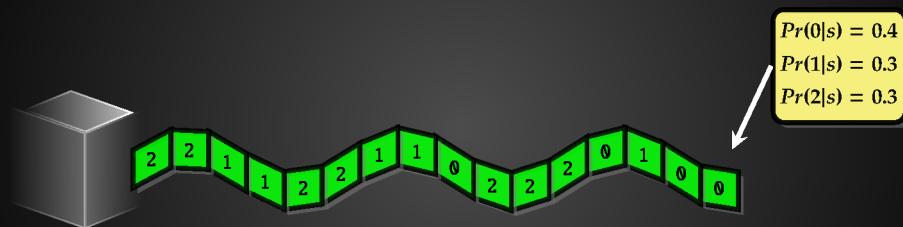
Dynamical System As A Symbol Generator





## The Black Box Approach

Dynamical System As A Symbol Generator



Ergodic Stationary Quantized Process  $\Longleftrightarrow$  Probability Measure on Infinite Strings



## System As A Symbol Generator

### Probability Space $(\Sigma^\omega, \mathcal{B}, \mu)$

- $\Sigma^\omega$  : Set of strictly infinite strings on alphabet  $\Sigma$
- $\mathcal{B}$  : smallest  $\sigma$ -algebra generated by the sets  $\{x\Sigma^\omega : x \in \Sigma^*$
- $\mu$  : Probability measure on infinite strings:

$$\begin{aligned}\mu(x\Sigma^\omega) &\mapsto [0, 1] \\ \sum_{x \in \Sigma^*} \mu(x\Sigma^\omega) &= 1\end{aligned}$$



## System As A Symbol Generator

### Probability Space $(\Sigma^\omega, \mathcal{B}, \mu)$

- $\Sigma^\omega$  : Set of strictly infinite strings on alphabet  $\Sigma$
- $\mathcal{B}$  : smallest  $\sigma$ -algebra generated by the sets  $\{x\Sigma^\omega : x \in \Sigma^*$
- $\mu$  : Probability measure on infinite strings:

$$\begin{aligned}\mu(x\Sigma^\omega) &\mapsto [0, 1] \\ \sum_{x \in \Sigma^*} \mu(x\Sigma^\omega) &= 1\end{aligned}$$

Probability Measure on Infinite Strings  $\implies$  Equivalence Relation on Finite Strings

$$\begin{aligned}\forall x_1, x_2 \in \Sigma^*, x_1 &\sim x_2 \\ \text{if } \forall x \in \Sigma^*, \mu(x_1 x \Sigma^\omega) &= \mu(x_2 x \Sigma^\omega)\end{aligned}$$

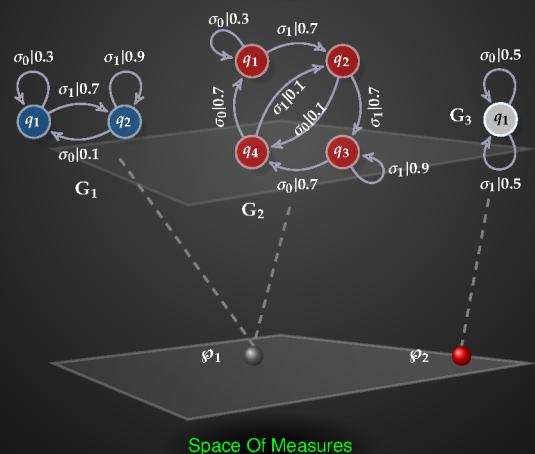
Equivalence Classes are causal states



## Probabilistic Finite State Automata

Models For Quantized Stationary Ergodic Stochastic Processes

Space Of PFSA



Space Of Measures



## Adding Probability Measures

Consider two measures:  $\begin{cases} \varphi_1 : \mathcal{B} \rightarrow [0, 1] \\ \varphi_2 : \mathcal{B} \rightarrow [0, 1] \end{cases}$

Define a binary operation:

$$\varphi_1 \oplus \varphi_2 \triangleq \varphi_3$$

where

$$\begin{aligned}\varphi_3(x\Sigma^\omega) &= \varphi_1(x\Sigma^\omega)\varphi_2(x\Sigma^\omega) \times \text{Constant} \\ \sum_{x \in \Sigma^*} \varphi_3(x\Sigma^\omega) &= 1\end{aligned}$$



## Adding Probability Measures

Consider two measures:  $\begin{cases} \varphi_1 : \mathcal{B} \rightarrow [0, 1] \\ \varphi_2 : \mathcal{B} \rightarrow [0, 1] \end{cases}$

Define a binary operation:

$$\varphi_1 \oplus \varphi_2 \triangleq \varphi_3$$

where

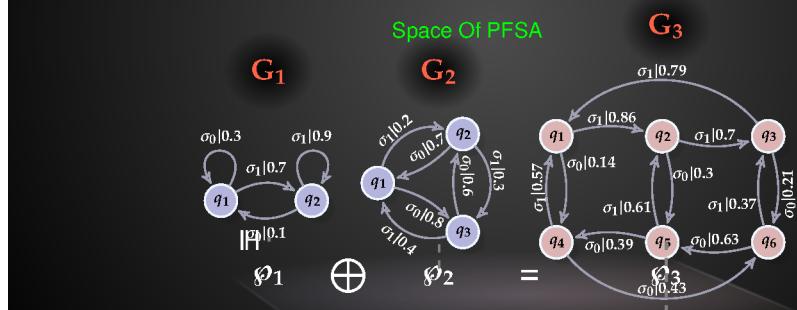
$$\begin{aligned} \varphi_3(x\Sigma^\omega) &= \varphi_1(x\Sigma^\omega)\varphi_2(x\Sigma^\omega) \times \text{Constant} \\ \sum_{x \in \Sigma^*} \varphi_3(x\Sigma^\omega) &= 1 \end{aligned}$$

- ⊕ Commutative
- ⊕ Closed
- ⊕ Unique Inverse
- ⊕ Unique Identity

Abelian Group



## Lifting Group Structure To PFSAs





## Mathematical Structure Of Model Space

---

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group



## Mathematical Structure Of Model Space

---

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group

$$1 + 2 = 3$$

$$2 - 2 = 0$$

$$3 + 0 = 3$$



## Mathematical Structure Of Model Space

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group

$$1 + 2 = 3$$

$$2 - 2 = 0$$

$$3 + 0 = 3$$

We can "add and subtract" models:

$$G + H = J$$

$$G - H = K$$



## Mathematical Structure Of Model Space

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group

$$1 + 2 = 3$$

$$2 - 2 = 0$$

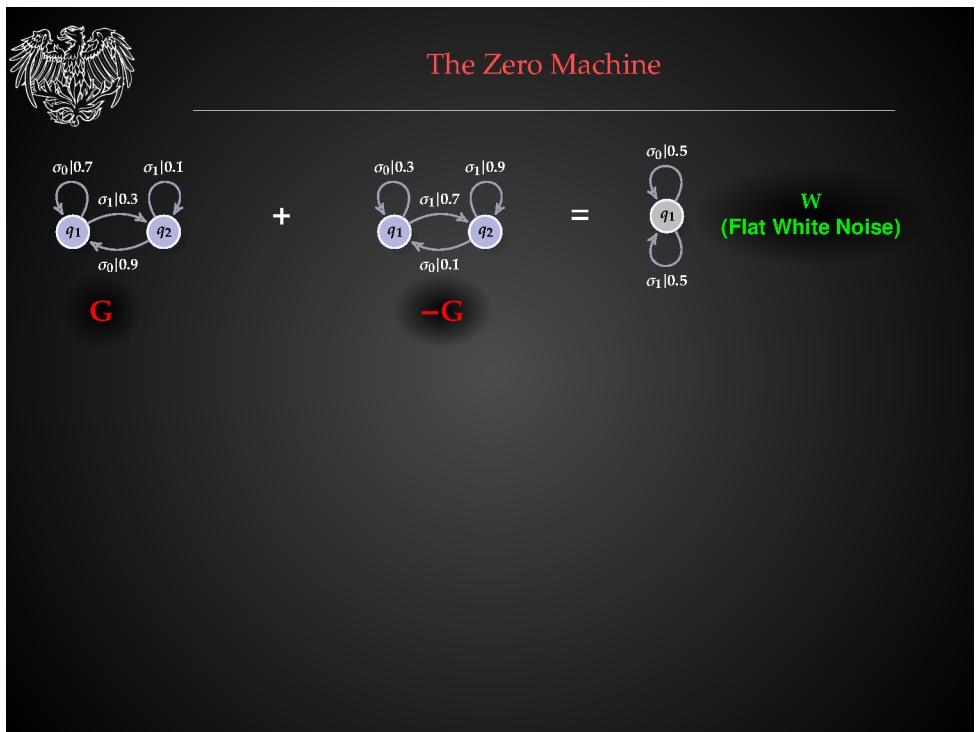
$$3 + 0 = 3$$

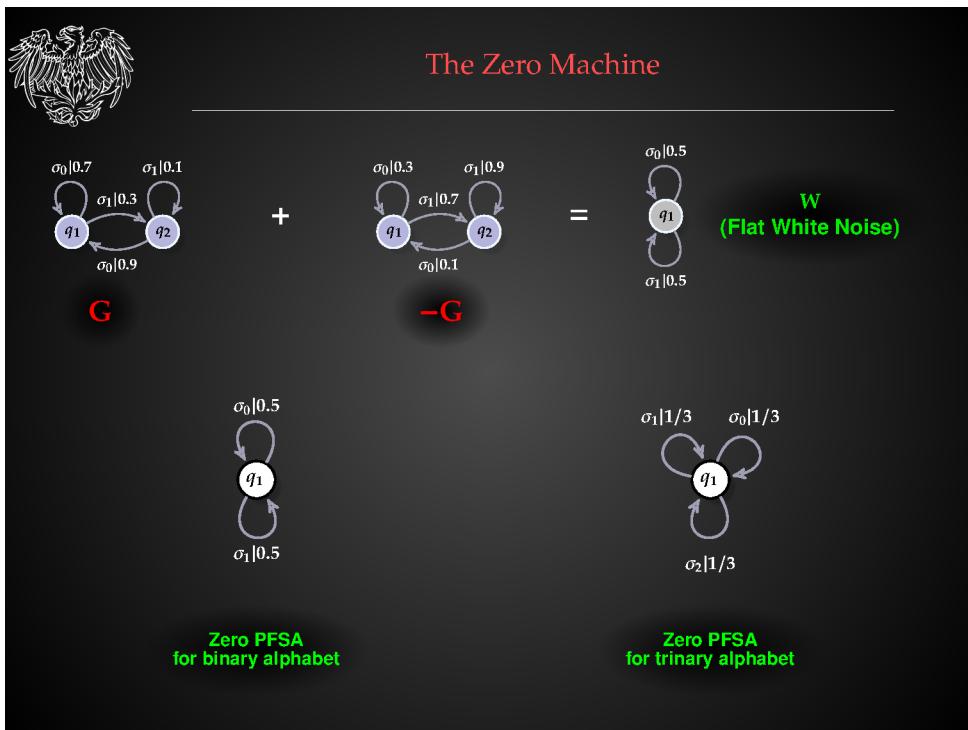
We can "add and subtract" models:

$$G + H = J$$

$$G - H = K$$

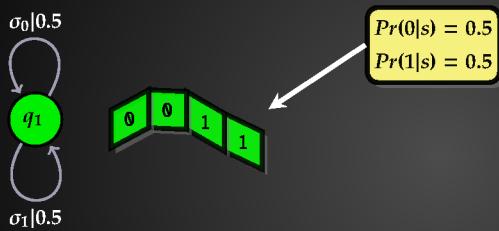
$$G - G = ?$$







## The Zero Machine



Zero PFSA  
binary alphabet

- Maximum entropy rate
- History is useless for prediction
- Encodes minimum information

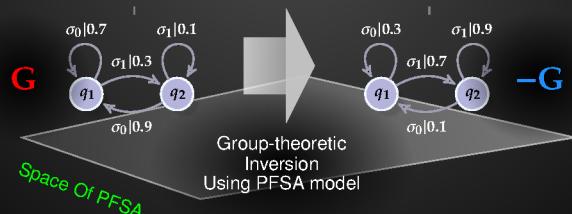


## Stream Inversion

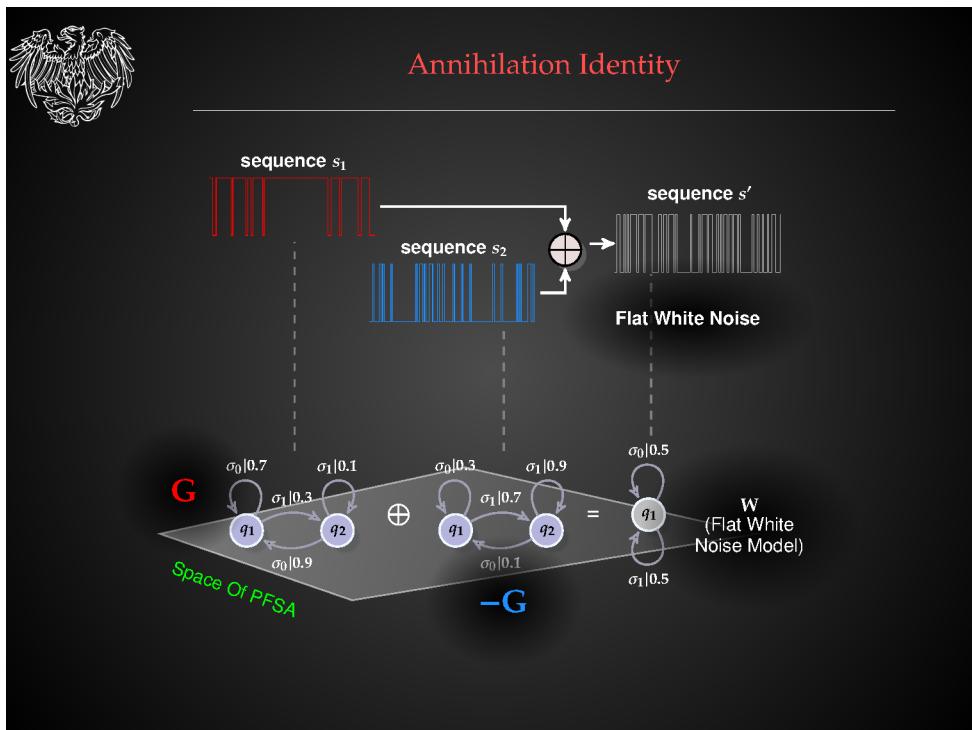
Direct Generation of Anti-streams



Stream Inversion  
via selective erasure of  $s_1$



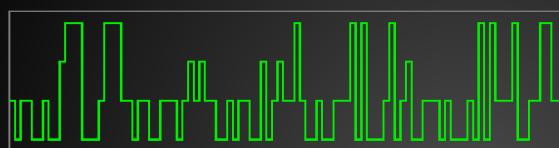
Space Of PFSA



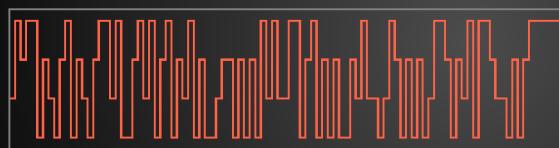


## Example Of Stream, Anti-stream, & FWN

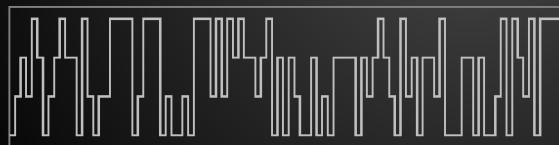
4 Letter Alphabet



**Signal**



**Inverse Signal**

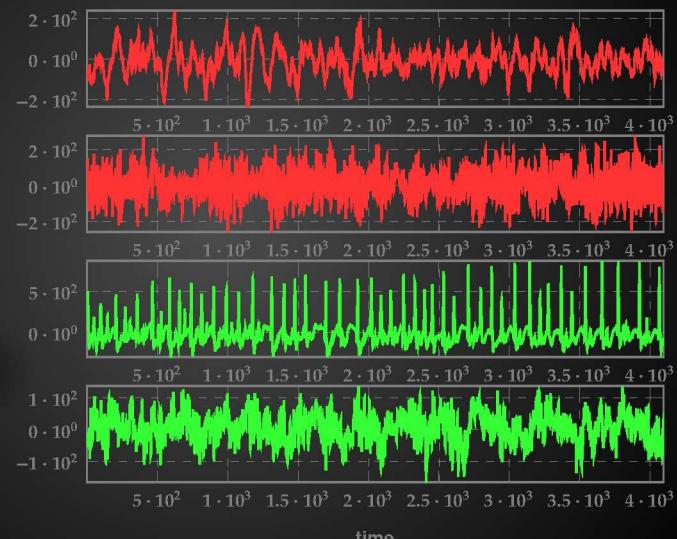


**Flat White Noise**

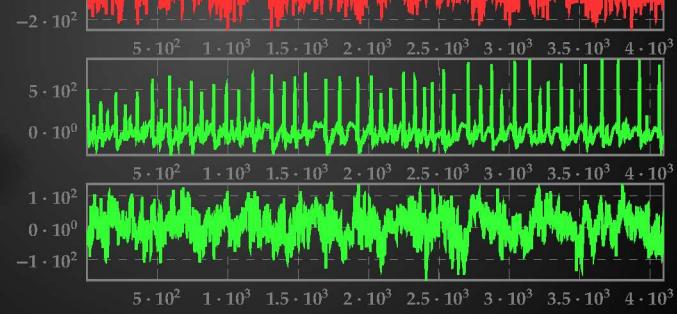


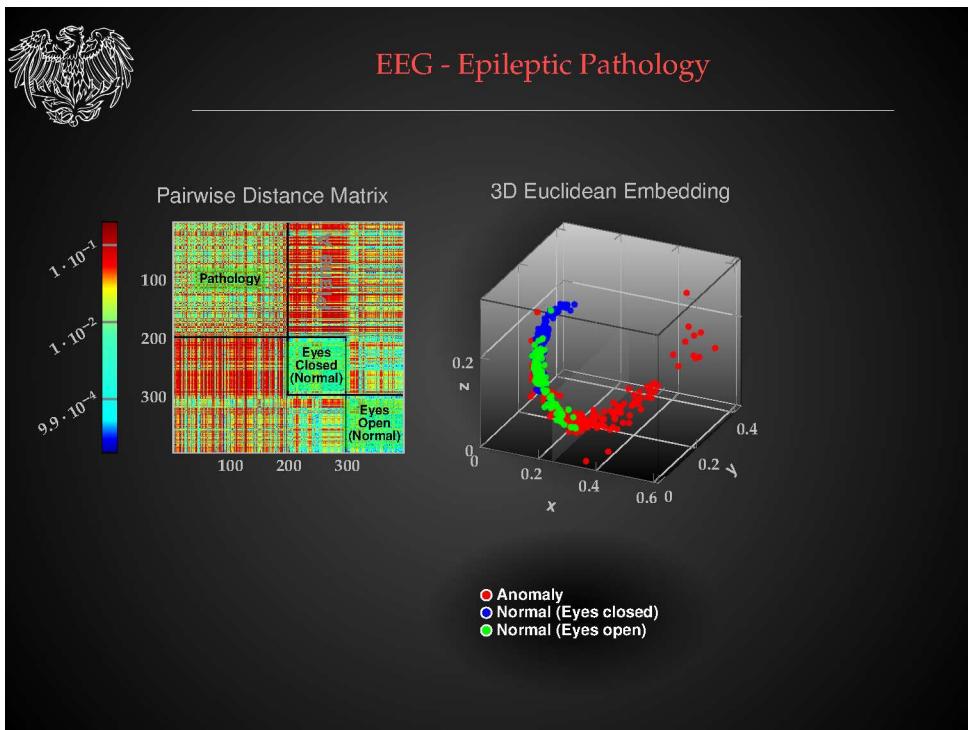
## EEG - Epileptic Pathology

Eyes Open



Epileptic Pathology



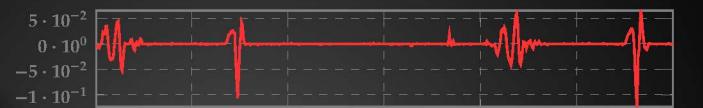




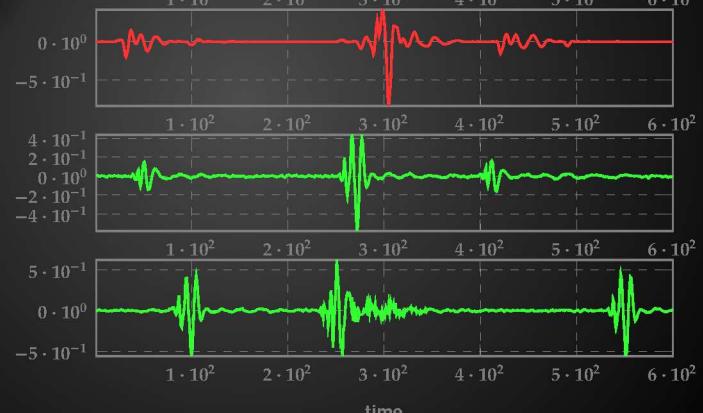
## Cardiac Pathology

Disambiguate Normal Rhythm from Murmur

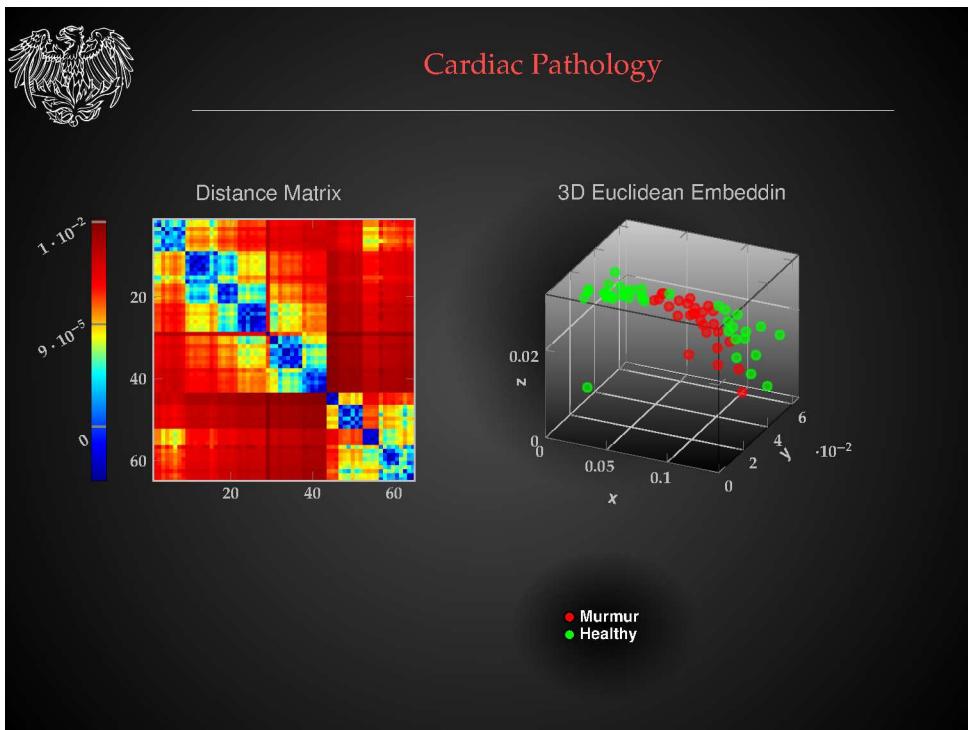
Normal Rhythm

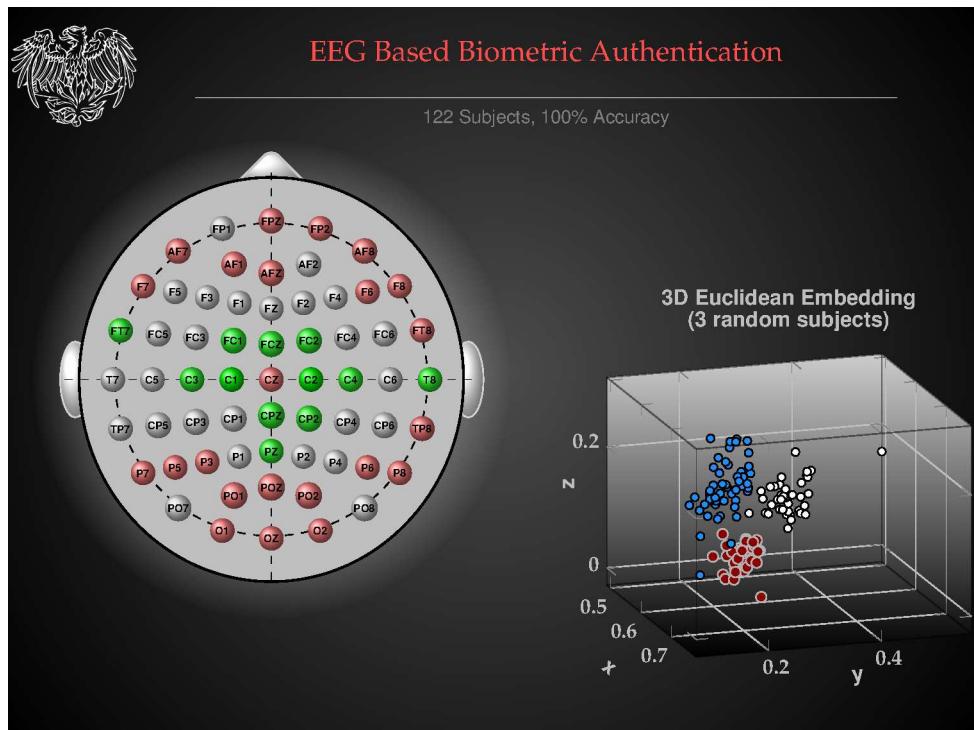


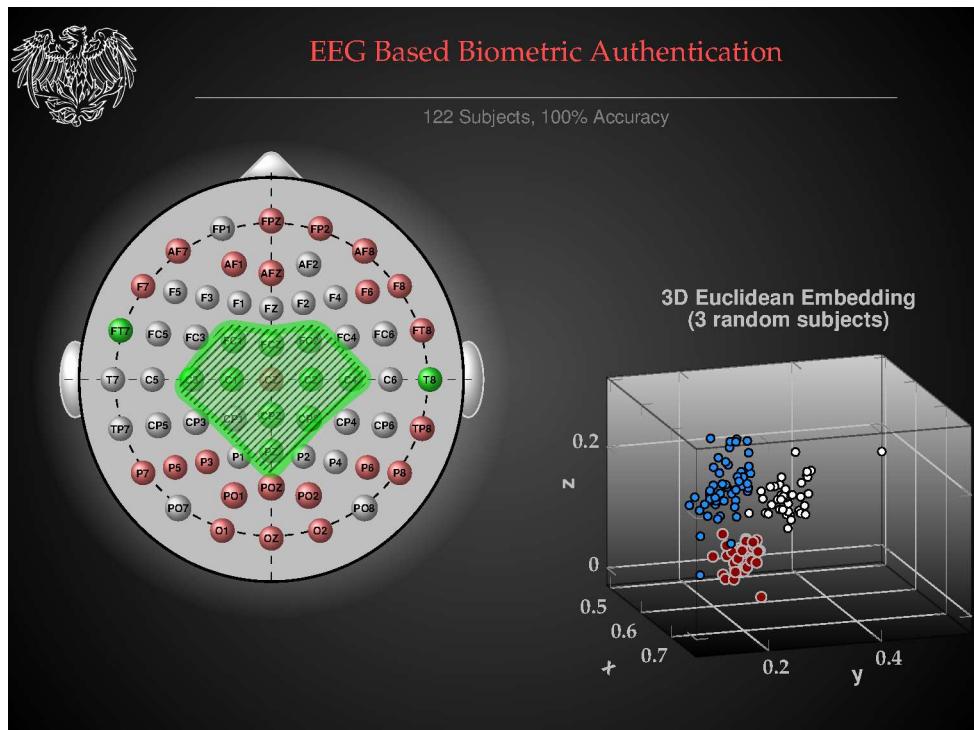
Murmur

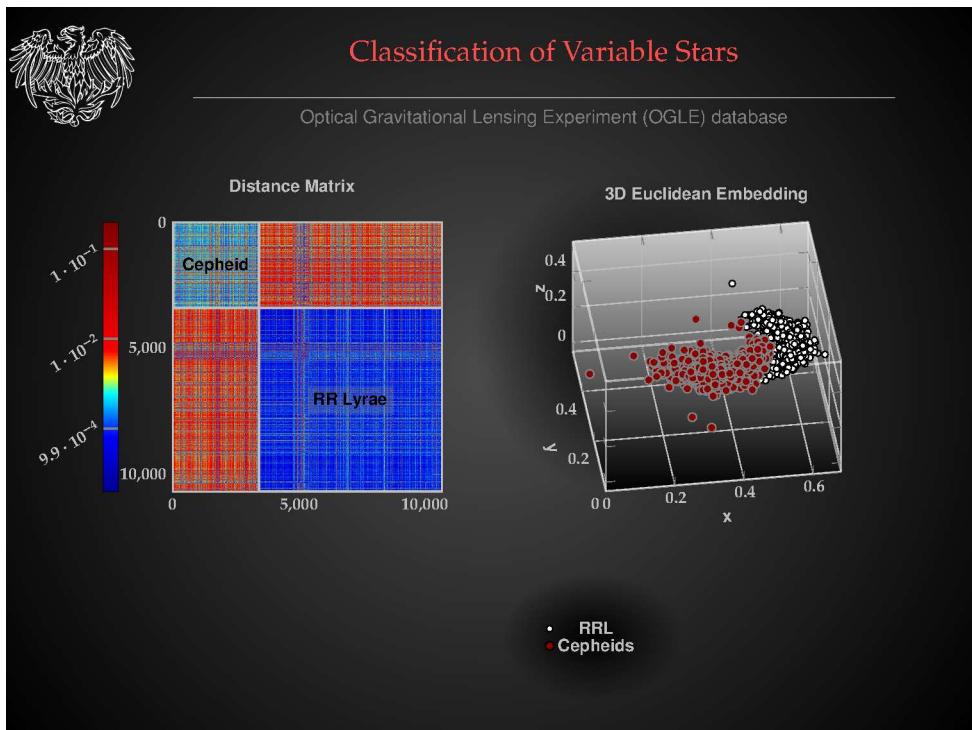


time



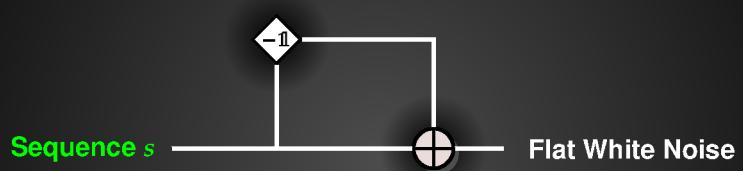






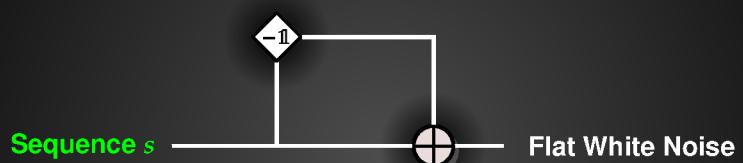


## Self Annihilation Error





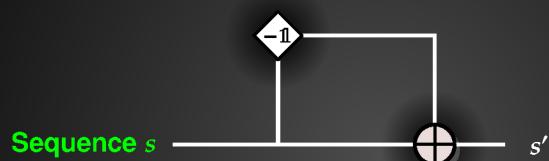
## Self Annihilation Error



Only if  $|s|$  is large enough !



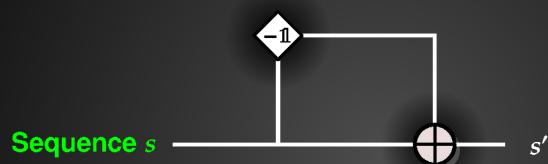
## Self Annihilation Error



$$\epsilon_s = \Theta(s', W)$$



## Self Annihilation Error

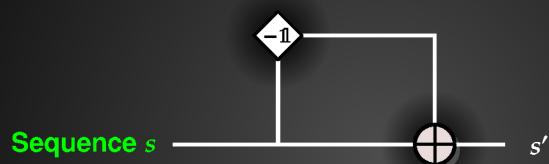


$$\epsilon_s = \Theta(s', W)$$

- $\epsilon_s \rightarrow 0$  exponentially fast with  $|s|$
- Information content of stream

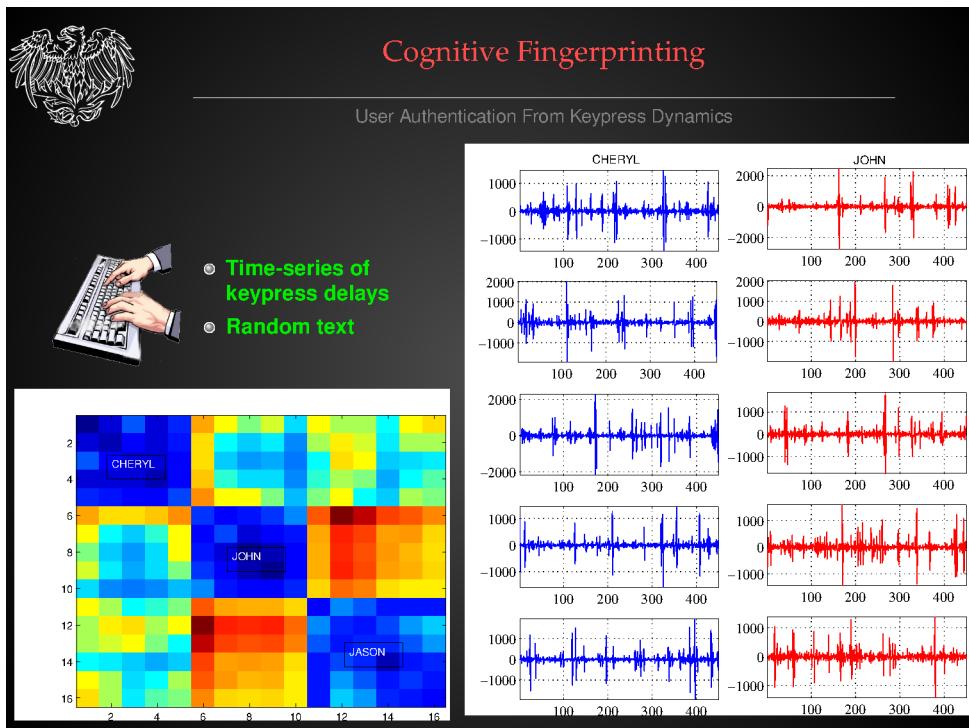


## Self Annihilation Error



$$\epsilon_s = \Theta(s', W)$$

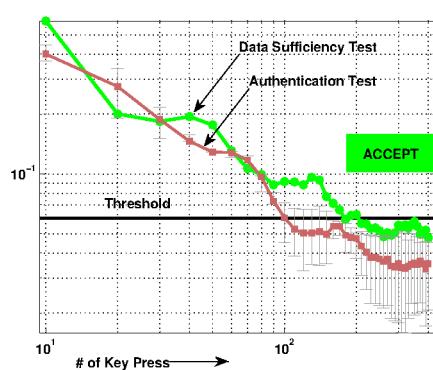
Auto-detect Data Sufficiency!



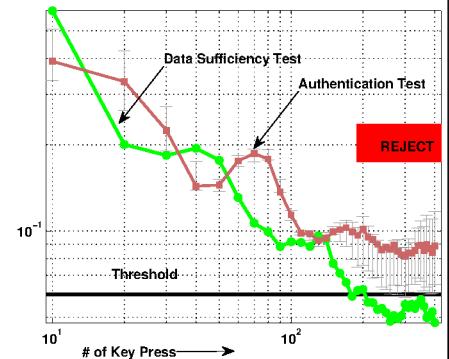


## Cognitive Fingerprinting

User Authentication From Keypress Dynamics



User Authorized

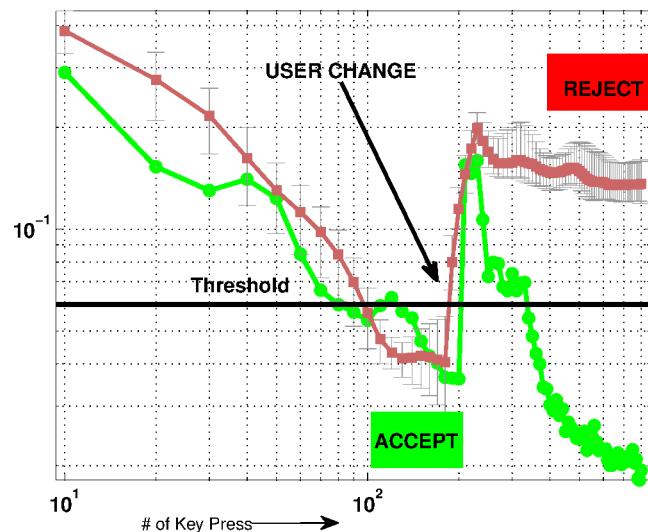


User Rejected



## Cognitive Fingerprinting

User Authentication From Keypress Dynamics

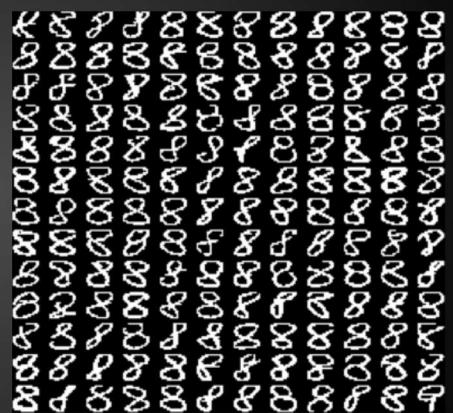
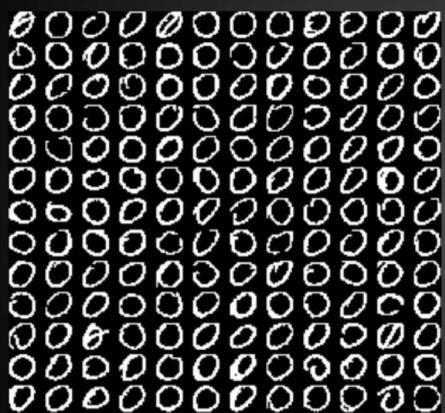


Authorization  
revoked  
on unexpected user  
change detection



## Handwritten Digits

Recognition / Classification Problem



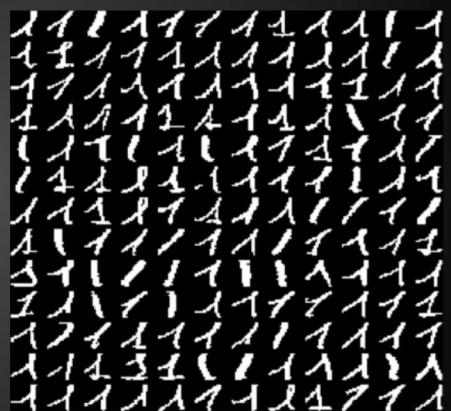
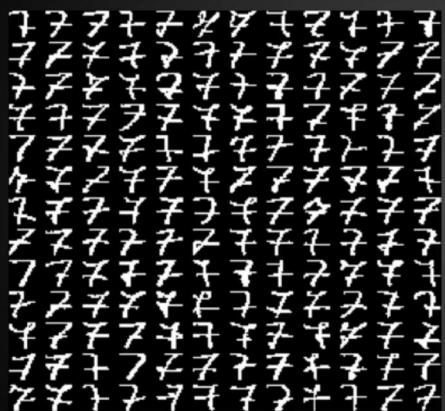


# Handwritten Digits



## Handwritten Digits

Recognition / Classification Problem



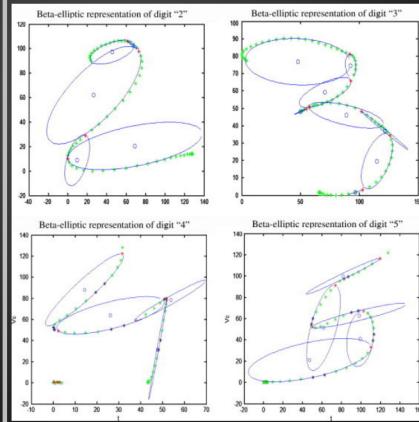
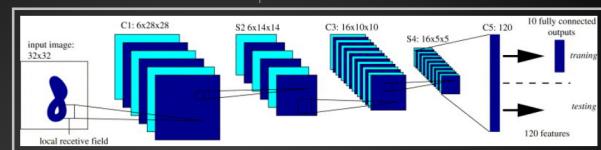


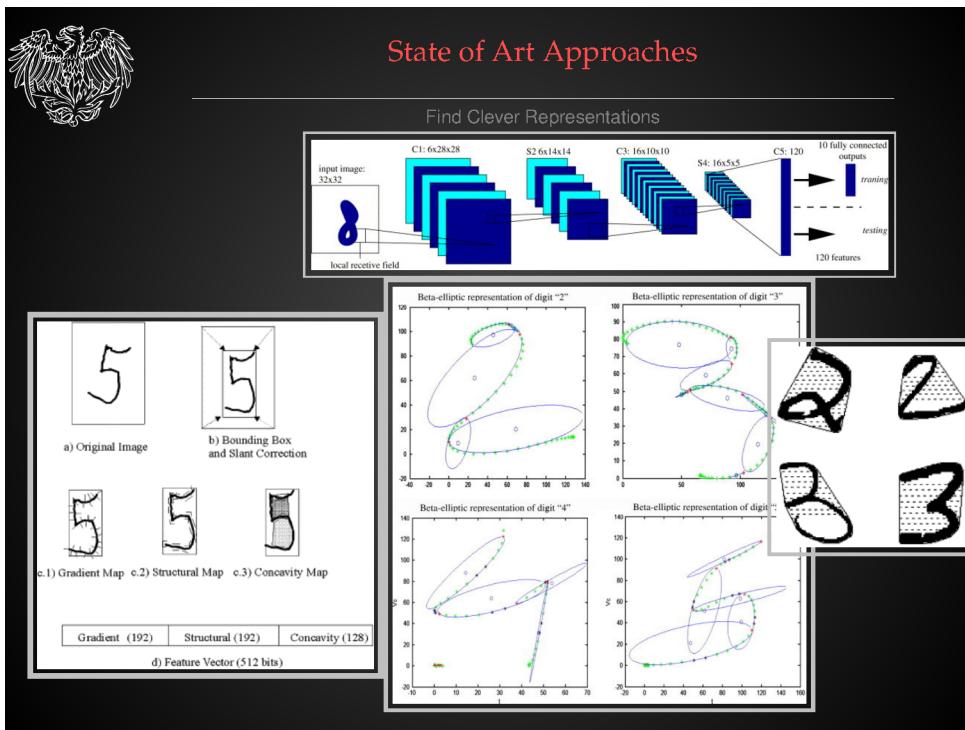
## Handwritten Digits

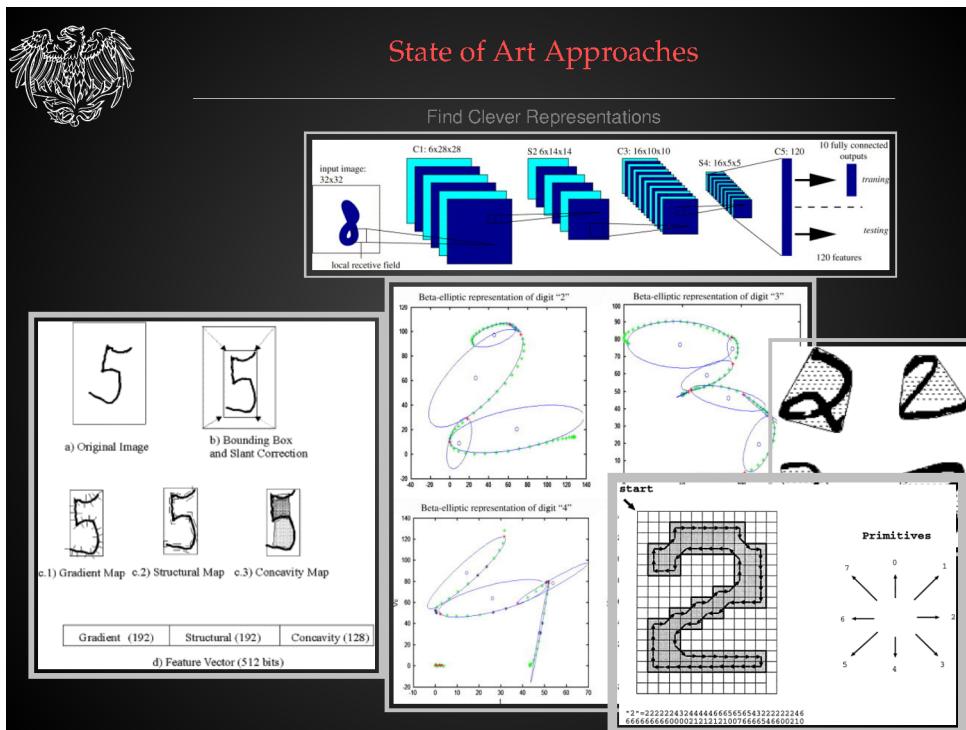


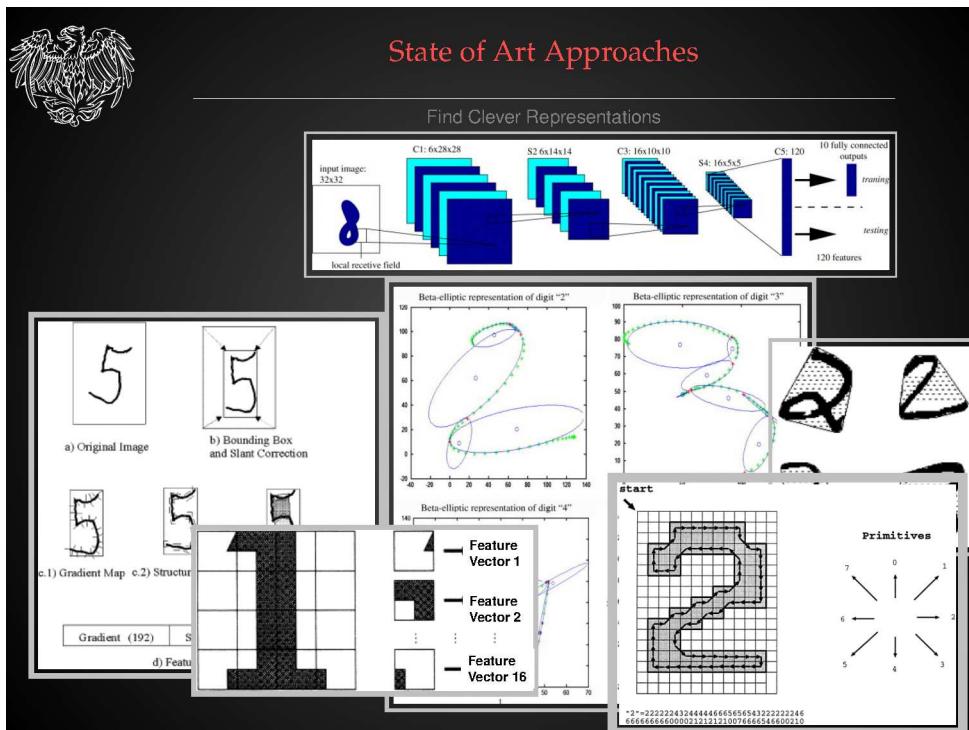
## State of Art Approaches

Find Clever Representations





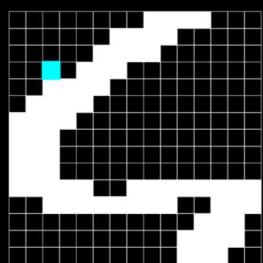




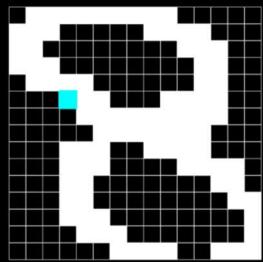


## Image To Symbol Stream

Random Walk



0



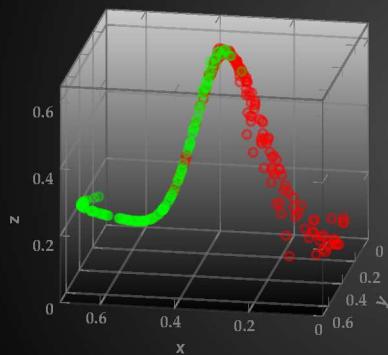
0



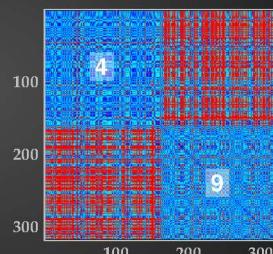
## Handwritten Digits

Recognition / Classification Problem

3D Euclidean Embedding



Pairwise Distance Matrix





## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	<input type="text"/>
0	<input type="text"/>

Copy 1:



## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	1
0	1

Copy 1: 1



## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	1	2
0	1	2

Copy 1: 1 2



## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	1	2	0
0	1	2	2

Copy 1: 1 2



## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	1	2	0	1
0	1	2	2	2

Copy 1: 1 2



## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	1	2	0	1	0	2	2	0	0	0	2	1	2	0	1	1	0	0	1	0	0	2	0	1	0	2
0	1	2	2	2	1	2	0	0	0	0	2	1	0	2	1	0	0	0	2	2	2	0	2	1	0	2

**Copy 1:** 1 2 2 0 0 0 1 0 0 1 0 2 1 0 2 2 2 1 2 2 1 2 2 1 2 1

Generating Second Independent Copy for String  $s_1$ :

2	
0	

**Copy 2:**



## Memory-less Stream Manipulation

Generating First Independent Copy for String  $s_1$ :

2	1	2	0	1	0	2	2	0	0	0	2	1	2	0	1	1	0	0	1	0	0	2	0	1	0	2
0	1	2	2	2	1	2	0	0	0	0	2	1	0	2	1	0	0	0	2	2	2	0	2	1	0	2

Copy 1: 1 2 2 0 0 0 1 0 0 1 0 2 1 0 2 2 2 1 2 2 1 2 2 1 2 1

Generating Second Independent Copy for String  $s_1$ :

2	1	2	0	1	0	2	2	0	0	2	1	2	0	1	1	0	0	1	0	0	2	0	1	0	2
0	0	2	0	0	2	0	2	2	2	2	1	2	1	0	1	0	0	0	2	0	2	2	0	2	0

Copy 2: 2 0 2 2 1 0 0 2 2 2 1 2 2 0 1 2 2 1 2 1 2



## Memory-less Stream Manipulation

Generating Inverse of String  $s_1$ :

1	
2	
0	

Inverse: 0



## Memory-less Stream Manipulation

Generating Inverse of String  $s_1$ :

1	2
2	0
0	1

Inverse: 0 1



## Memory-less Stream Manipulation

Generating Inverse of String  $s_1$ :

1	2	2
2	0	2
0	1	

Inverse: 0 1



## Memory-less Stream Manipulation

Generating Inverse of String  $s_1$ :

1	2	2	0
2	0	2	2
0	1		1

Inverse: 0 1 1



## Memory-less Stream Manipulation

Generating Inverse of String  $s_1$ :

1	2	2	0	0	0	1	0	0	1	0	2	1	0	2	2	2	1	2	2	1	2	1	2	1
2	0	2	2	1	0	0	2	2	2	1	2	2	0	1	2	2	1	2	1	2	1	2	1	2
0	1	1	2	2	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Inverse: 0 1 1 2 2 1 1 0 2 0 0 0 0

Summing  $s_2$  with Inverse of  $s_1$ :

2	2	1	1	2	2	1	1	0	2	2	2	0	1	0	0	2	1	2	2	1	2	2	1	
0	1	1	2	2	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Result: 1 2 1 0