Neural Networks Supporting Persistent Percepts

Shaul Druckmann and Mitya Chklovskii
Janelia Farm Research Campus

NIPS 2010

Working Memory

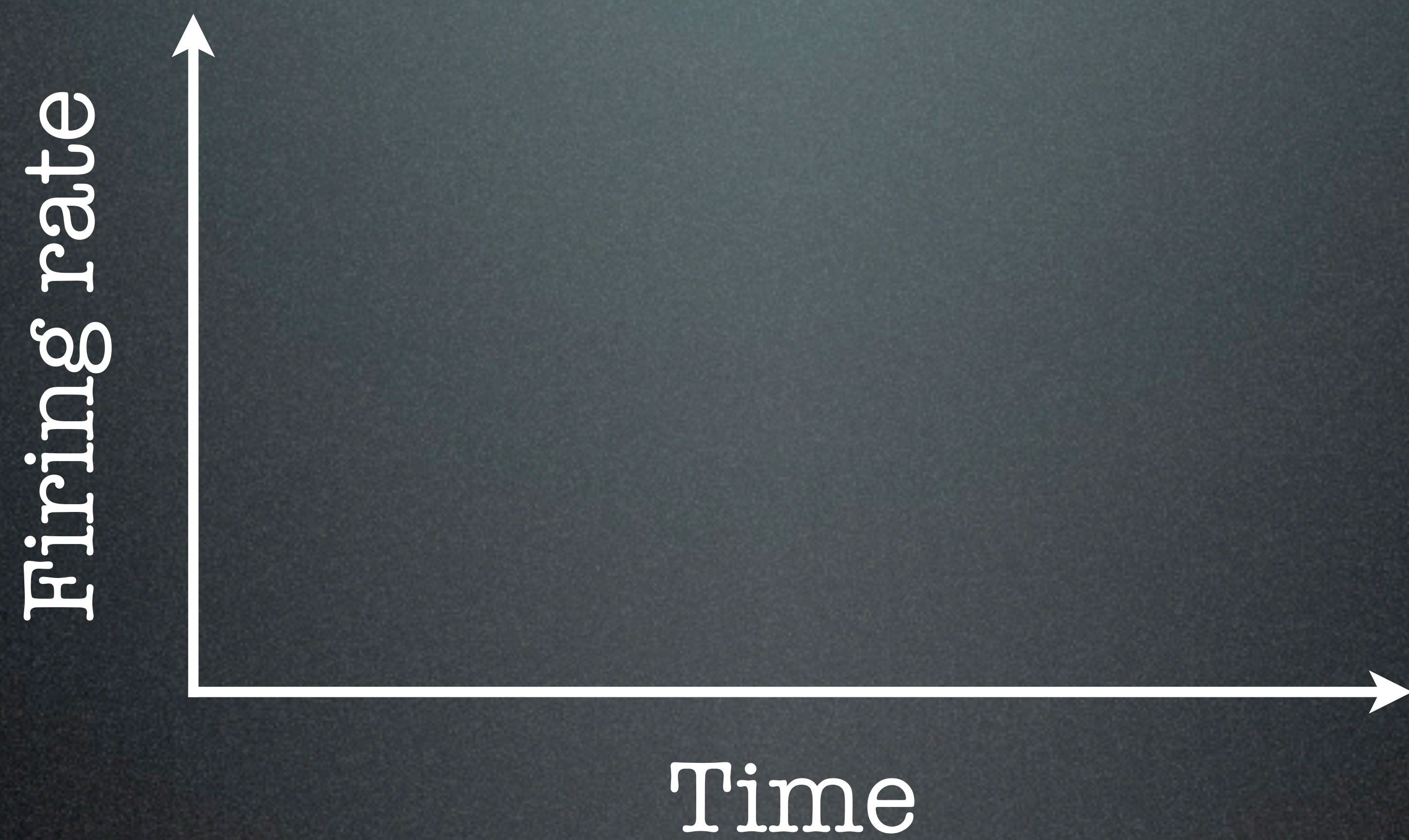
persistent representation from transient stimuli

Working Memory

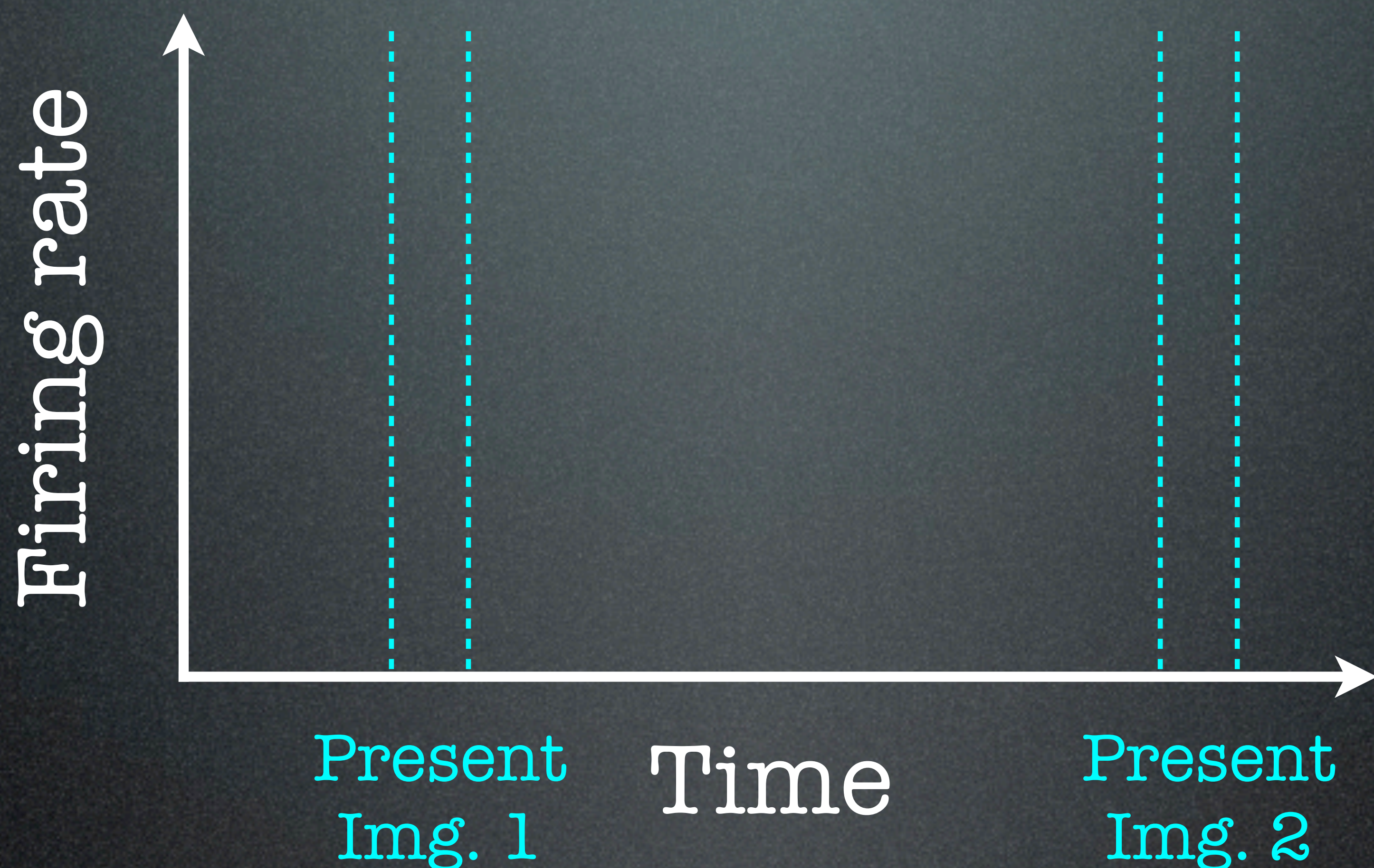
persistent representation from transient stimuli

Traditional Explanation:
Constant Percept by Constant Activity

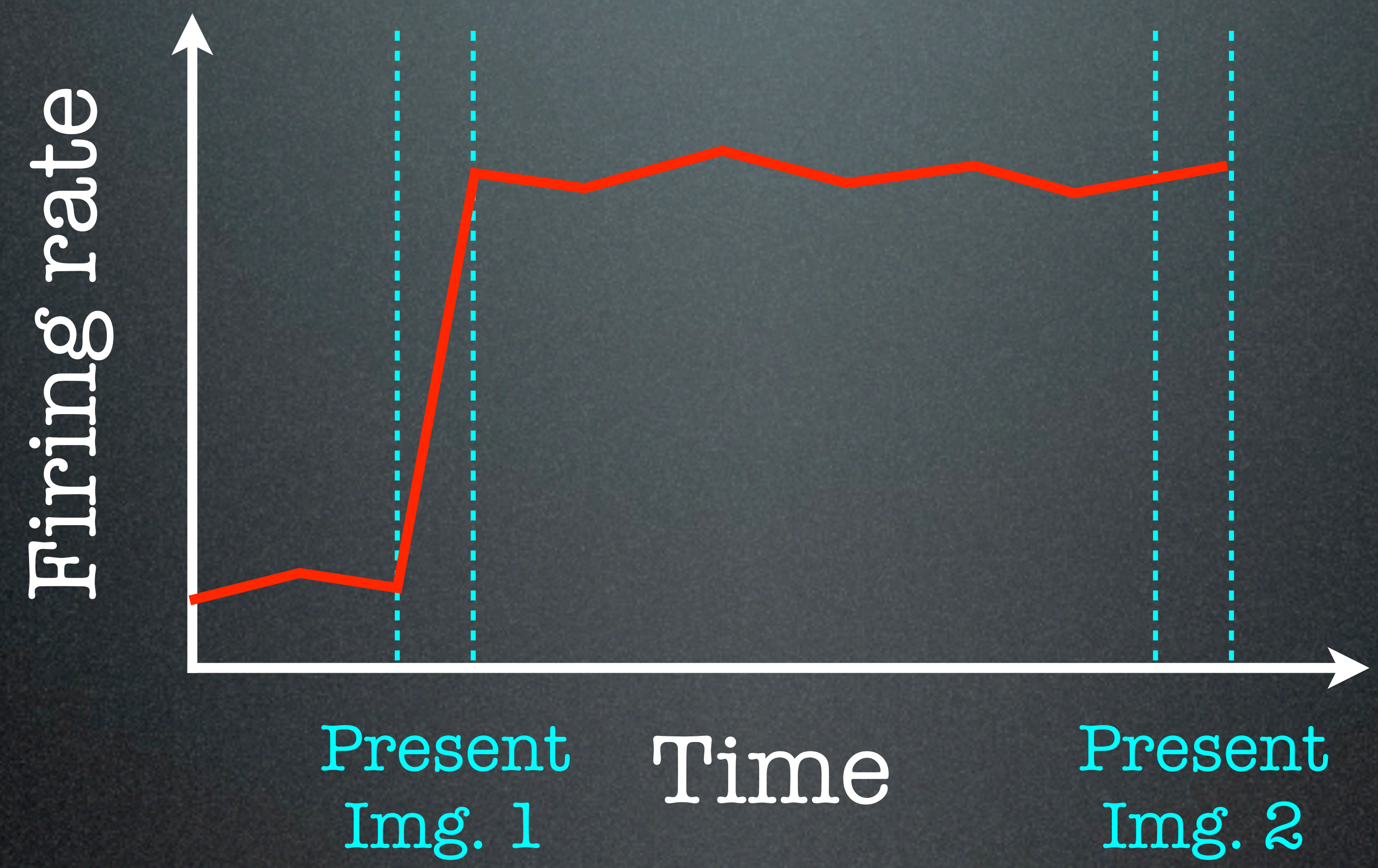
Traditional Explanation: Constant Percept by Constant Activity



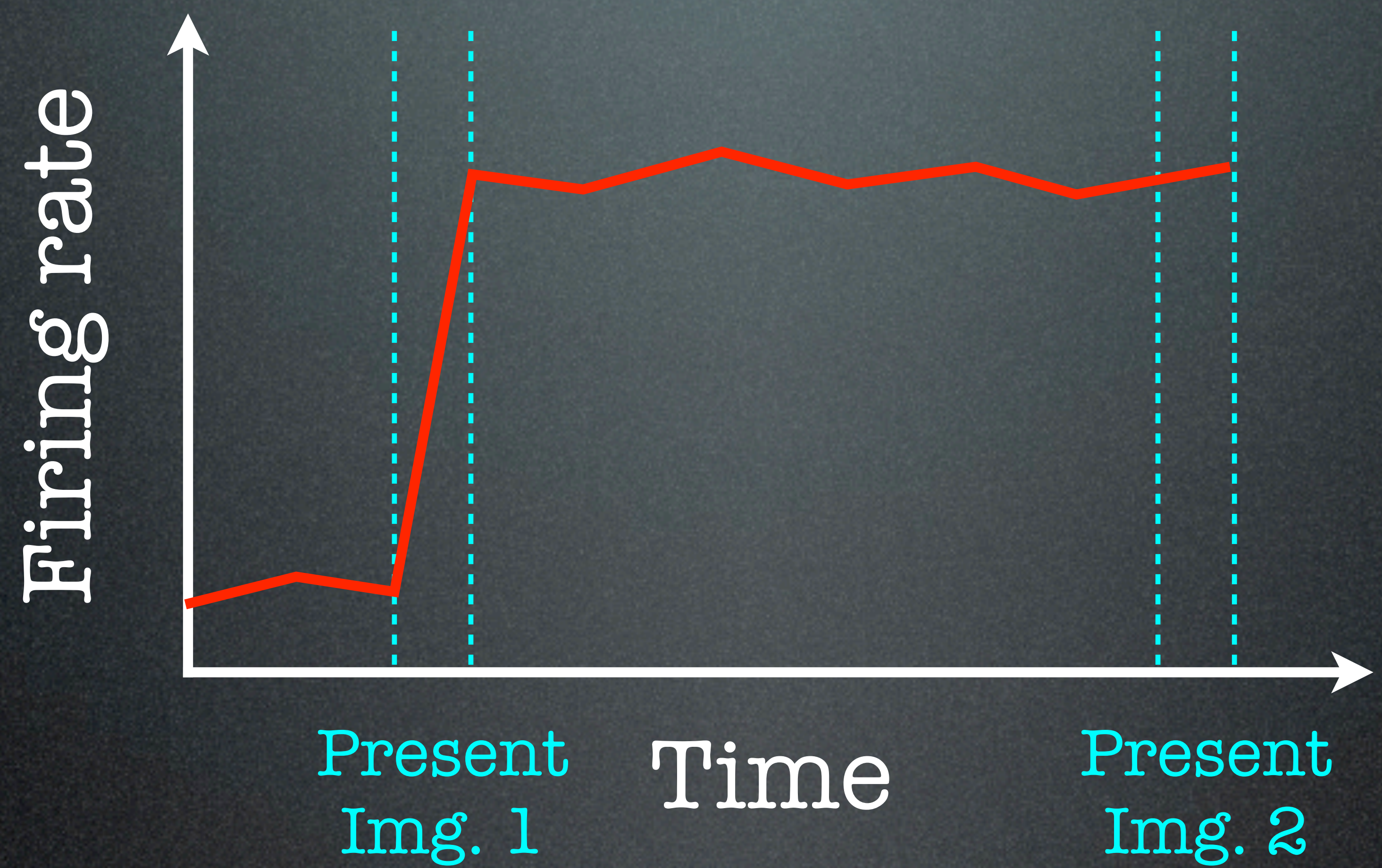
Traditional Explanation: Constant Percept by Constant Activity

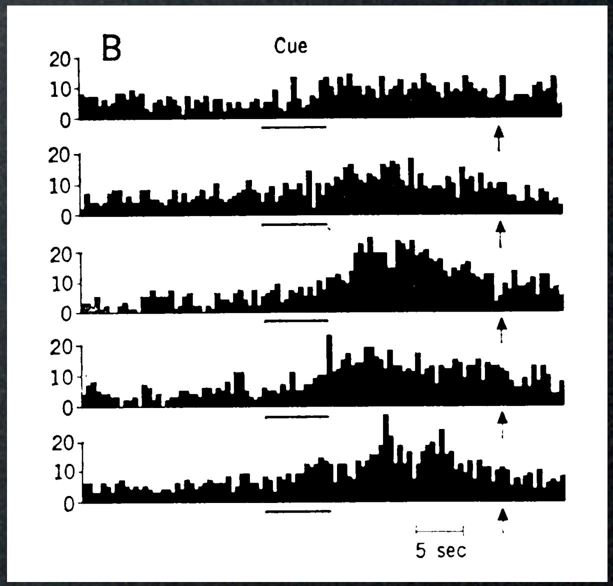


Traditional Explanation: Constant Percept by Constant Activity



Traditional Explanation: Constant Percept by Constant Activity

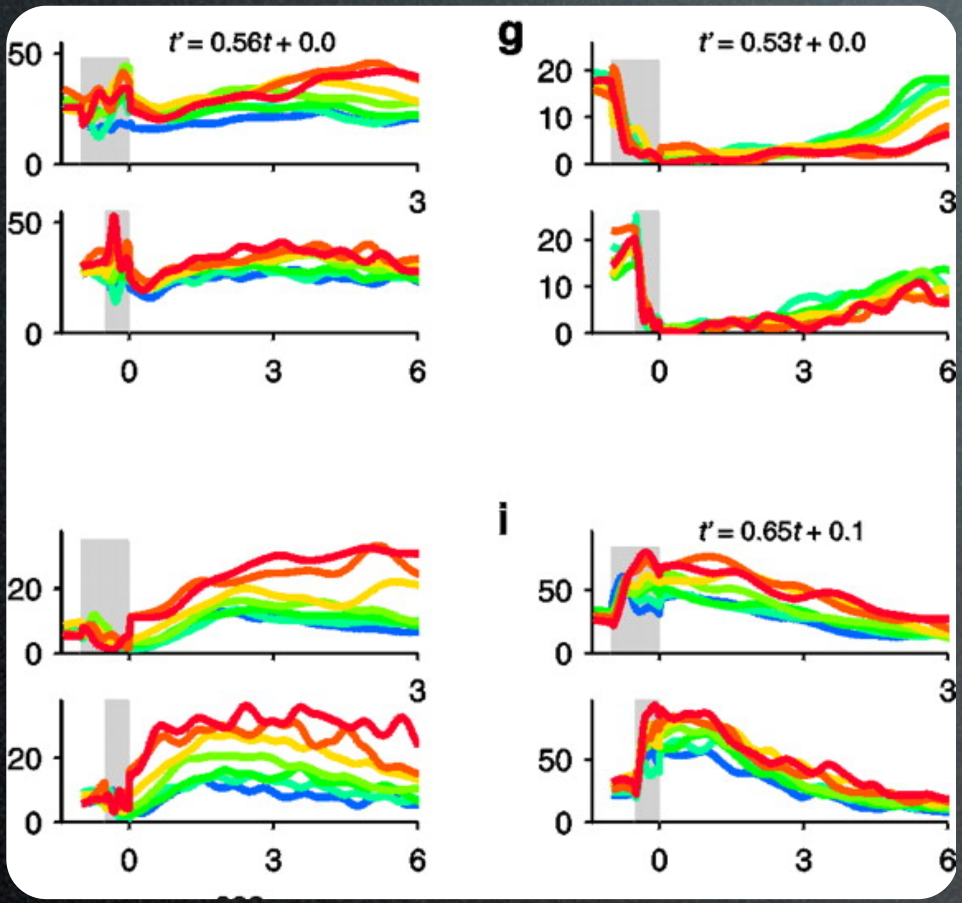




Fuster and Alexander (1971)

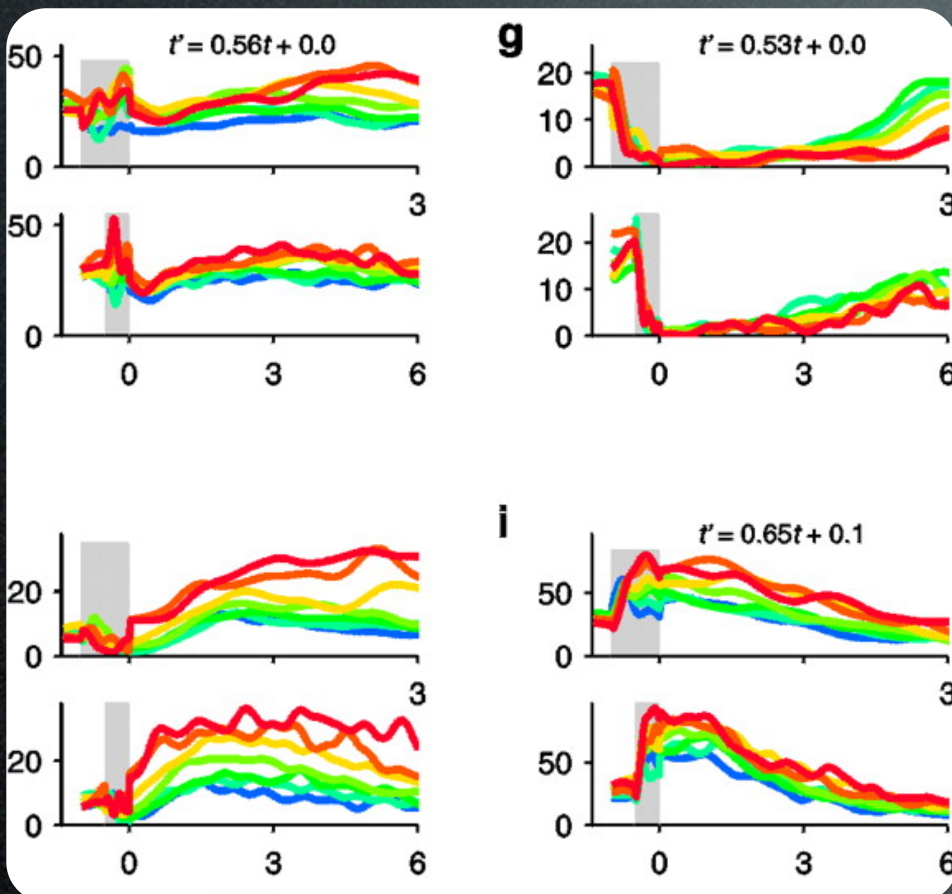
Problem: Time-invariant Activity is rare

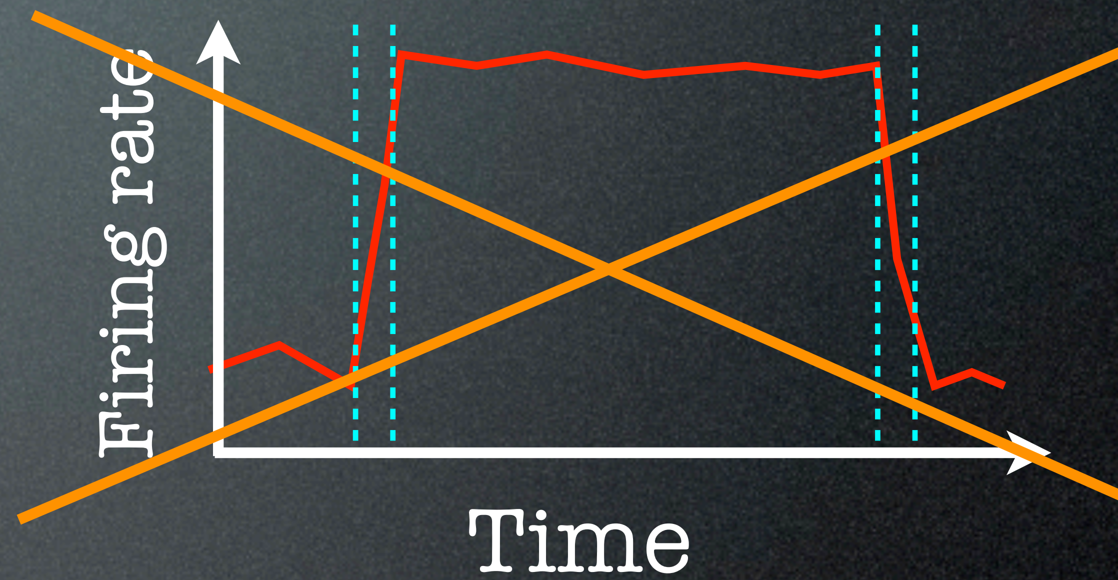
Problem: Time-invariant Activity is rare



Brody, alii, et Romo, 2003

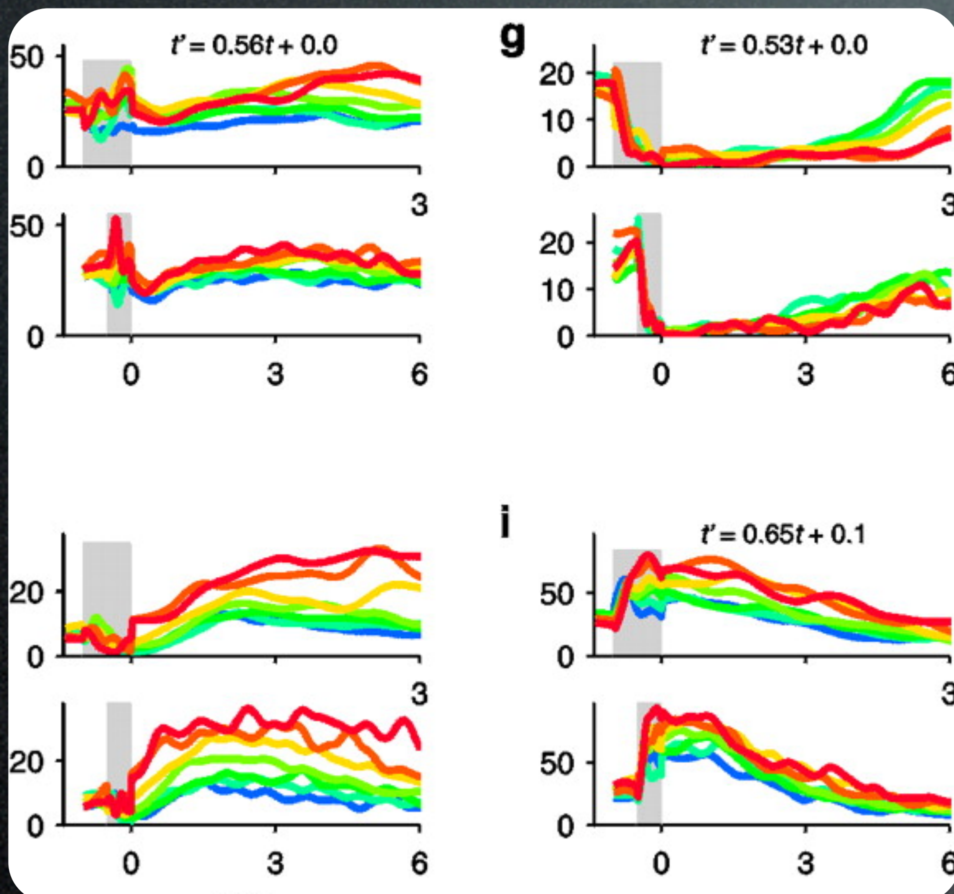
Problem: Time-invariant Activity is rare



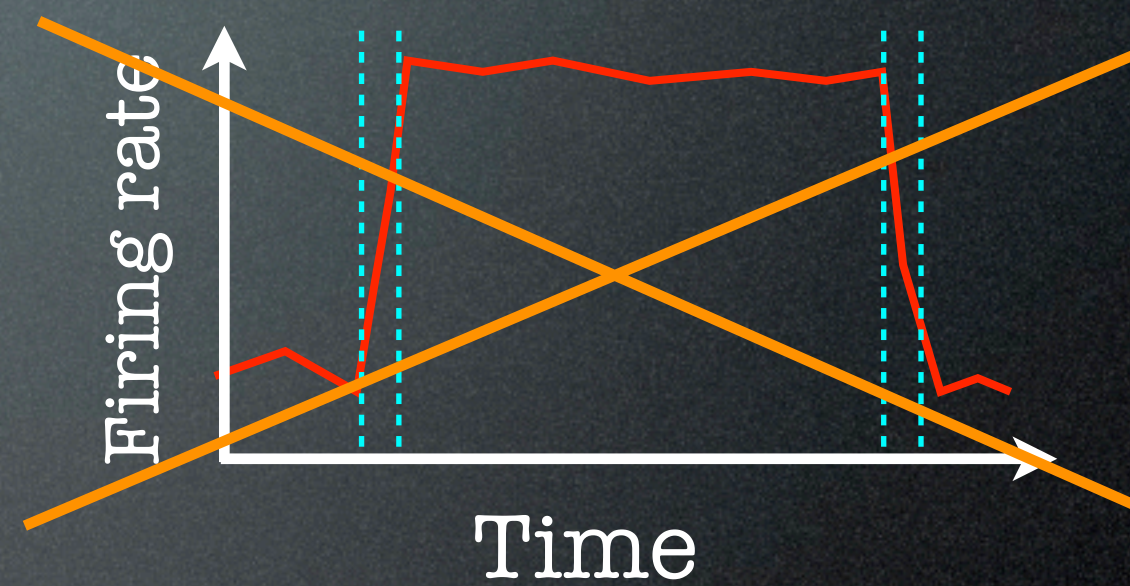


Brody, alii, et Romo, 2003

Problem: Time-invariant Activity is rare



Brody, alii, et Romo, 2003

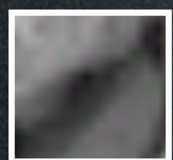
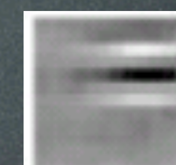
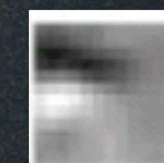


Can time-variant neuronal activity represent time invariant percepts?

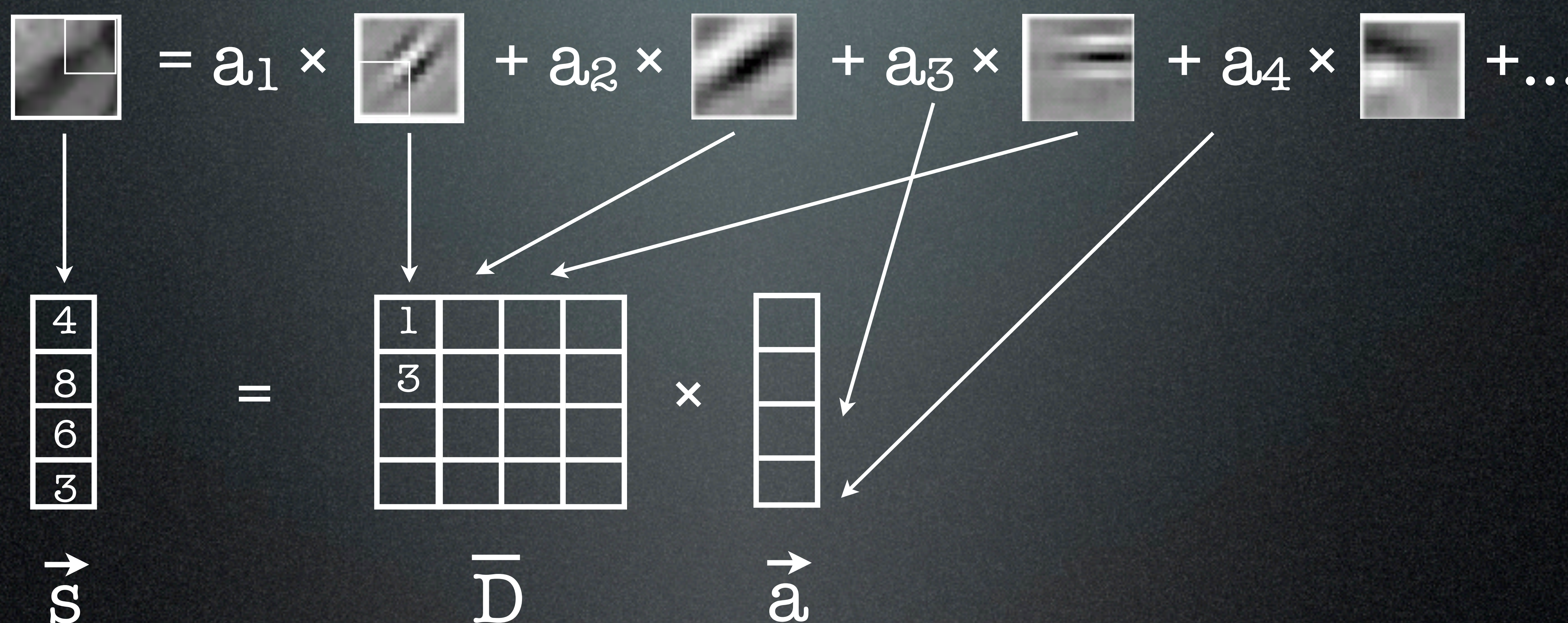
Traditional answer: No!

Linear encoding with orth. basis: persistent percepts → persistent activity

Linear encoding with orth. basis: persistent percepts → persistent activity

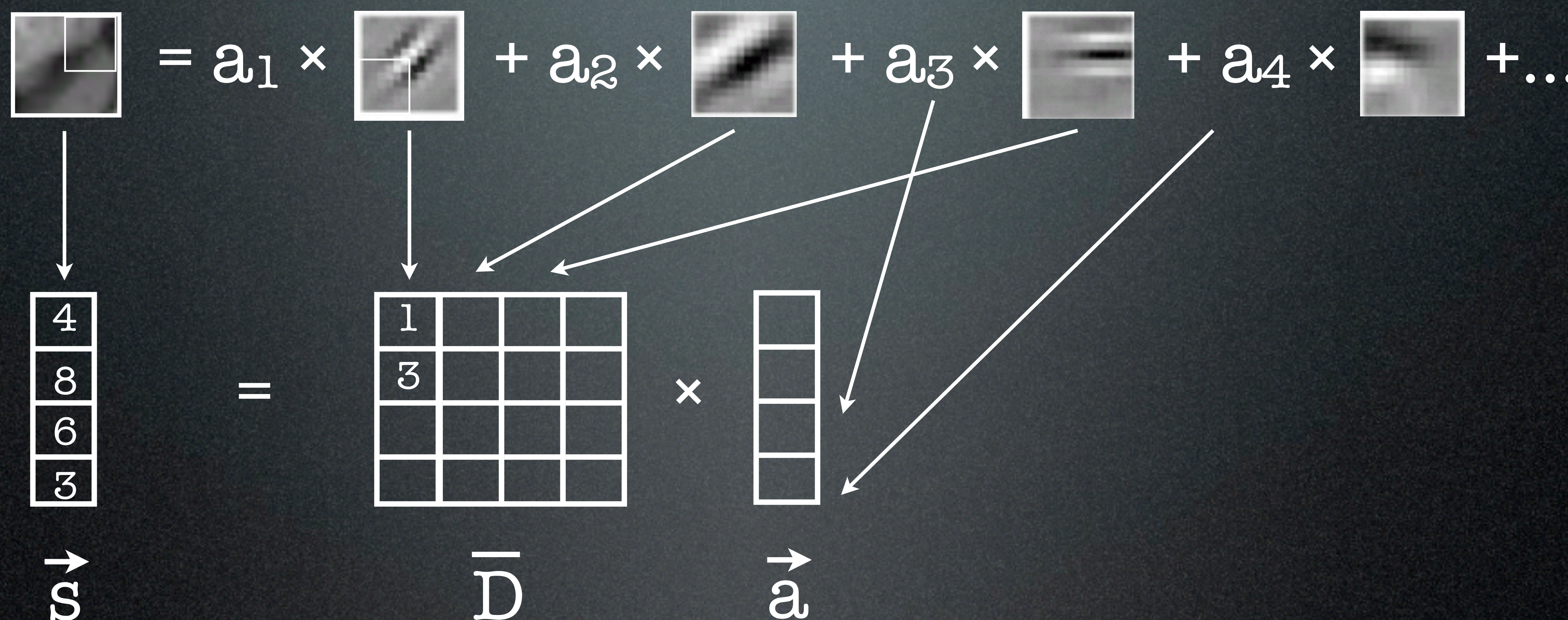
 $= a_1 \times$  $+ a_2 \times$  $+ a_3 \times$  $+ a_4 \times$  $+ \ldots$

Linear encoding with orth. basis: persistent percepts → persistent activity



$$s = Da$$

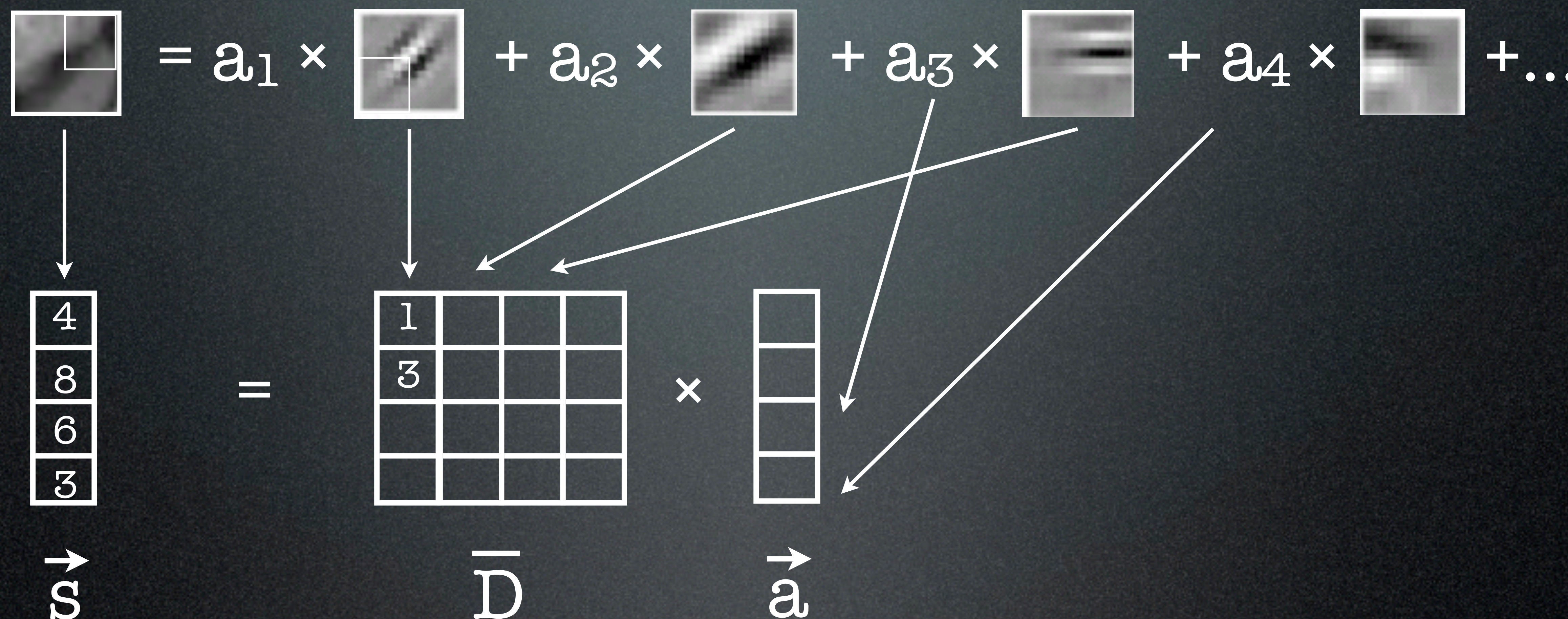
Linear encoding with orth. basis: persistent percepts → persistent activity



$$s = Da$$

Persistent percept: $\frac{ds}{dt} = 0$

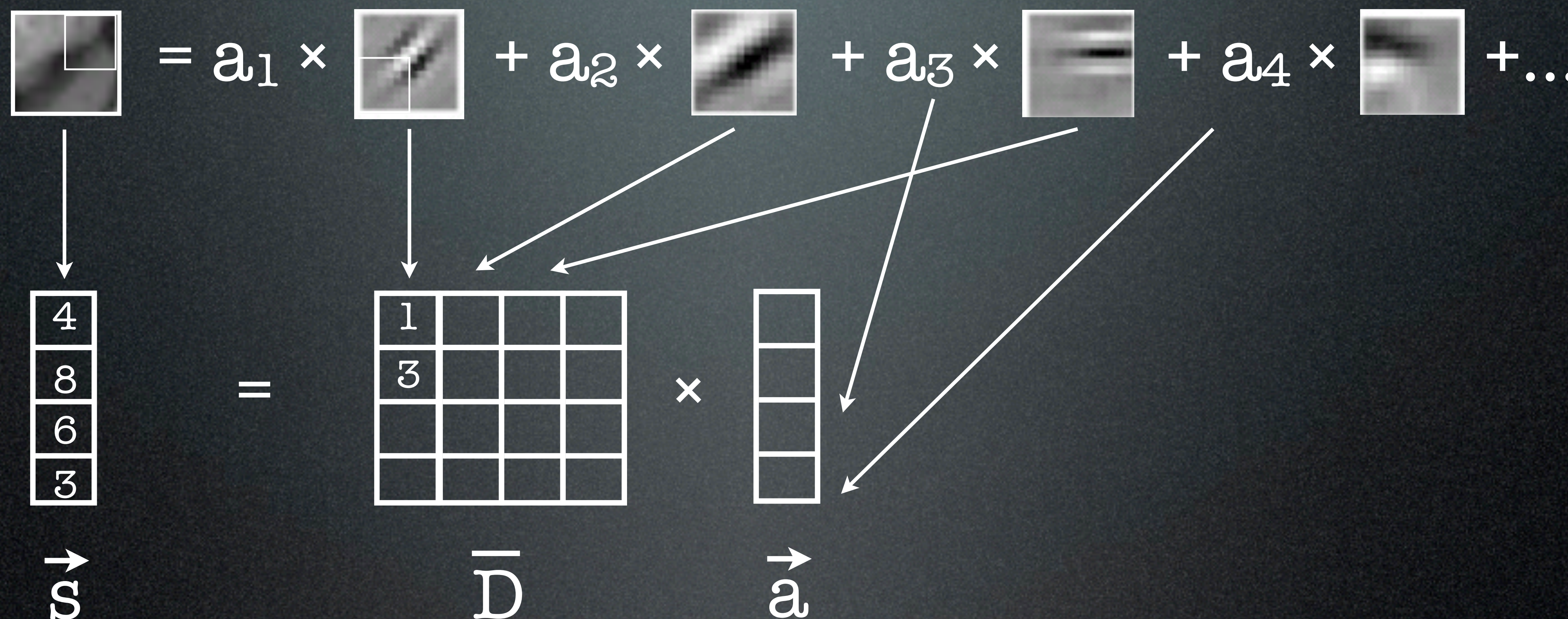
Linear encoding with orth. basis: persistent percepts → persistent activity



$$s = Da$$

Persistent percept: $\frac{ds}{dt} = 0 \Longrightarrow$

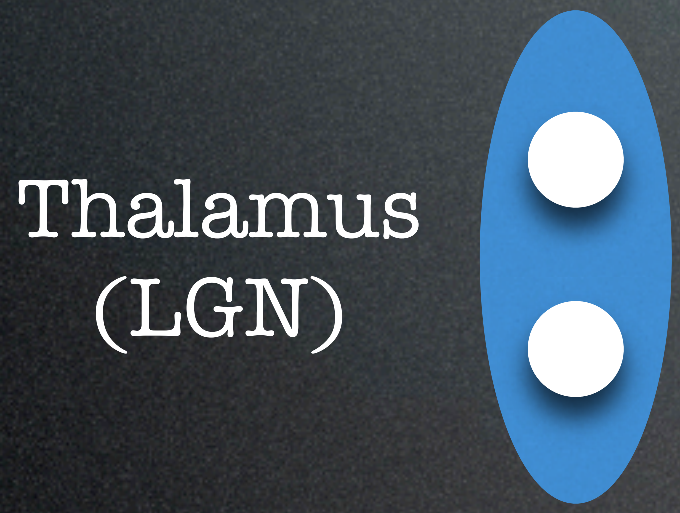
Linear encoding with orth. basis: persistent percepts → persistent activity



$$s = Da$$

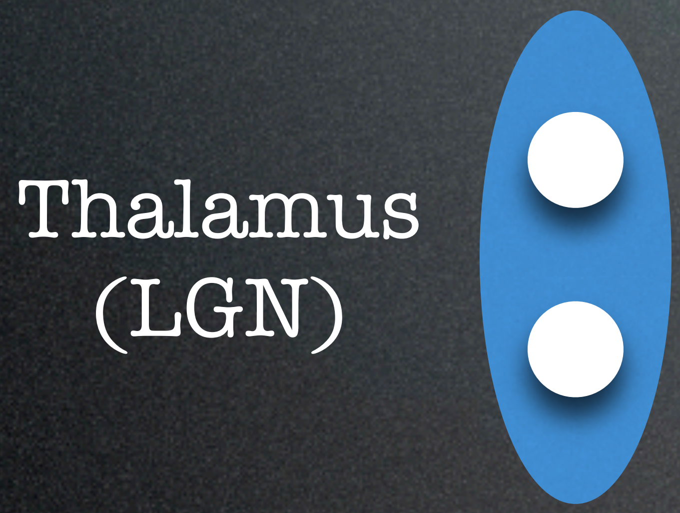
Persistent percept: $\frac{ds}{dt} = 0 \Longrightarrow$ Persistent activity: $\frac{da}{dt} = 0$

Thalamus-Cortex Divergence



Thalamus-Cortex Divergence

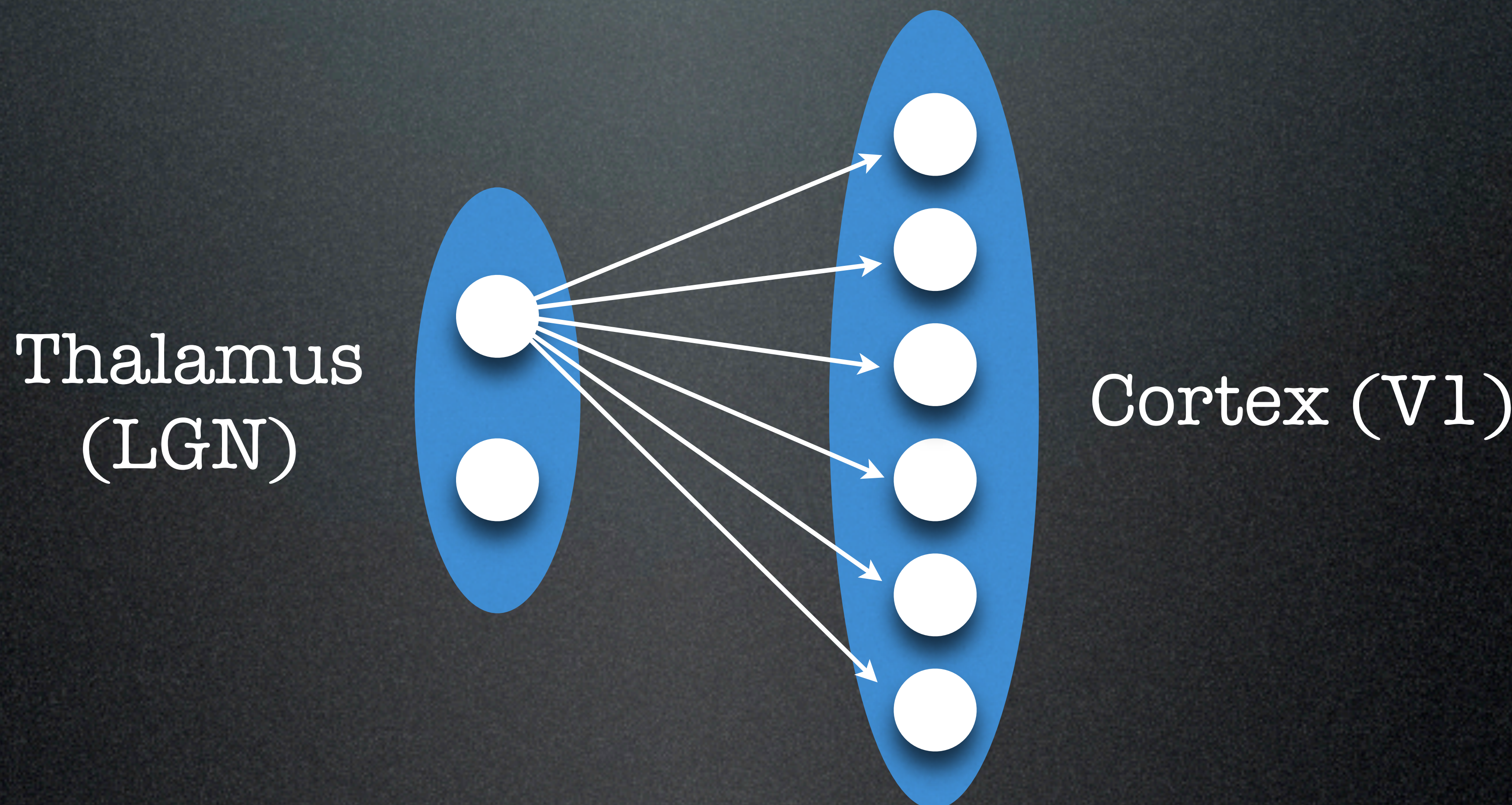
Input structure to Cortex is the Thalamus



Thalamus-Cortex Divergence

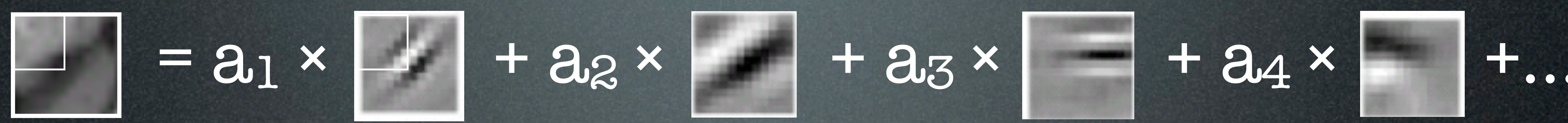
Input structure to Cortex is the Thalamus

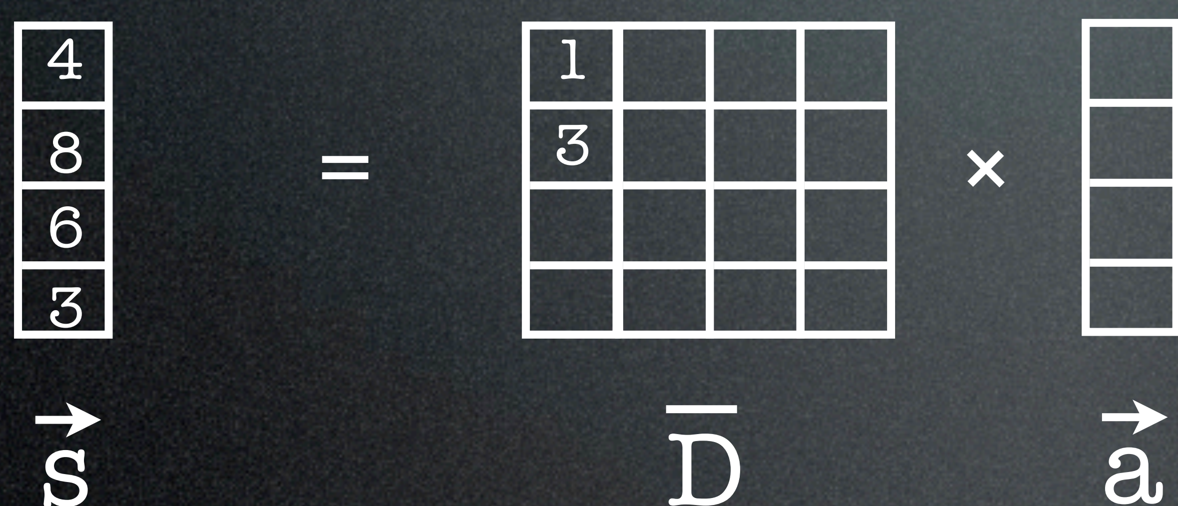
Number of cortical neurons much greater than the number of thalamic input neurons



Cortex uses a non-orthogonal (over-complete) representation

Linear Encoding in Overcomplete Representation?

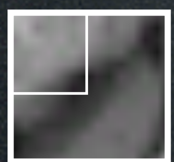
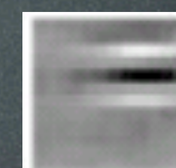
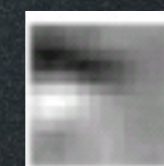


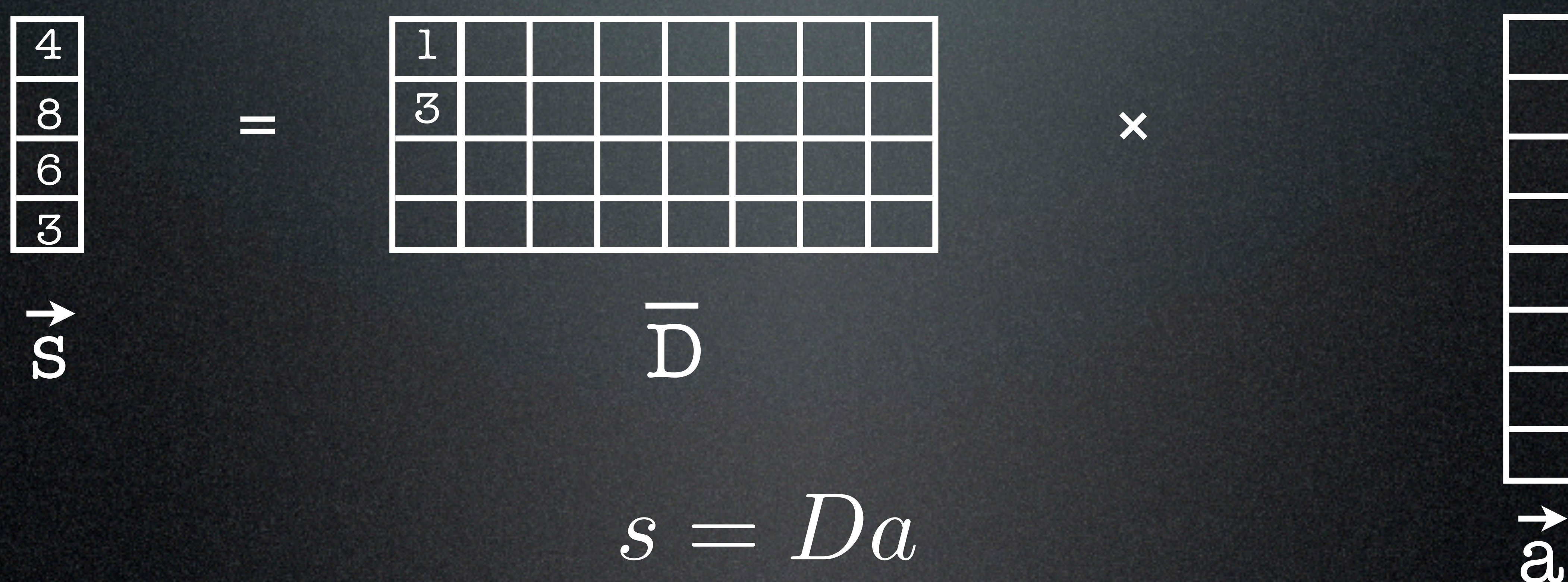


$$s = Da$$

Persistent percept: $\frac{ds}{dt} = 0 \Longrightarrow$

Linear Encoding in Overcomplete Representation?

 $= a_1 \times$  $+ a_2 \times$  $+ a_3 \times$  $+ a_4 \times$  $+ \ldots$



$$s = Da$$

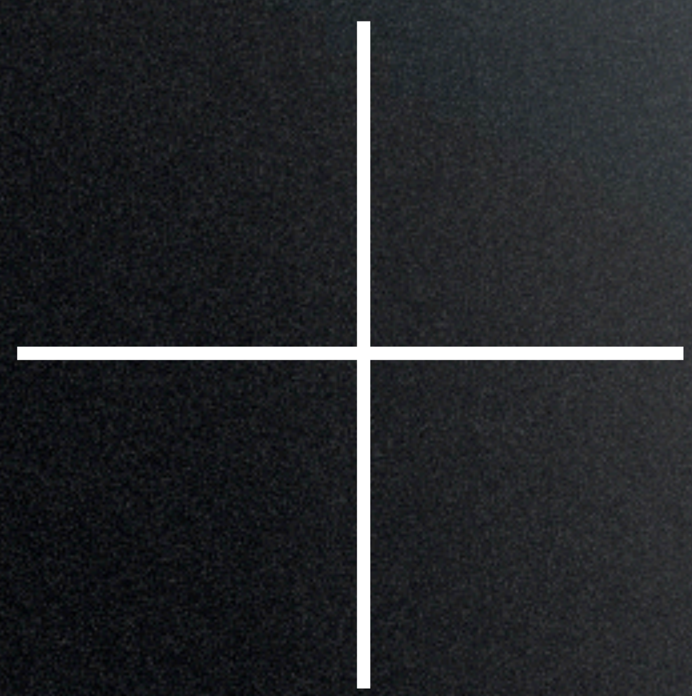
Persistent percept: $\frac{ds}{dt} = 0 \Longrightarrow$?

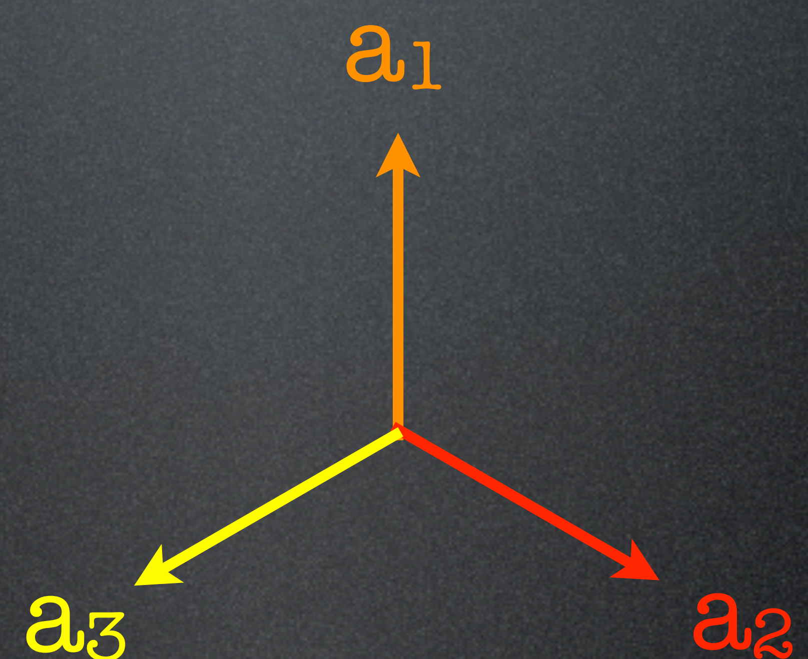
Over-complete Representation is Non-unique

Over-complete Representation is Non-unique

Stimulus dimension: 2

Number of neurons (activity dimension): 3

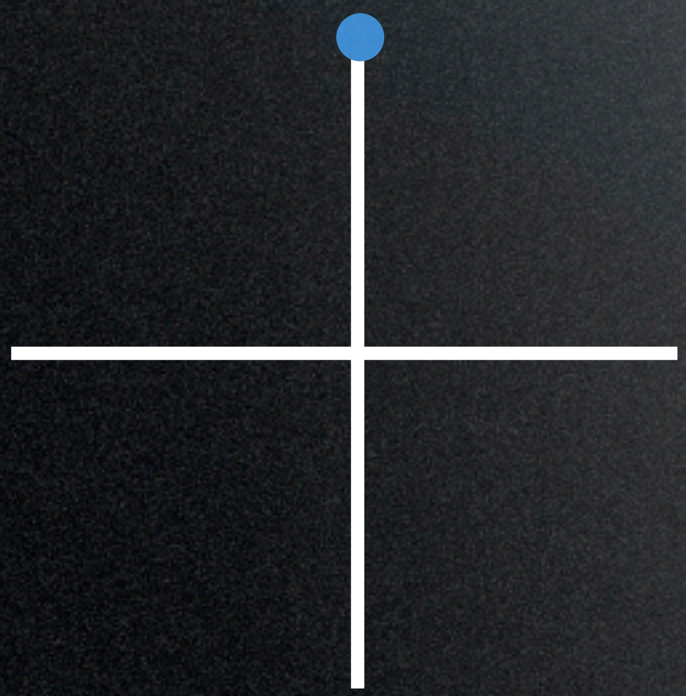


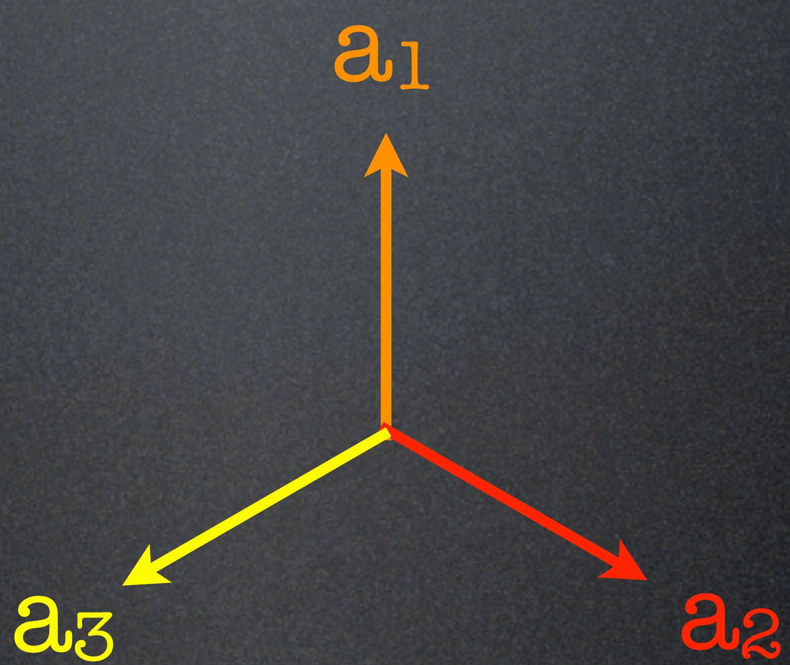


Over-complete Representation is Non-unique

Stimulus dimension: 2

Number of neurons (activity dimension): 3

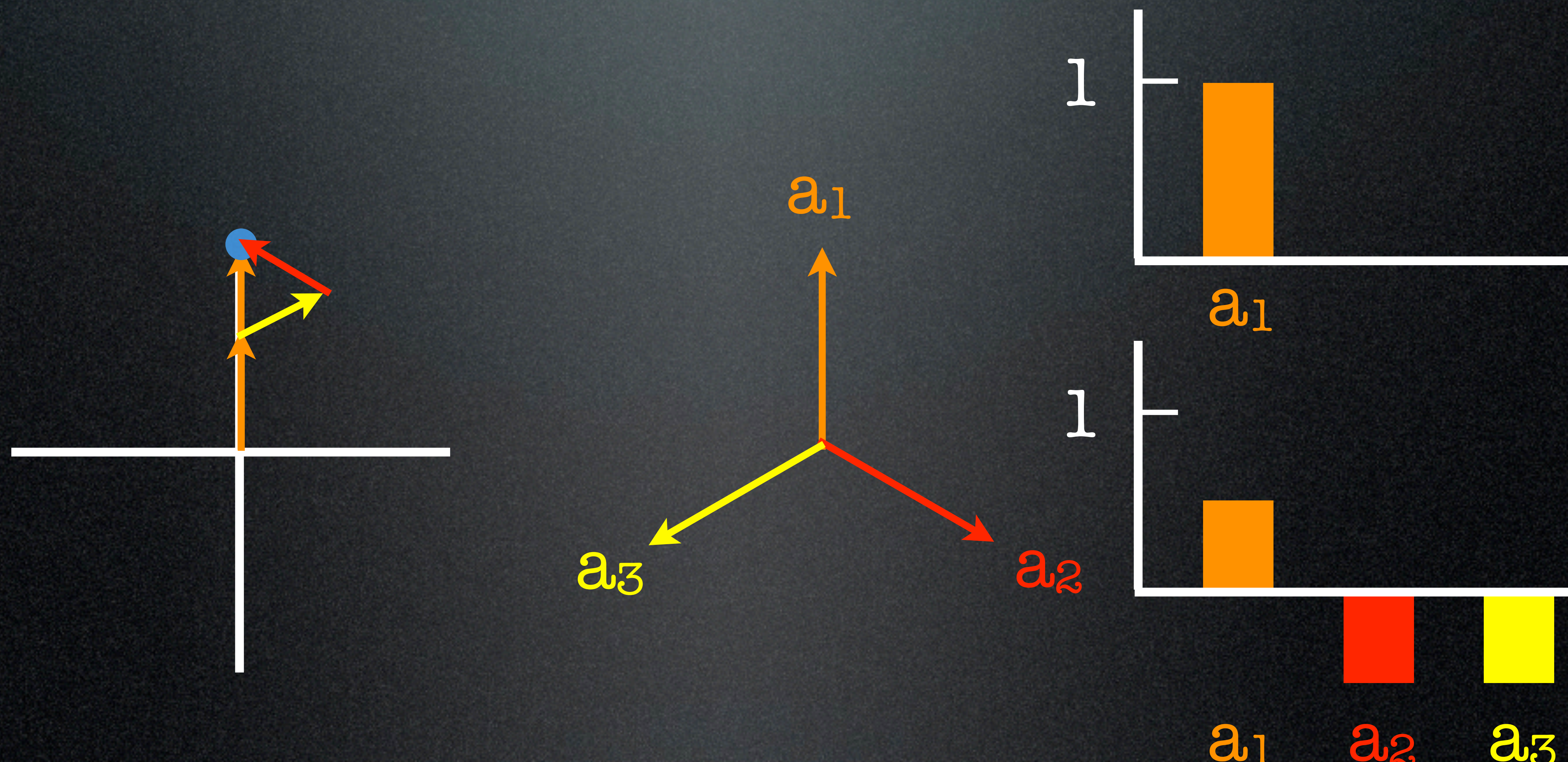




Over-complete Representation is Non-unique

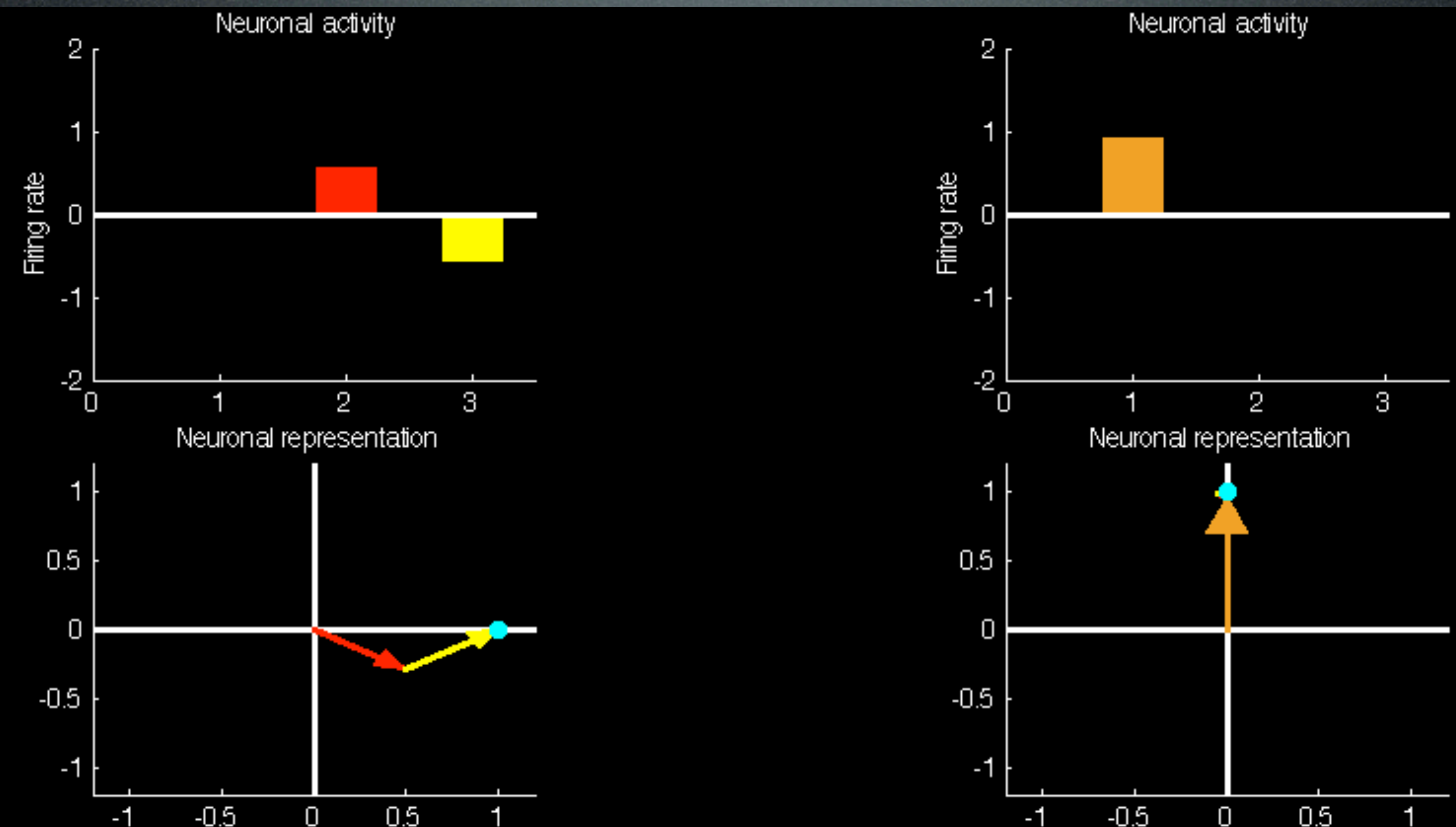
Stimulus dimension: 2

Number of neurons (activity dimension): 3

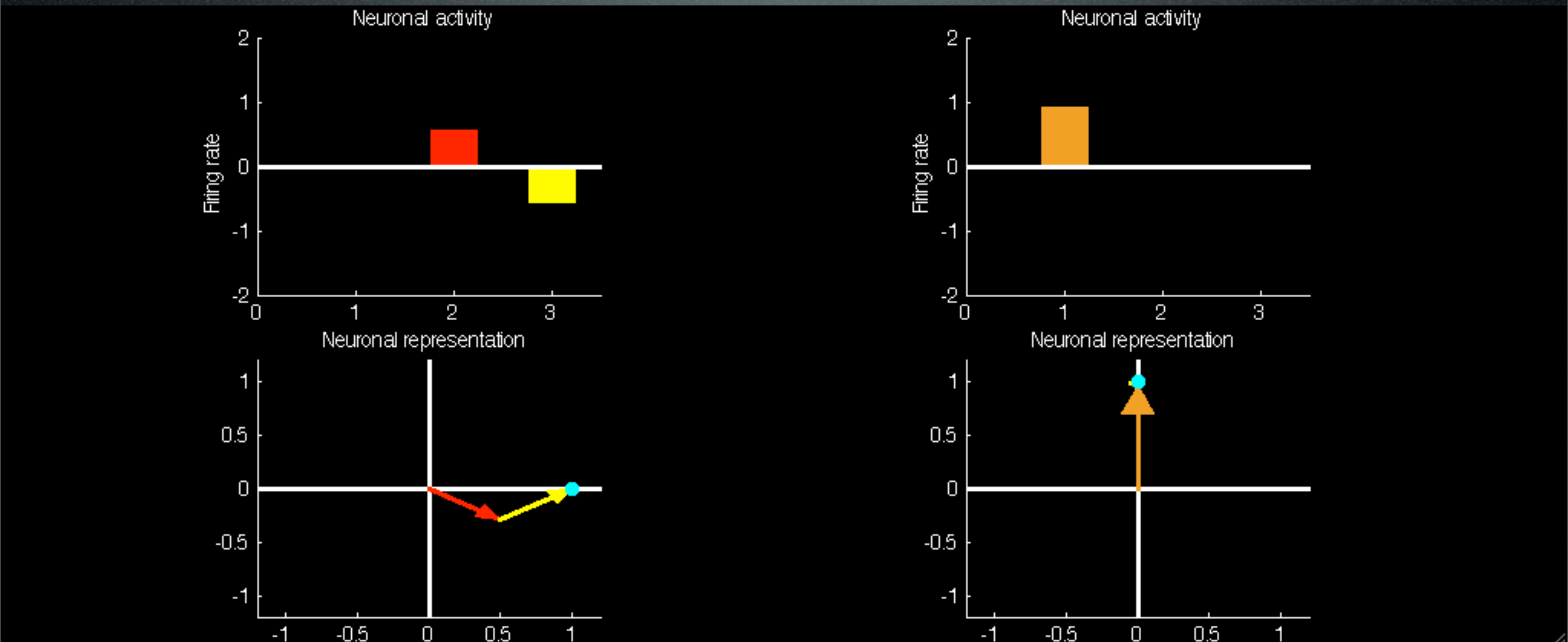


Over-complete Representation is Non-unique

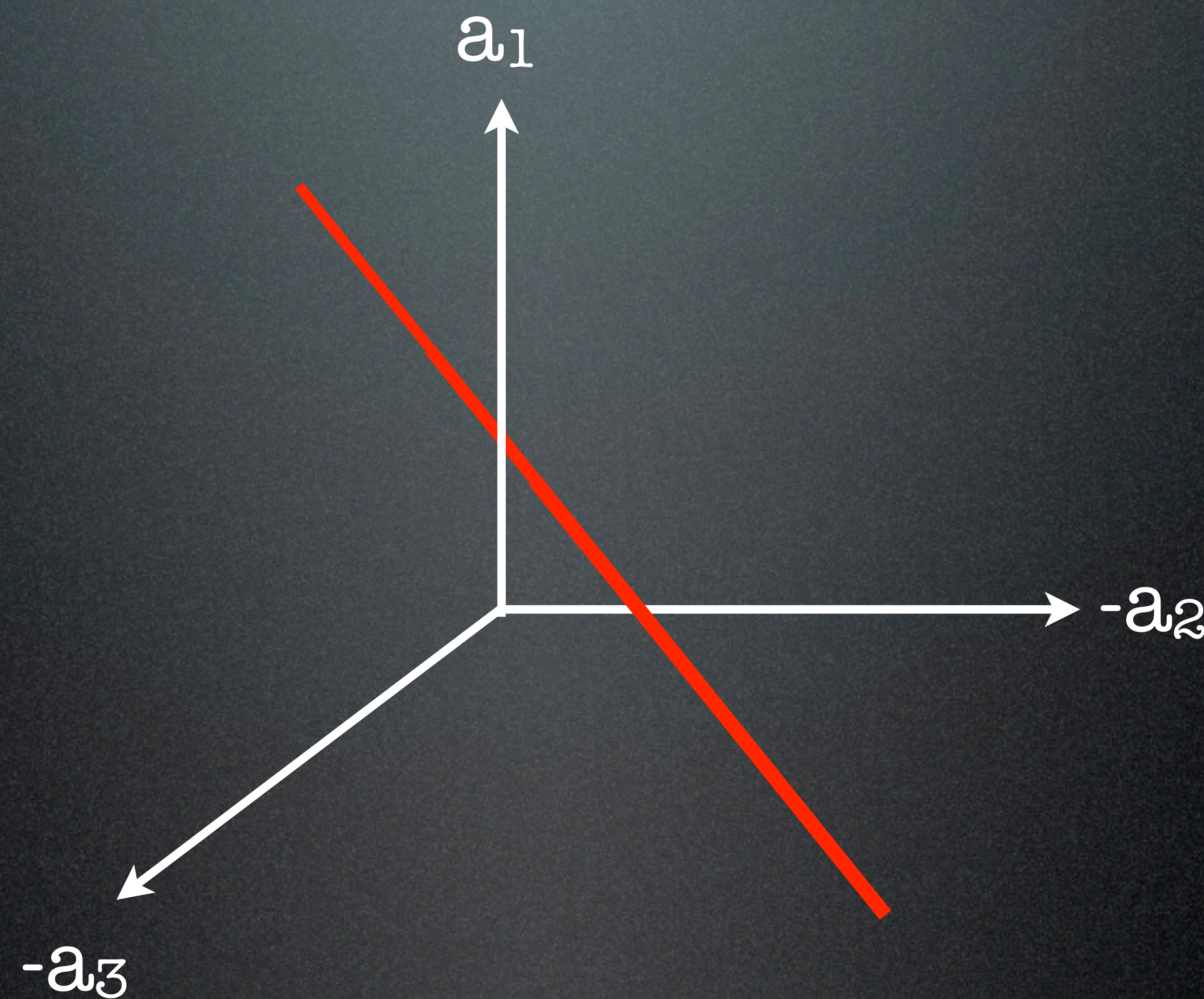
Over-complete Representation is Non-unique



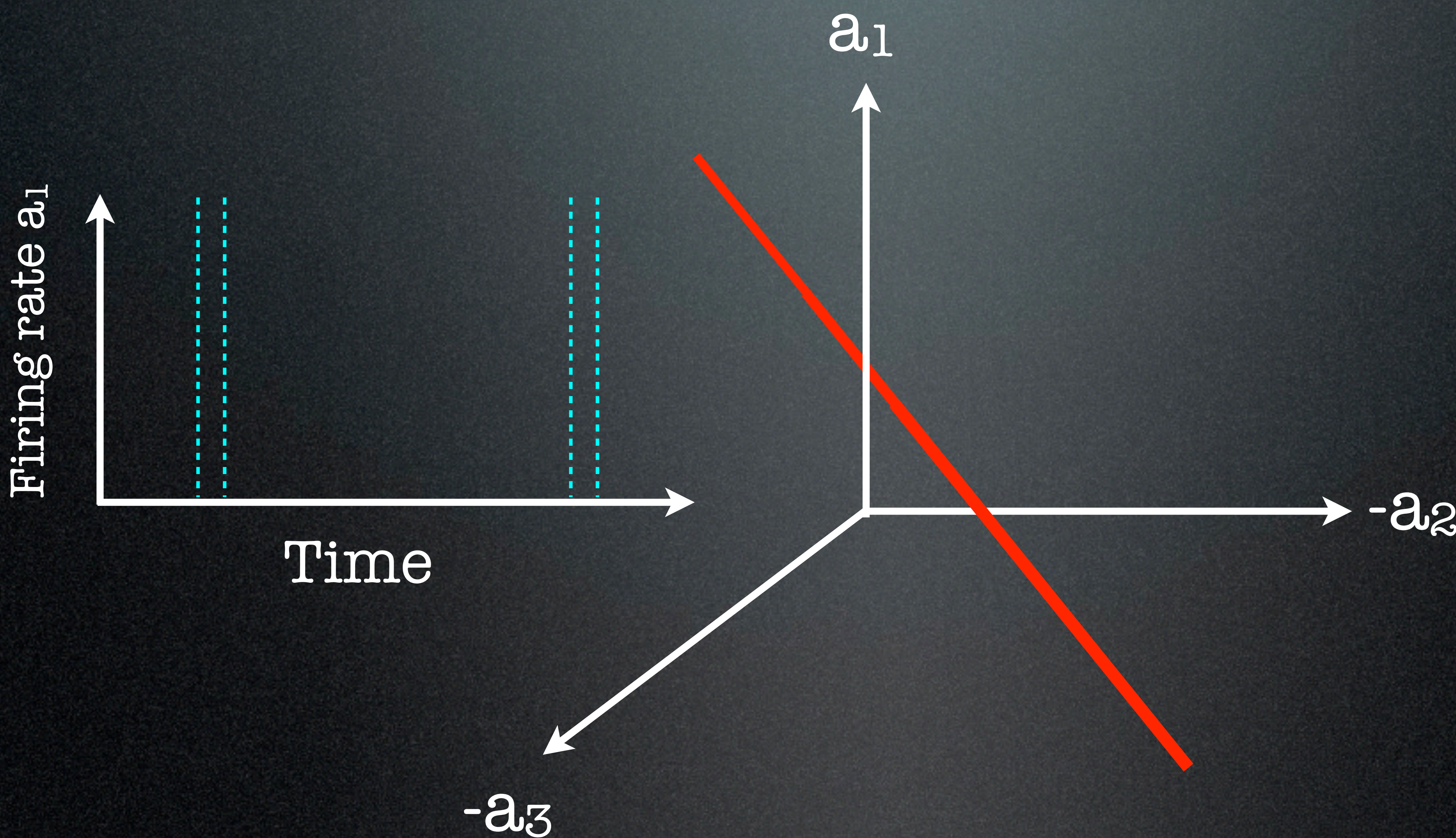
Over-complete Representation is Non-unique



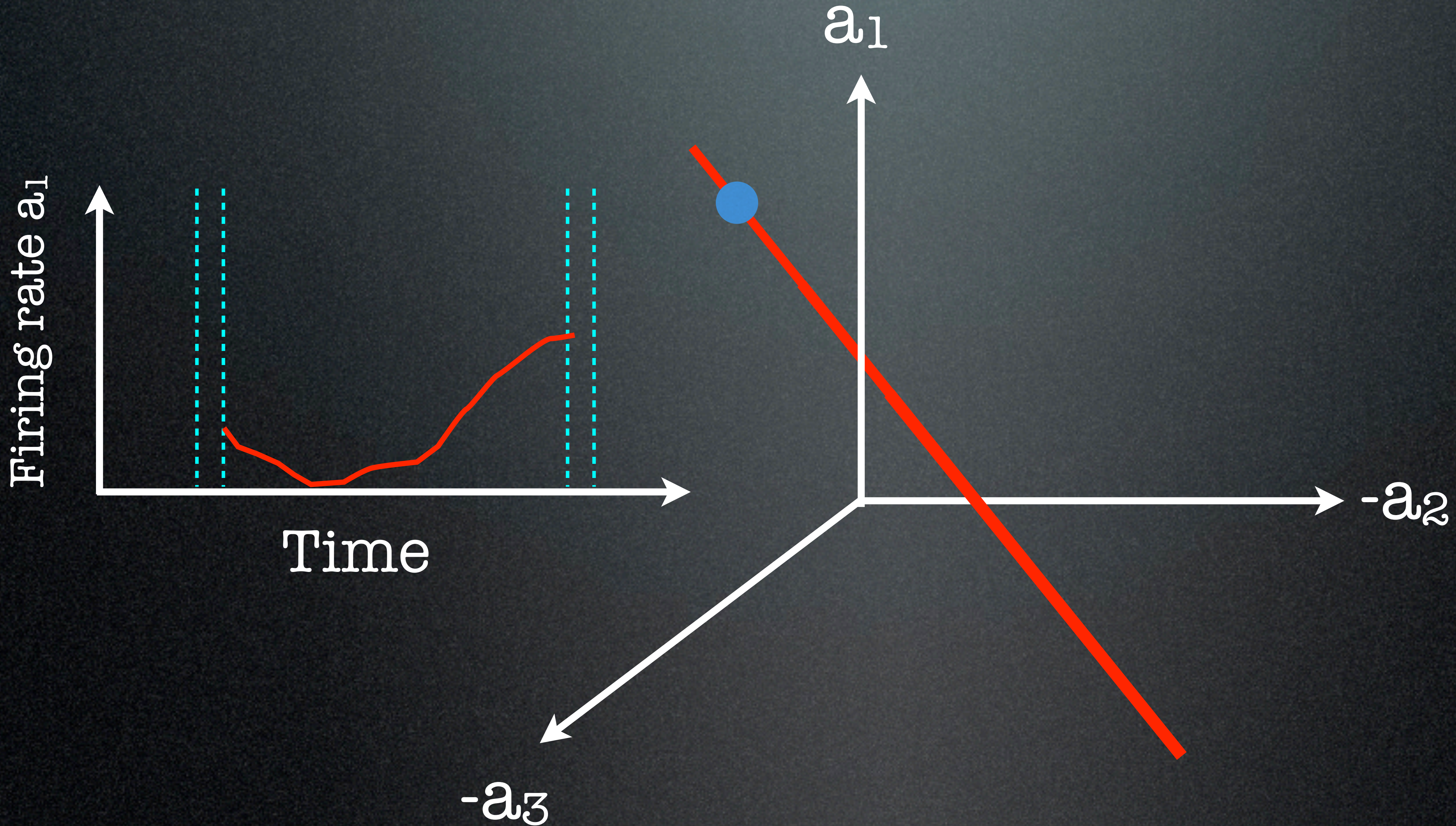
Freedom in Representation in an Over-complete Frame



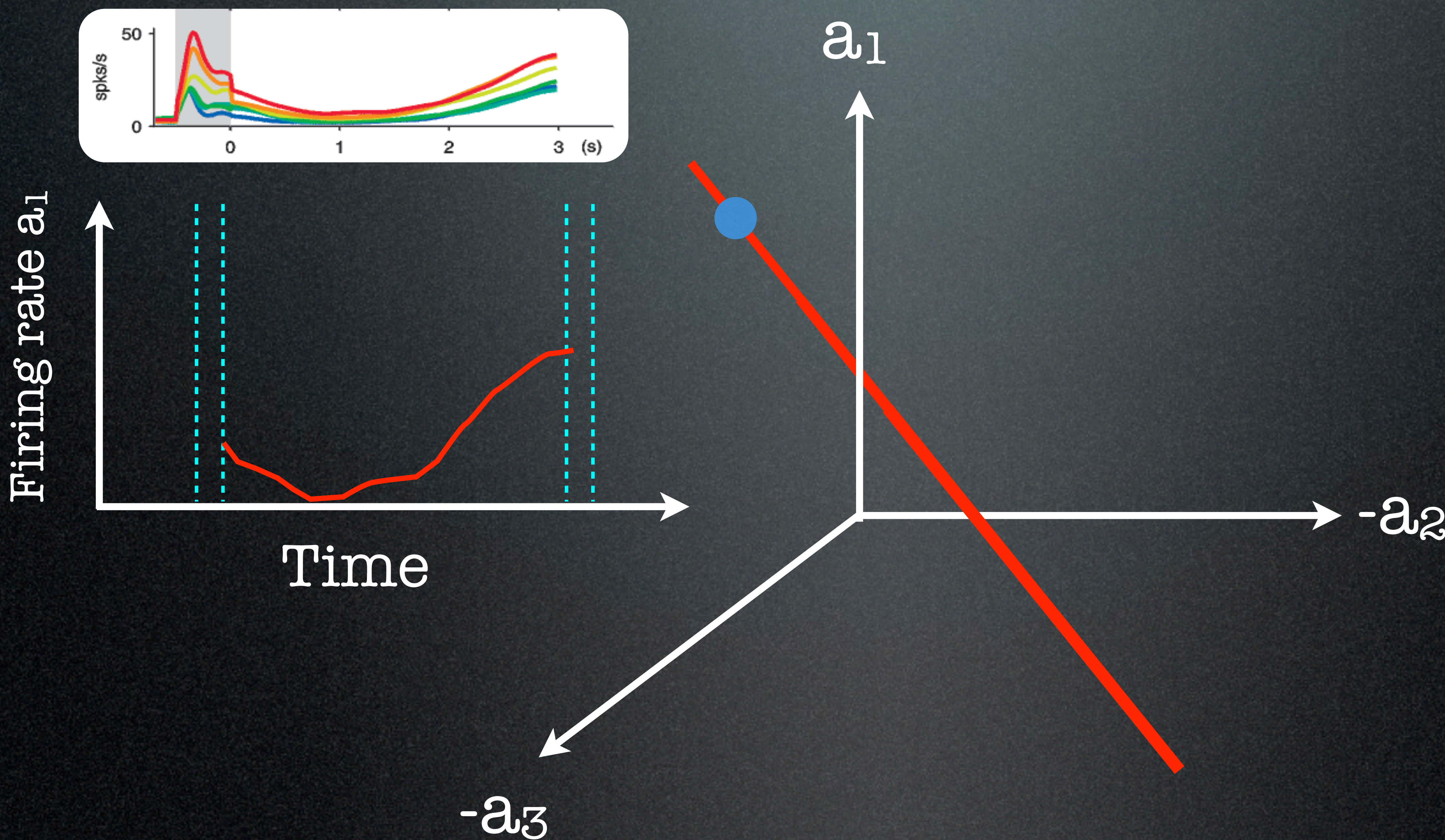
Freedom in Representation in an Over-complete Frame



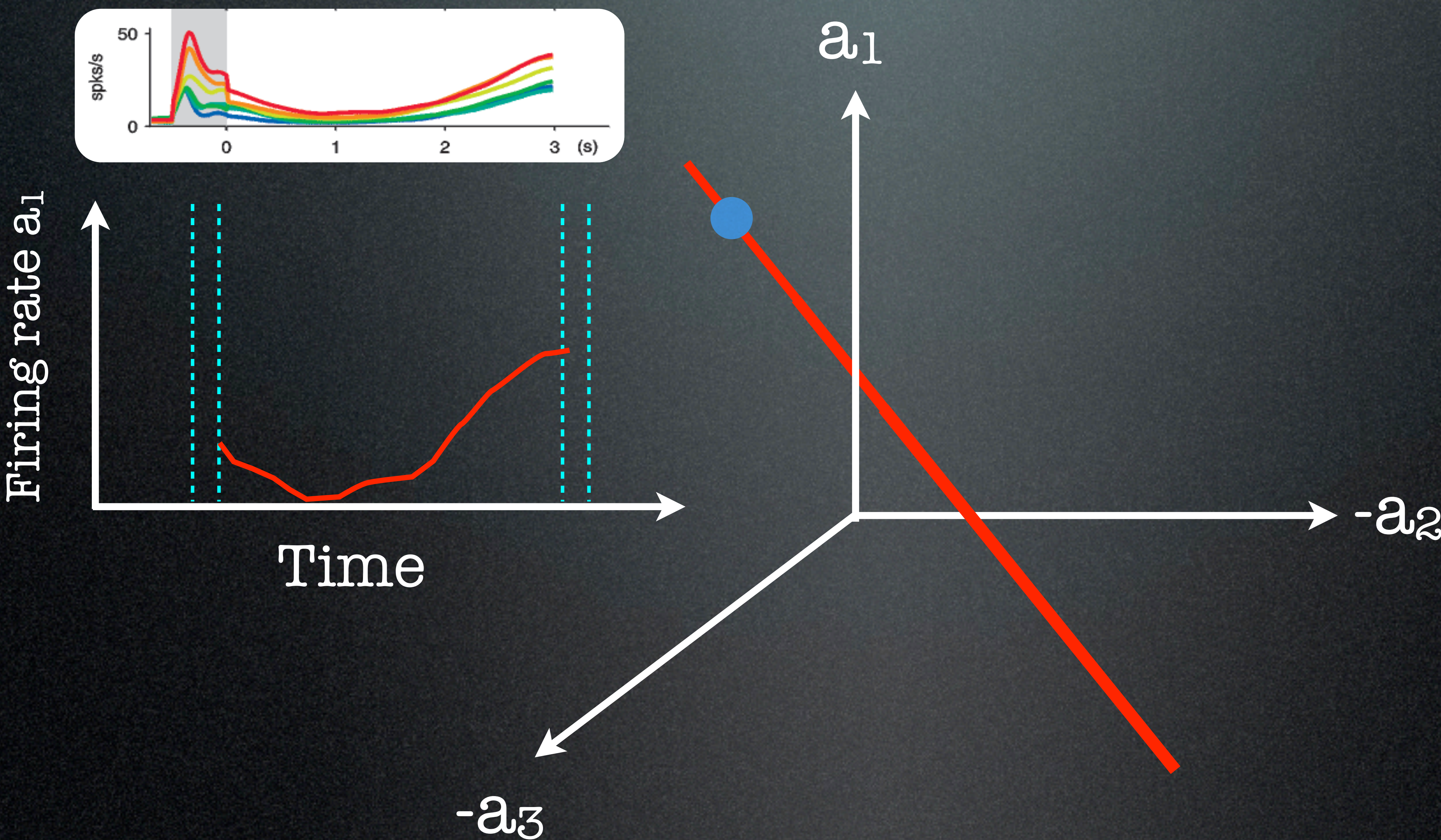
Freedom in Representation in an Over-complete Frame



Freedom in Representation in an Over-complete Frame



Freedom in Representation in an Over-complete Frame



Time-variant neuronal activity can represent a time invariant percept

Lateral connectivity maintains persistency

s: stimulus, a: activity

D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

s: stimulus, a: activity

D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

$\dot{a} = -a + La$ Rate dynamics

s: stimulus, a: activity
D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

$\dot{a} = -a + La$ Rate dynamics

$\dot{s} = D\dot{a} = 0 = D(-a + La)$ Persistency

s: stimulus, a: activity

D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

$\dot{a} = -a + La$ Rate dynamics

$\dot{s} = D\dot{a} = 0 = D(-a + La)$ Persistency

$Da = DLa$ If to hold for all a

s: stimulus, a: activity
D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

$\dot{a} = -a + La$ Rate dynamics

$\dot{s} = D\dot{a} = 0 = D(-a + La)$ Persistency

$Da = DLa$ If to hold for all a

$D = DL$ Family of Solutions

s: stimulus, a: activity
D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

$\dot{a} = -a + La$ Rate dynamics

$\dot{s} = D\dot{a} = 0 = D(-a + La)$ Persistency

$Da = DLa$ If to hold for all a

$D = DL$ Family of Solutions

$L = I$ Trivial solution

s: stimulus, a: activity

D: dictionary (feature vectors), L: lateral connections

Lateral connectivity maintains persistency

$s = Da$ Linear encoding

$\dot{a} = -a + La$ Rate dynamics

$D = DL$ Family of Solutions

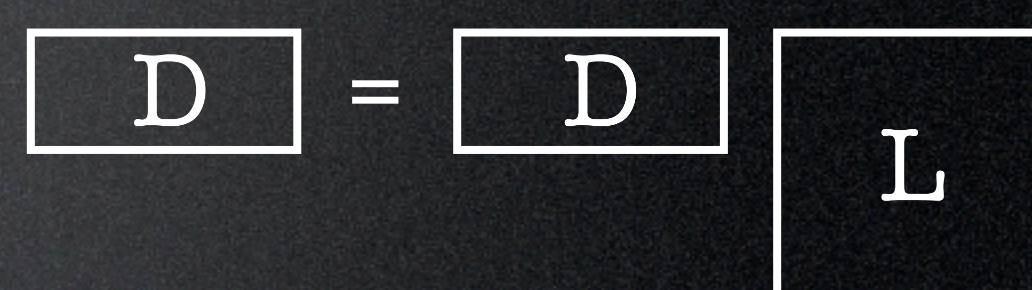
s: stimulus, a: activity
D: dictionary (feature vectors), L: lateral connections

Our Solution: sparse solution

$$D = DL$$

Our Solution: sparse solution

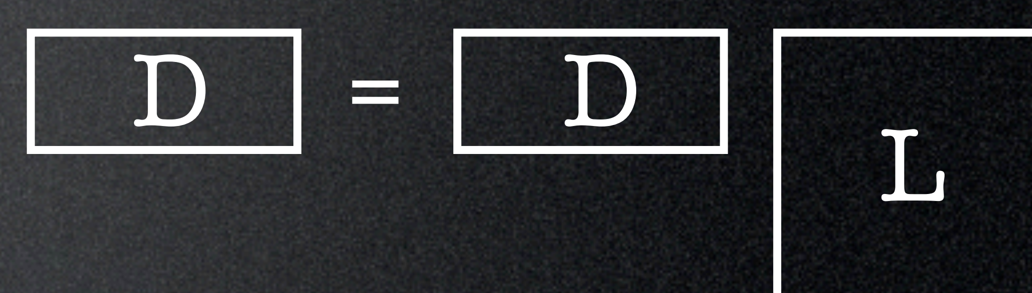
$$D = DL$$



Our Solution: sparse solution

Entries in L represent synaptic connections

$$D = DL$$

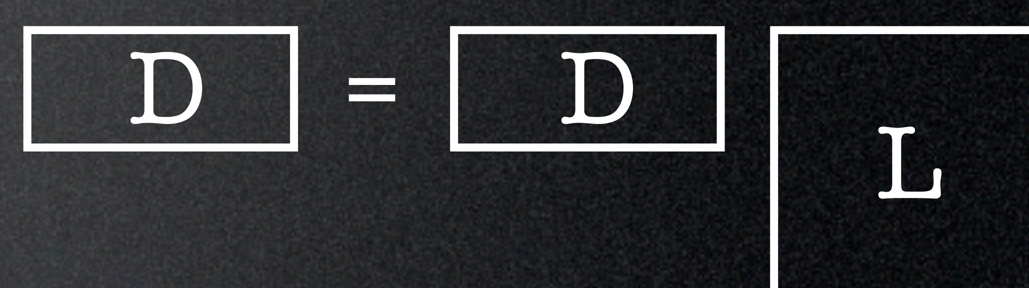


Our Solution: sparse solution

Entries in L represent synaptic connections

We pick the most economic solution, in terms of the resources taken up by synapses

$$D = DL$$



Our Solution: sparse solution

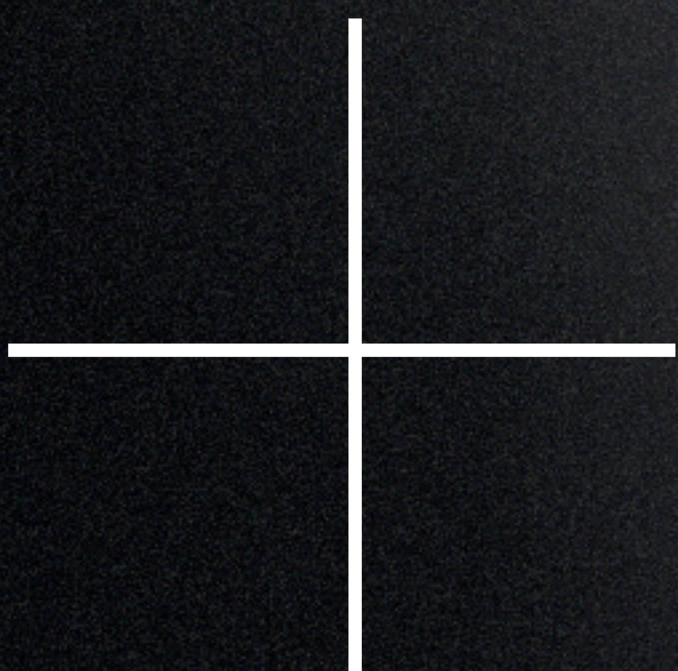
Entries in L represent synaptic connections

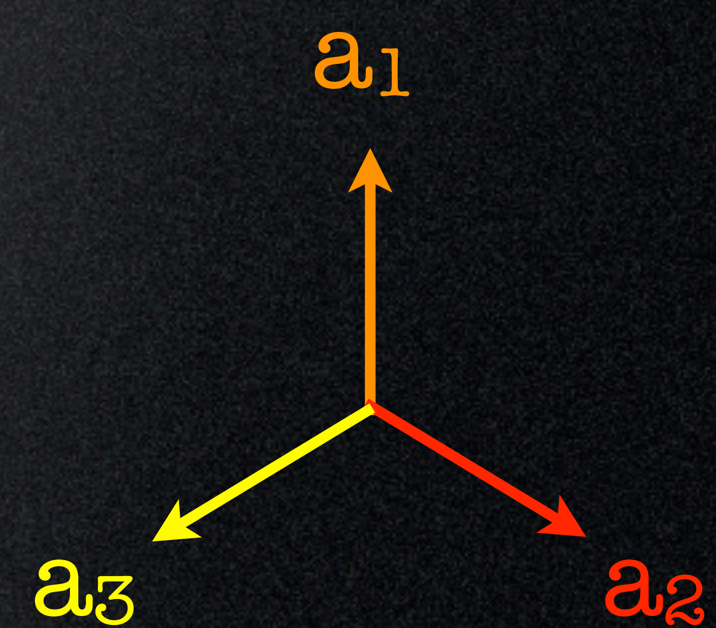
We pick the most economic solution, in terms of the resources taken up by synapses

$$min_{(L)} : \left[\underbrace{(D - DL)^2}_{\text{Reconstruction Error}} + \underbrace{\lambda |L|_1}_{\text{Sparsity}}\right]$$

$$D = DL$$

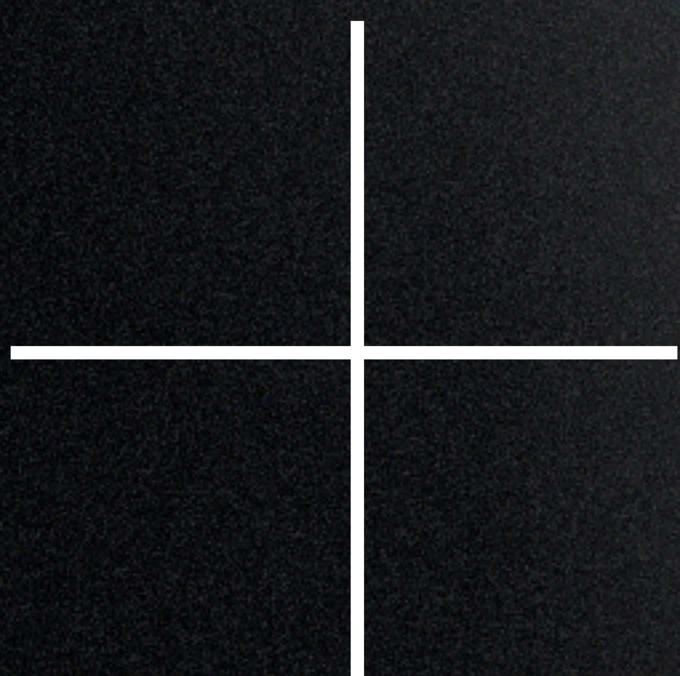
Neurons compensate for their changing representation by modifying post-synaptic neuronal activity

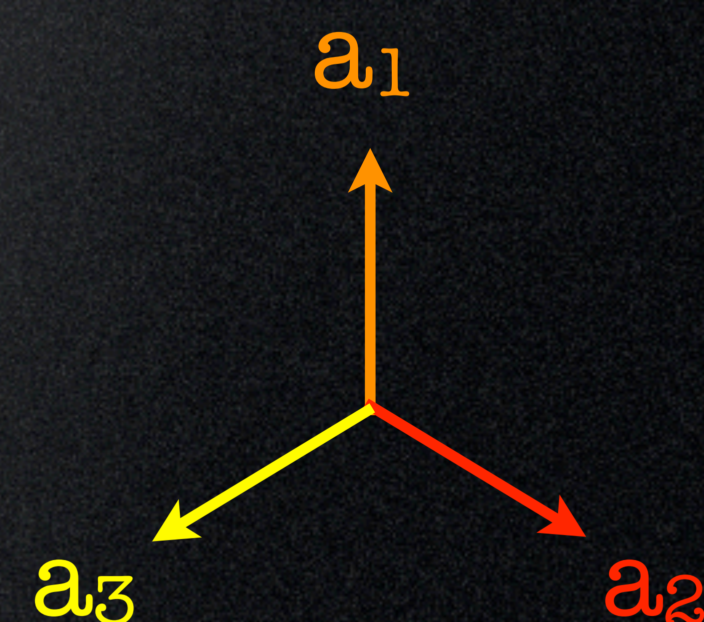




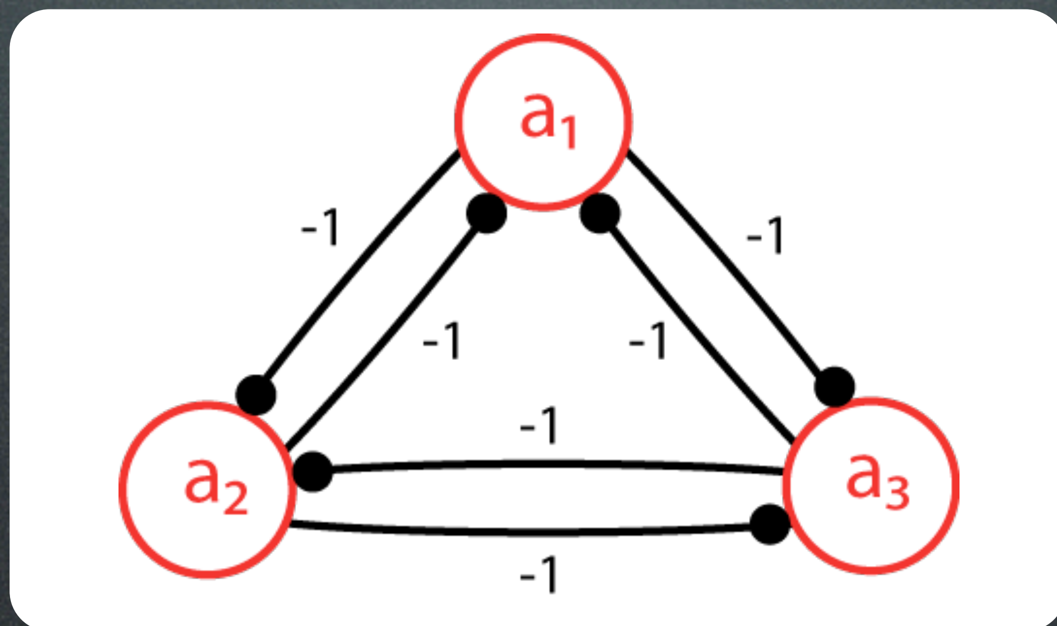
Neurons compensate for their changing representation by modifying post-synaptic neuronal activity

$$D = DL$$

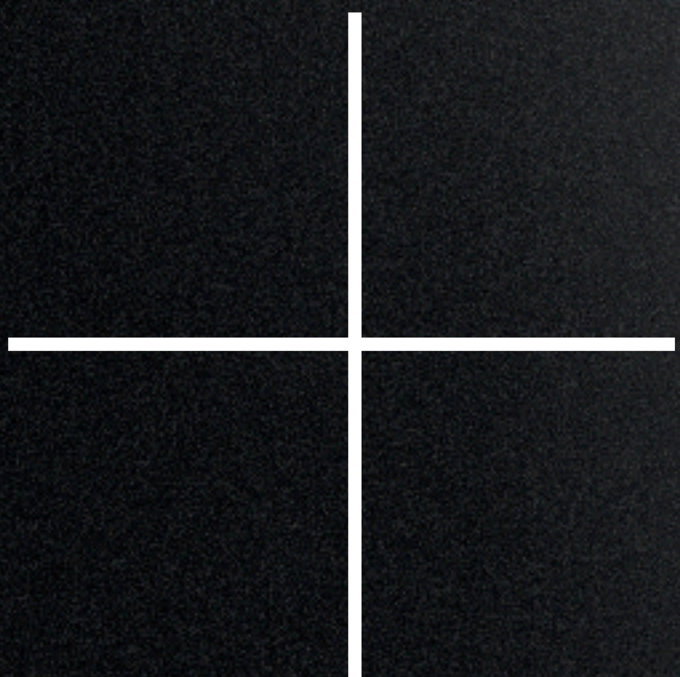


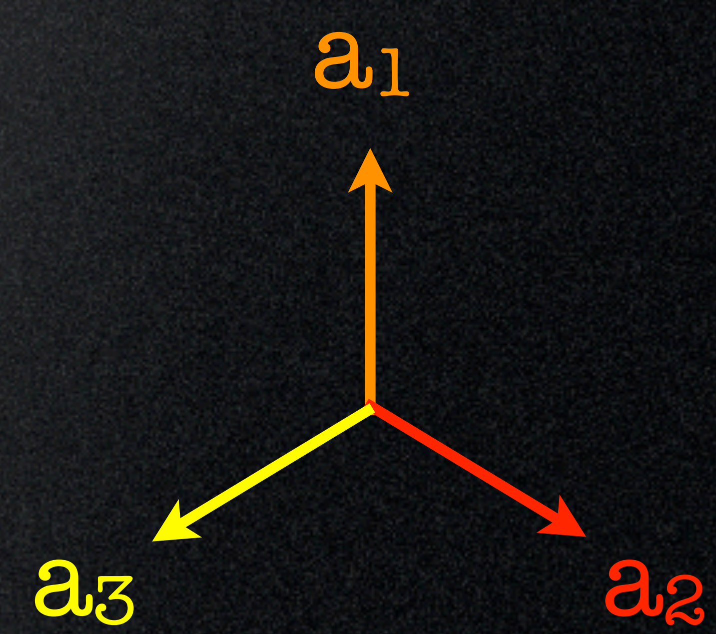


Neurons compensate for their changing representation by modifying post-synaptic neuronal activity

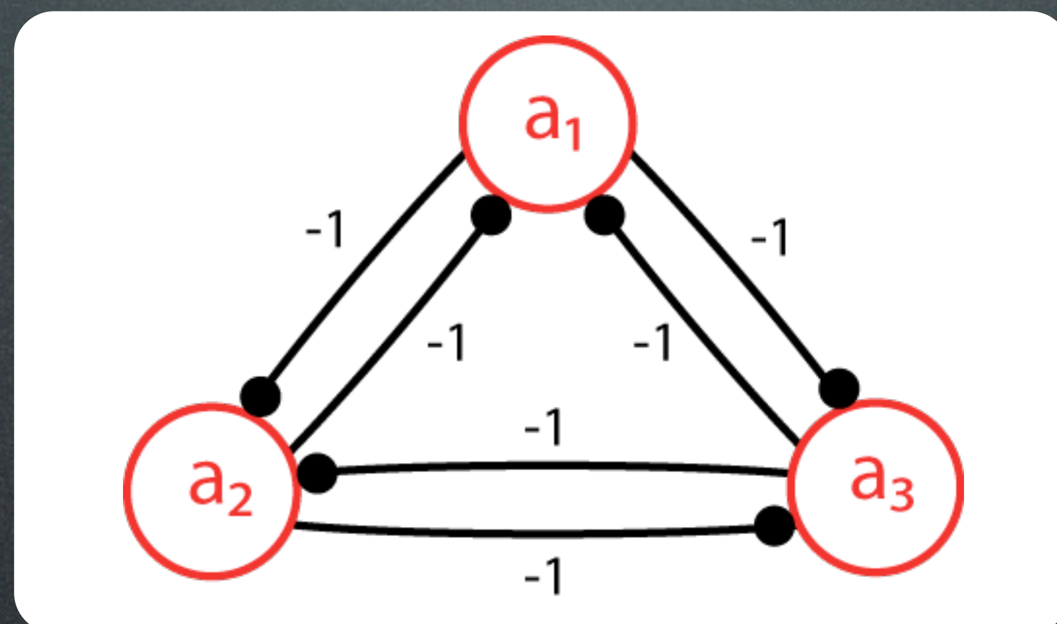


$$D = DL$$



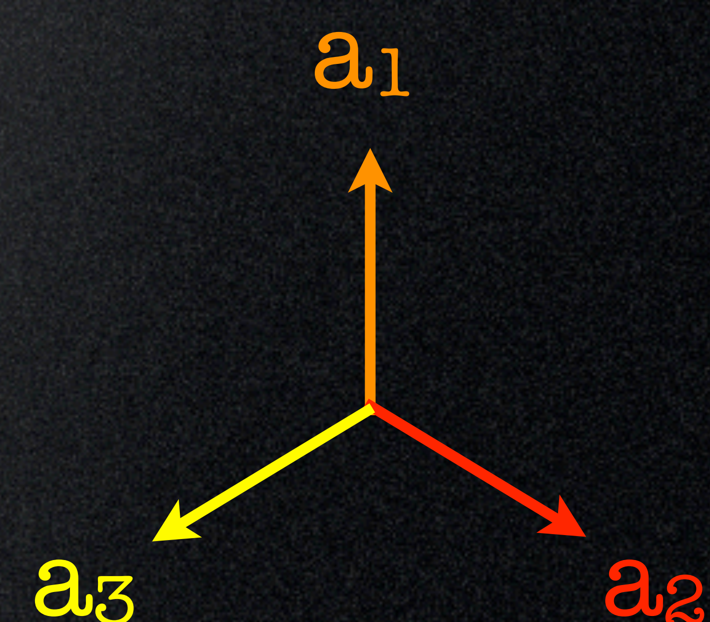


Neurons compensate for their changing representation by modifying post-synaptic neuronal activity

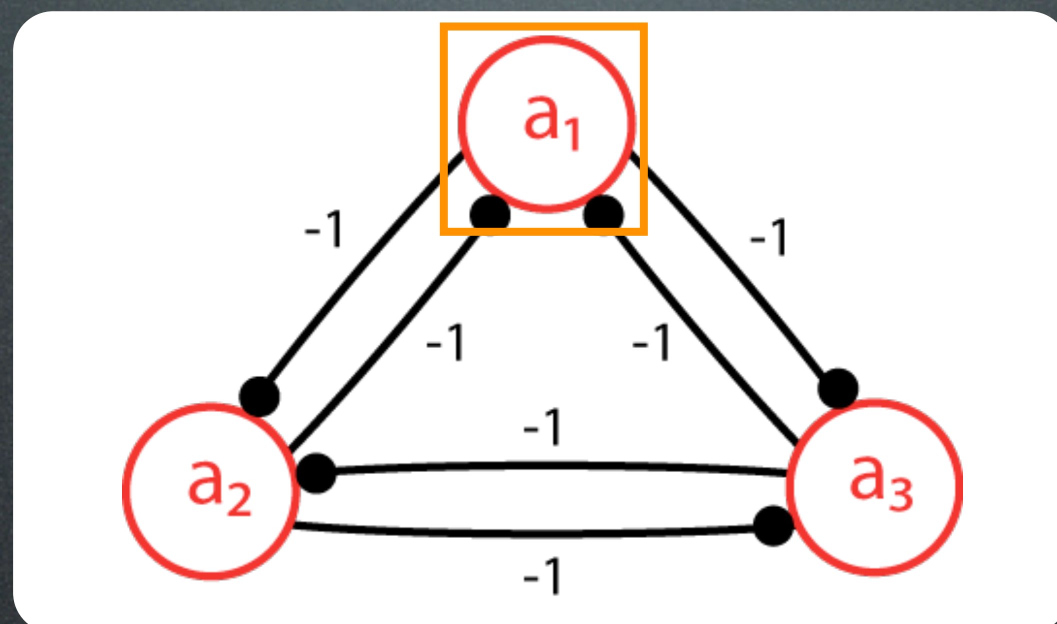


$$D = DL$$

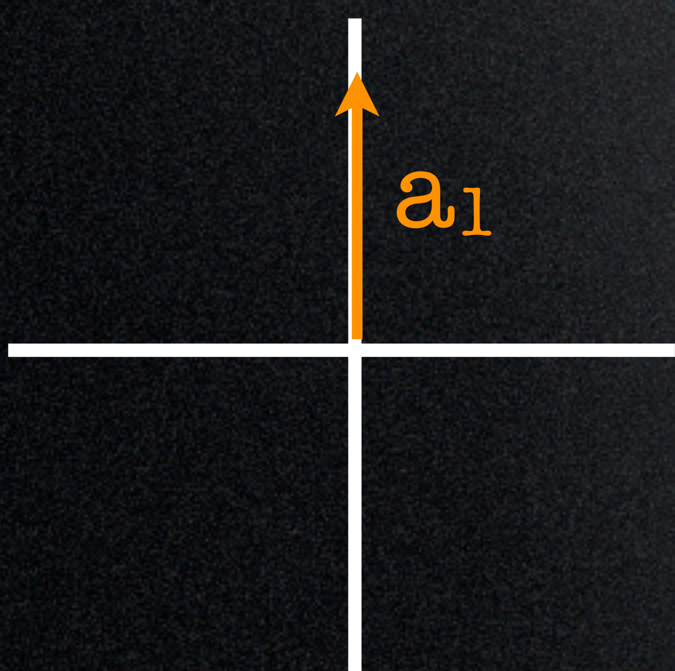
Sum of outgoing synapses times post-syn. receptive field equals neuron's own receptive field



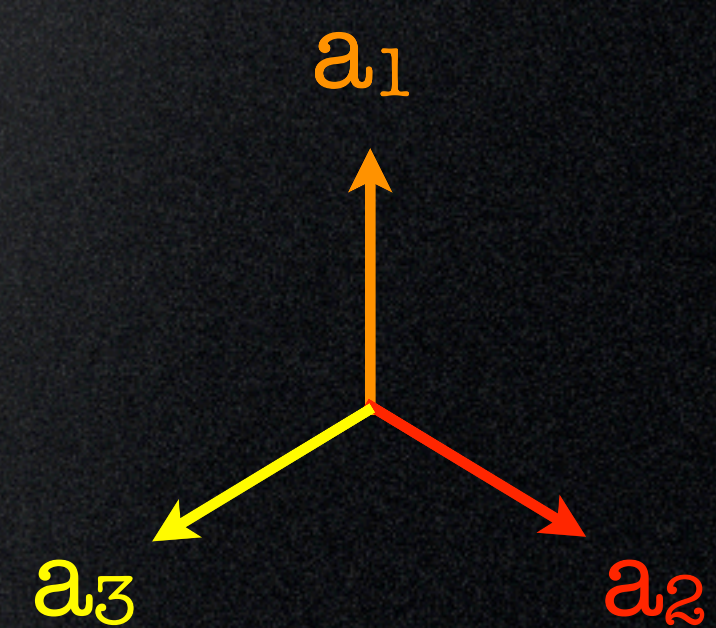
Neurons compensate for their changing representation by modifying post-synaptic neuronal activity



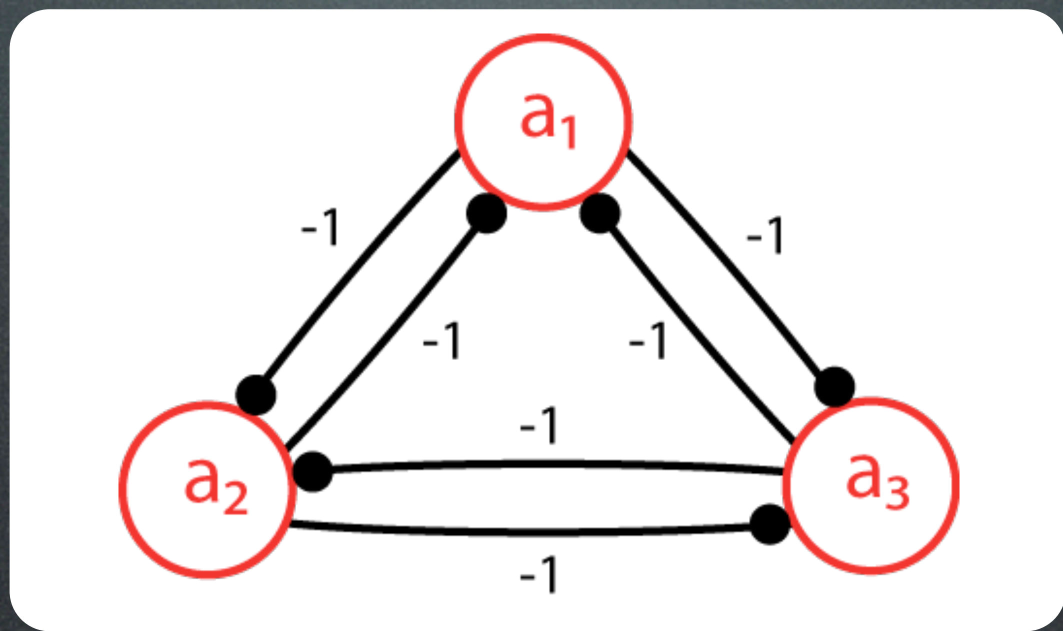
$$D = DL$$



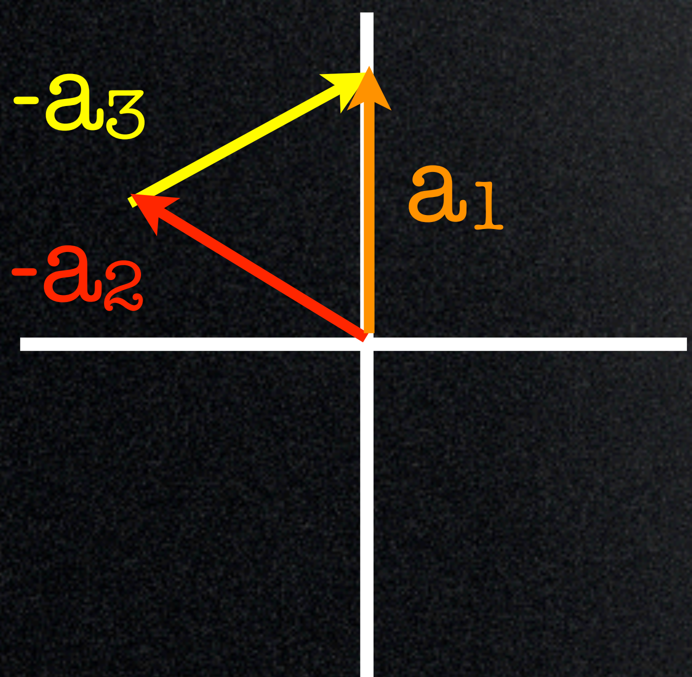
Sum of outgoing synapses times post-syn. receptive field equals neuron's own receptive field



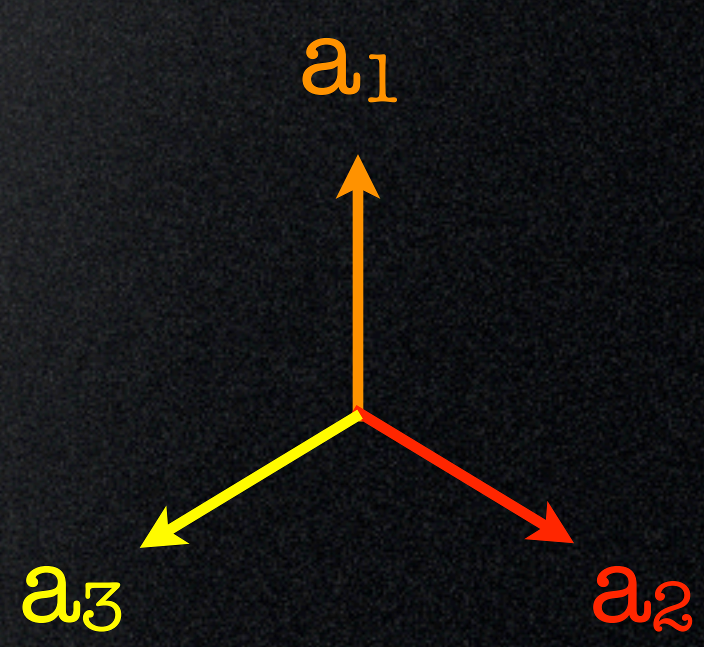
Neurons compensate for their changing representation by modifying post-synaptic neuronal activity



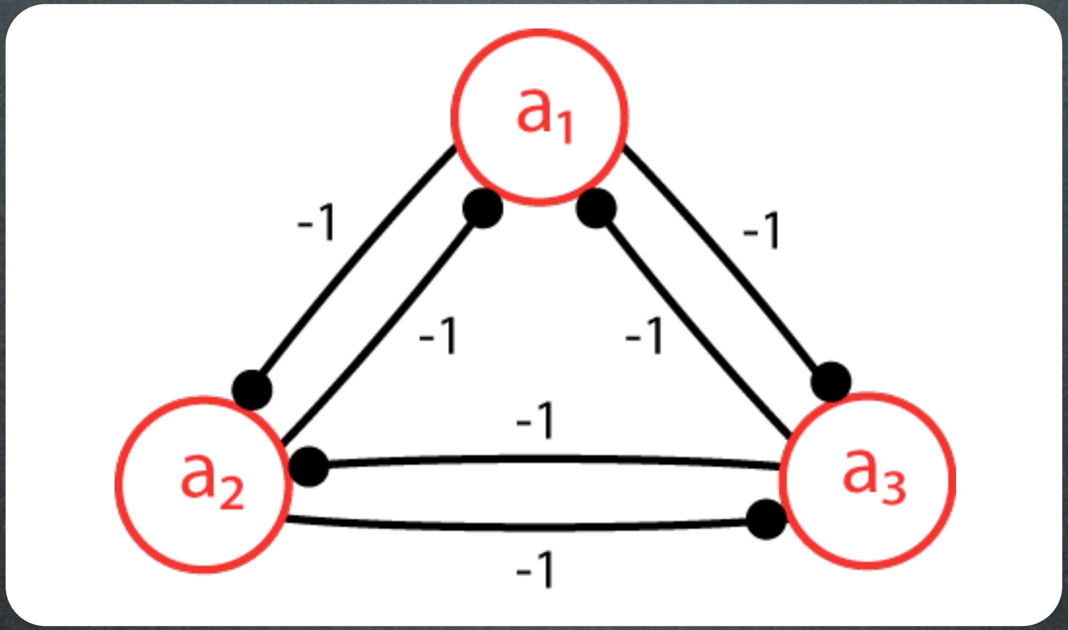
$$D = DL$$



Sum of outgoing synapses times post-syn. receptive field equals neuron's own receptive field



REceptive FIeld RE-combination (REFIRE) guarantees persistent percepts

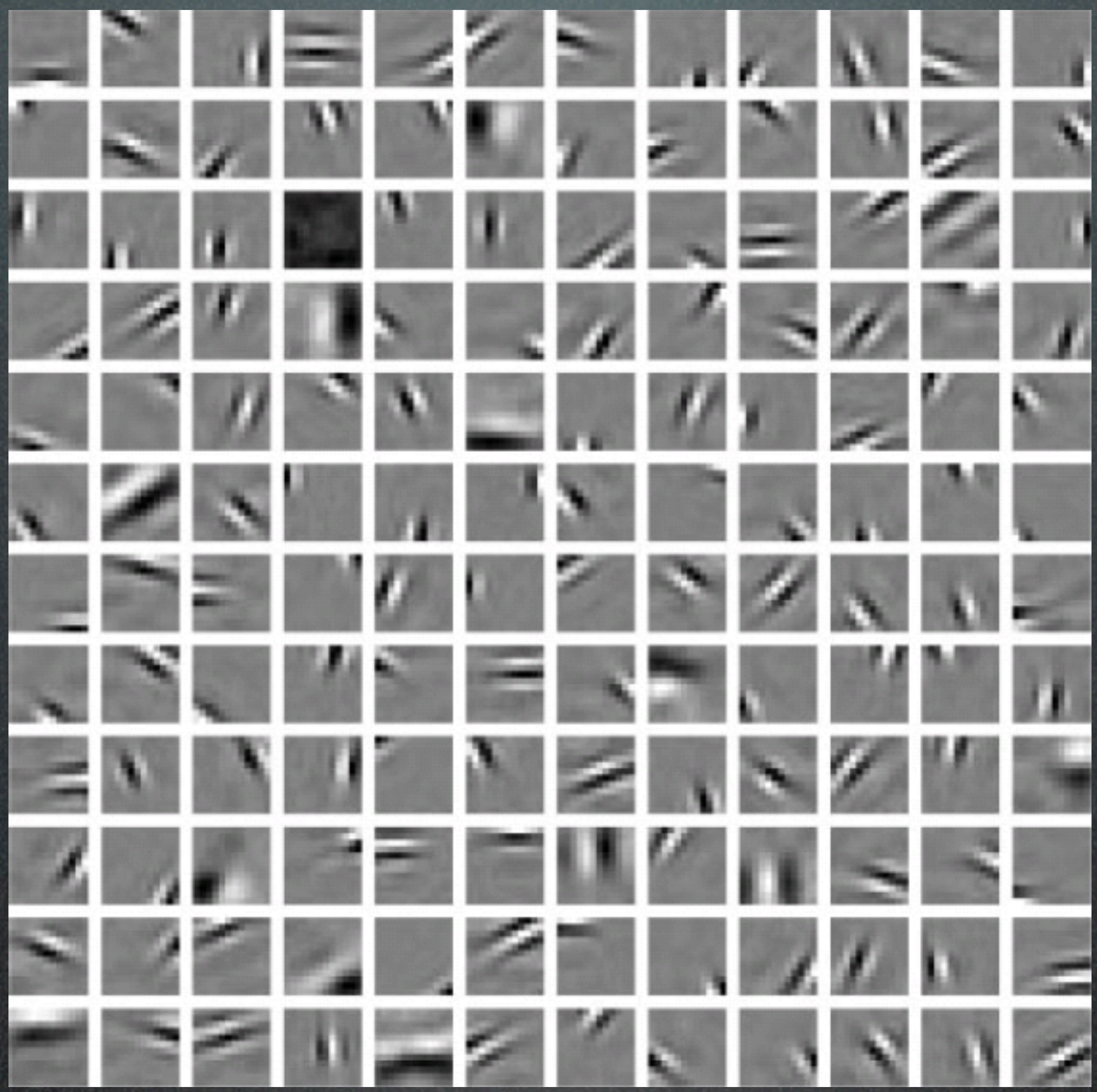


$$D = DL$$

Sum of outgoing synapses times post-syn. receptive field equals neuron's own receptive field

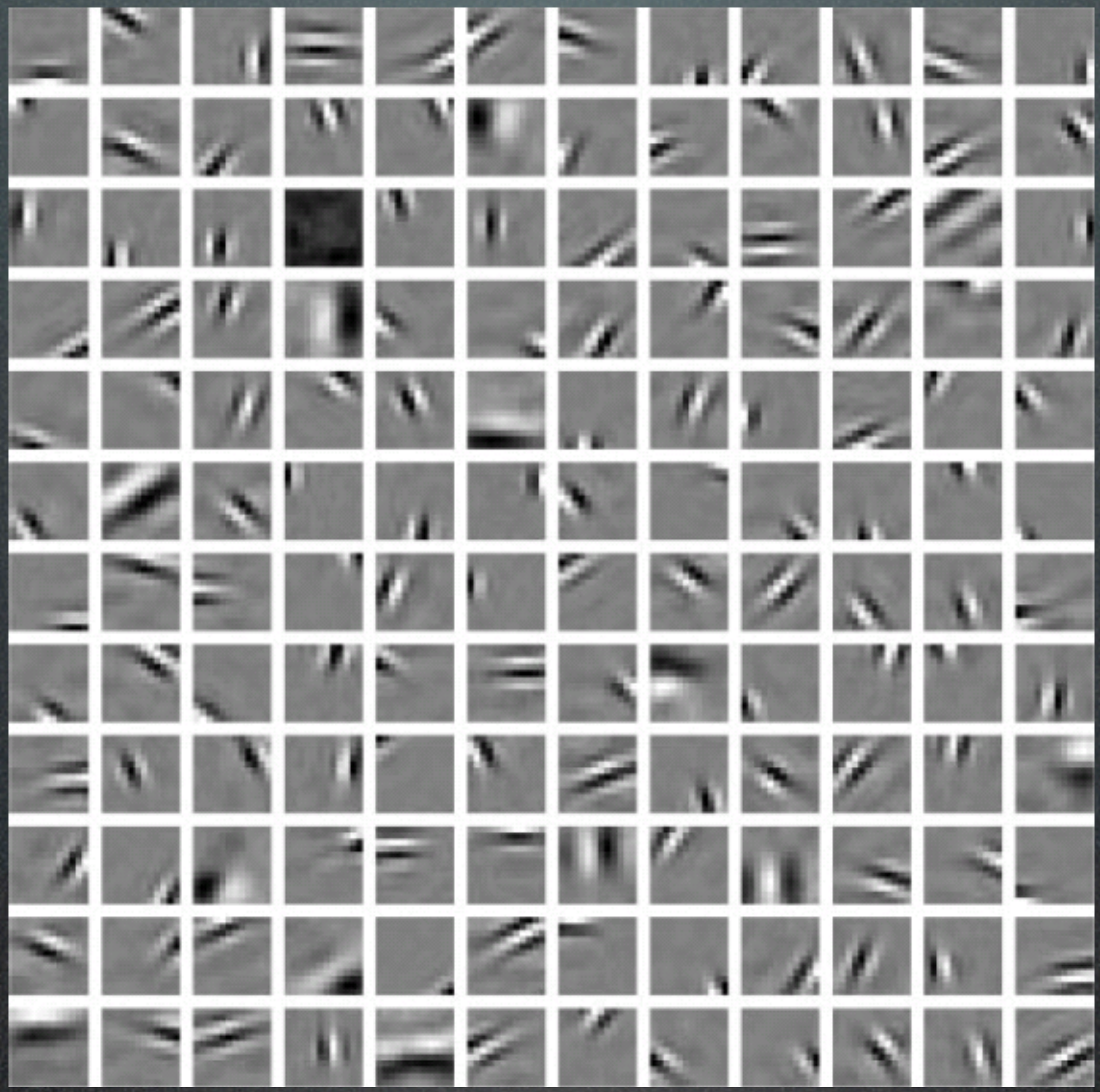
REFIRE with V1 receptive fields

REFIRE with V1 receptive fields



Receptive fields after:
Olshausen and Field, 1997

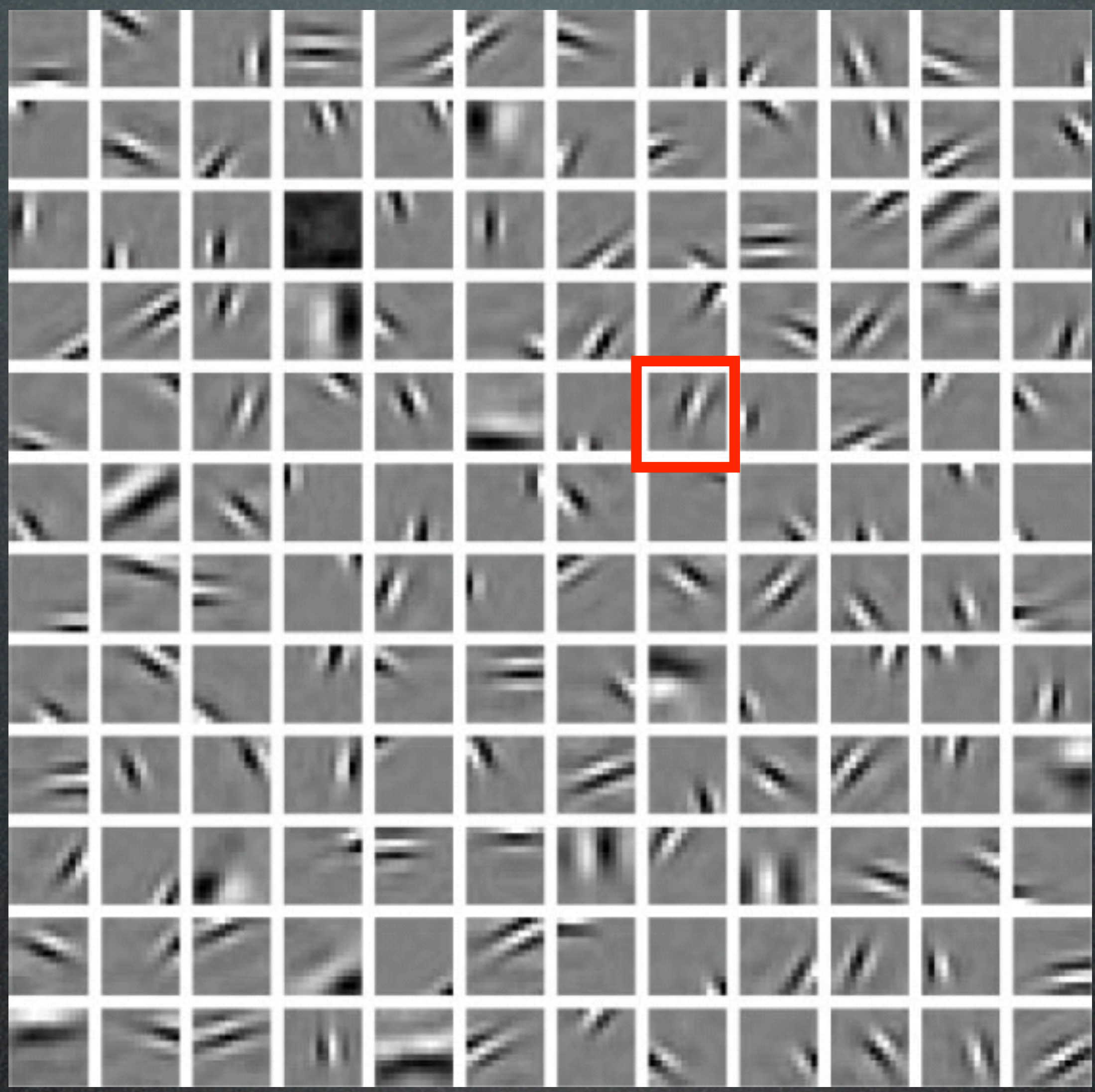
REFIRE with V1 receptive fields



$$D = DL$$

Receptive fields after:
Olshausen and Field, 1997

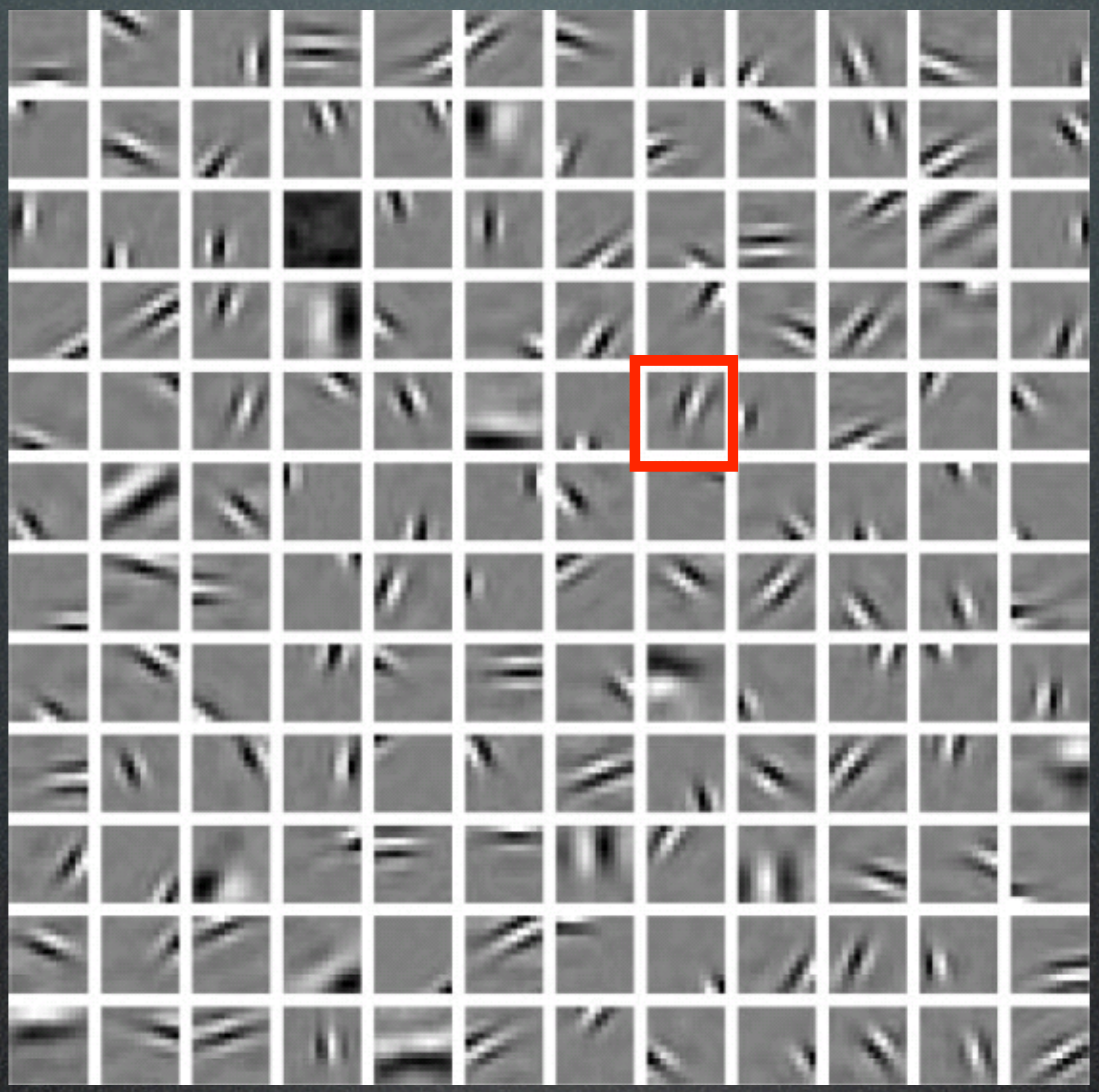
REFIRE with V1 receptive fields



$$D = DL$$

Receptive fields after:
Olshausen and Field, 1997

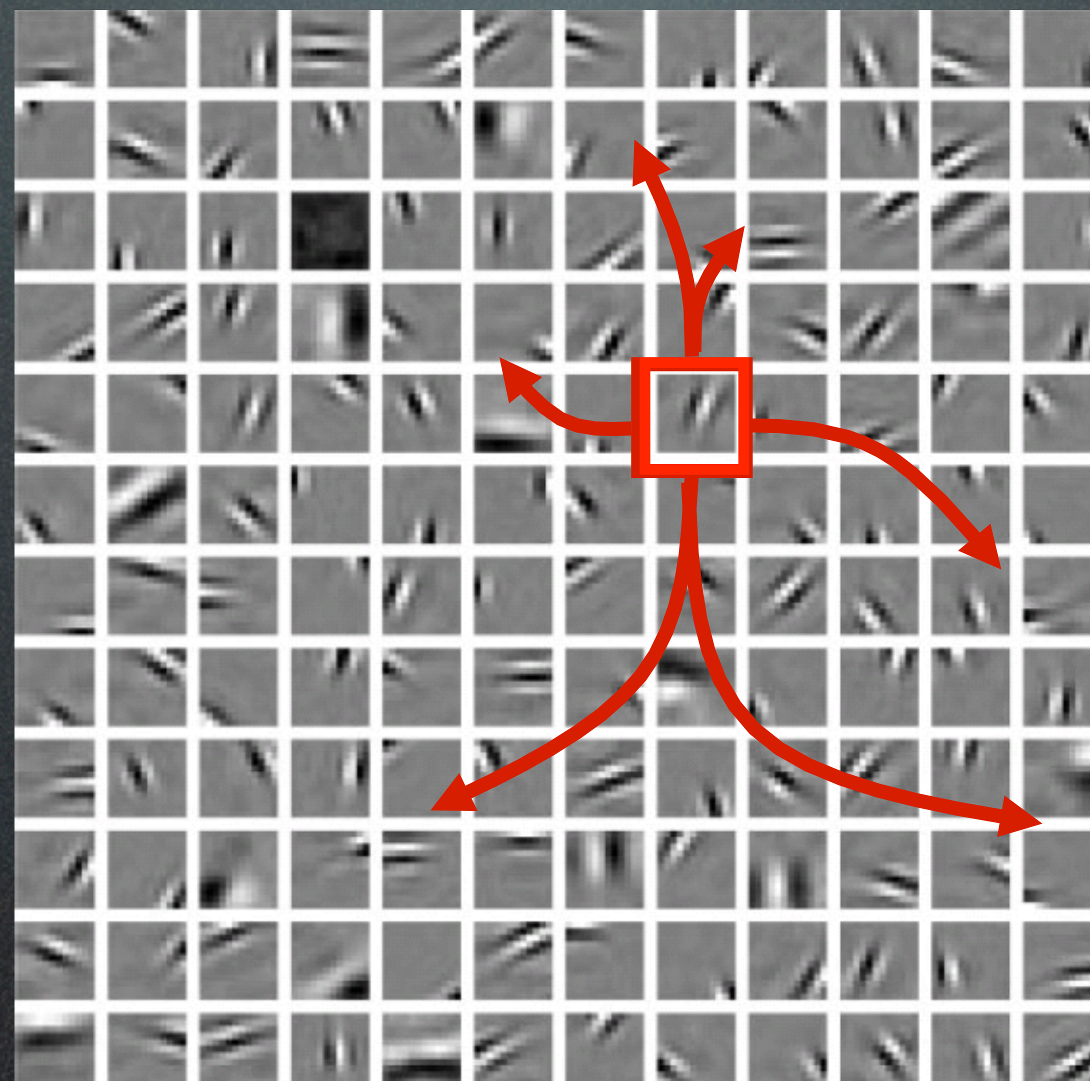
REFIRE with V1 receptive fields



Receptive fields after:
Olshausen and Field, 1997

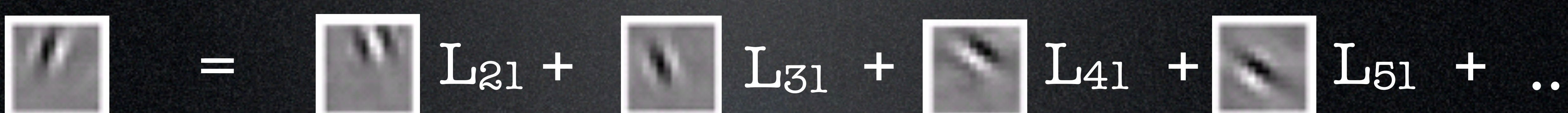
$$D = DL$$

REFIRE with V1 receptive fields

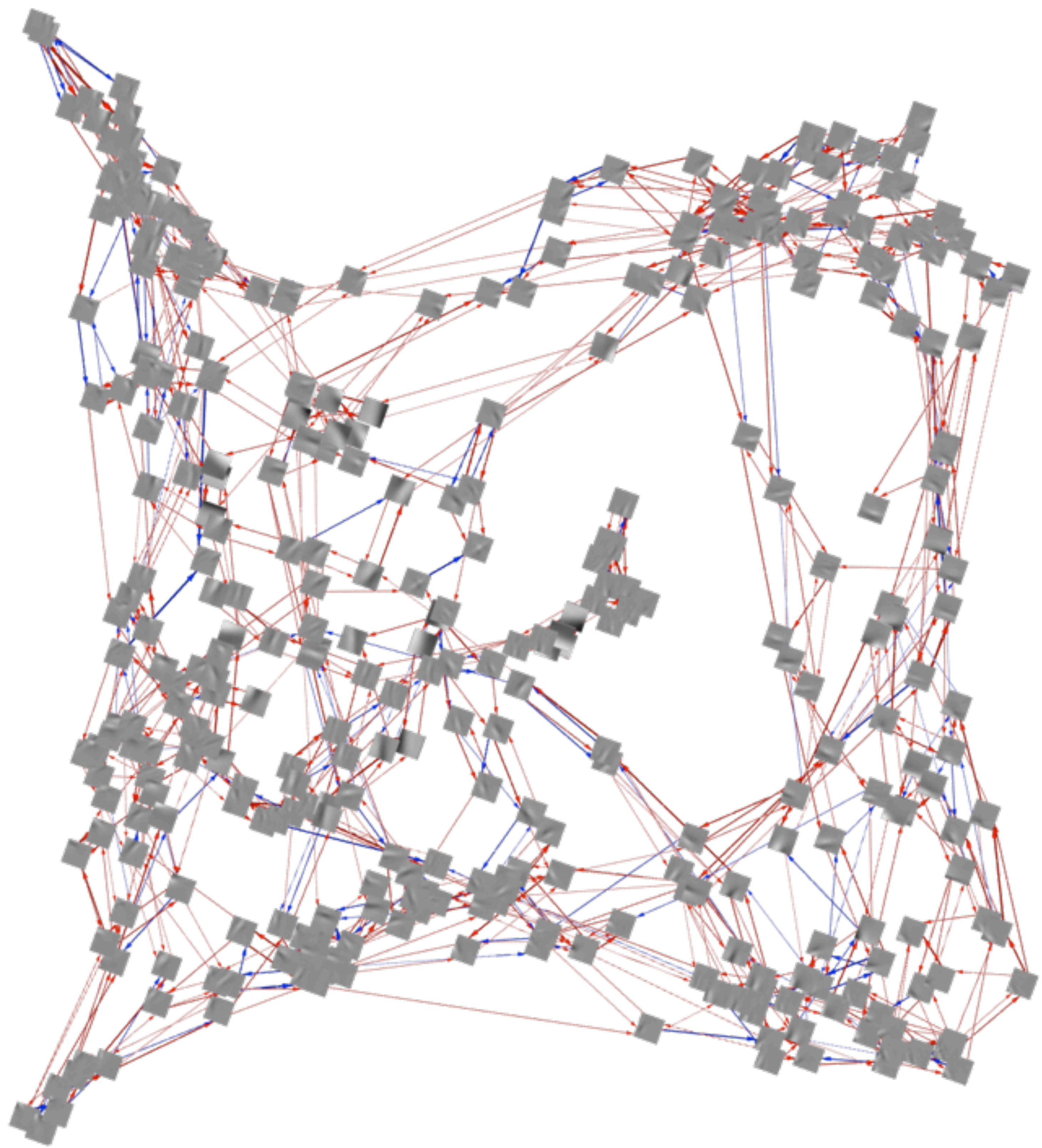


$$D = DL$$

Receptive fields after:
Olshausen and Field, 1997

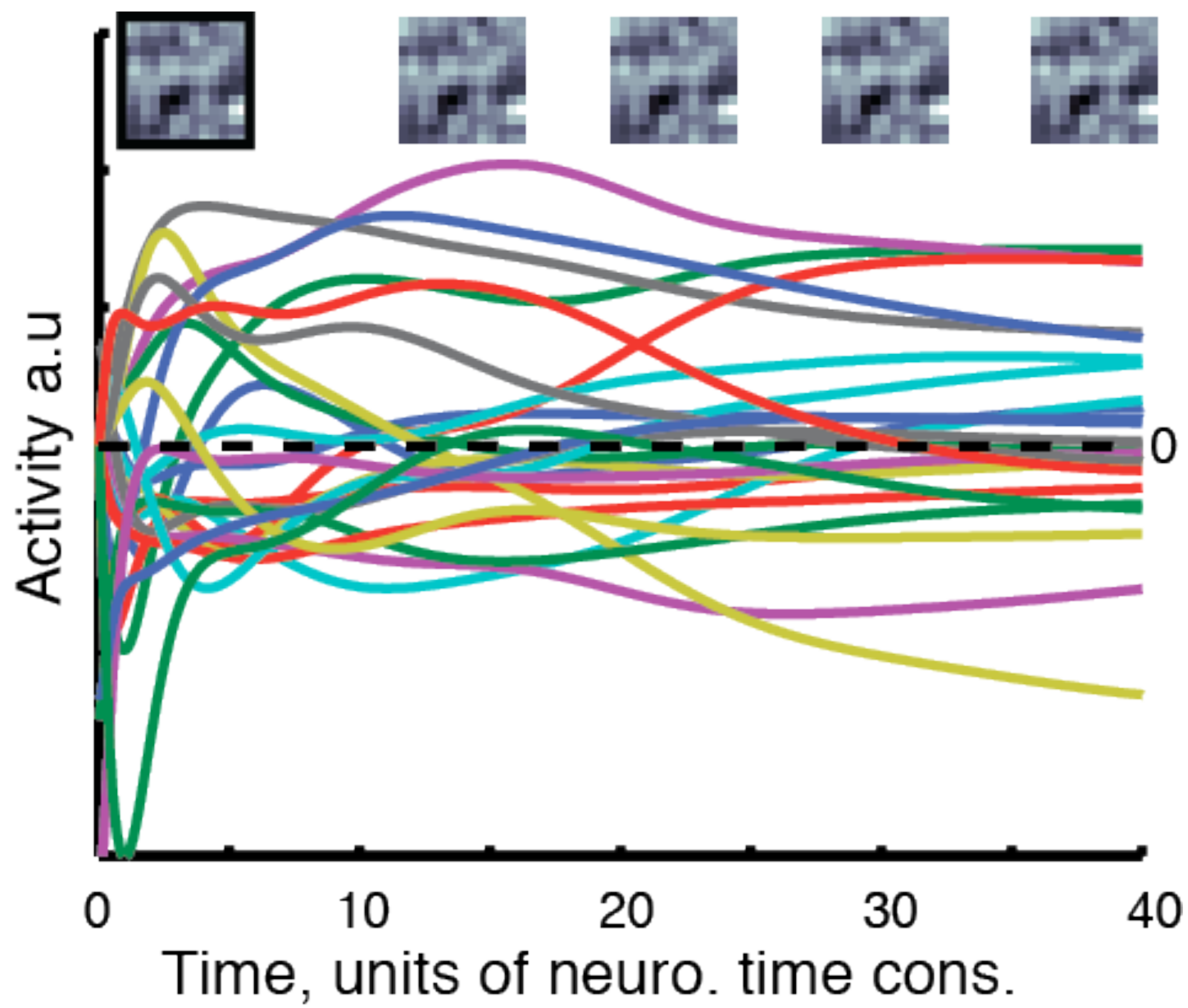


Network Structure



Network Structure

Numerical validation: Percepts are Persistent



Neuron 1

Neuron 2

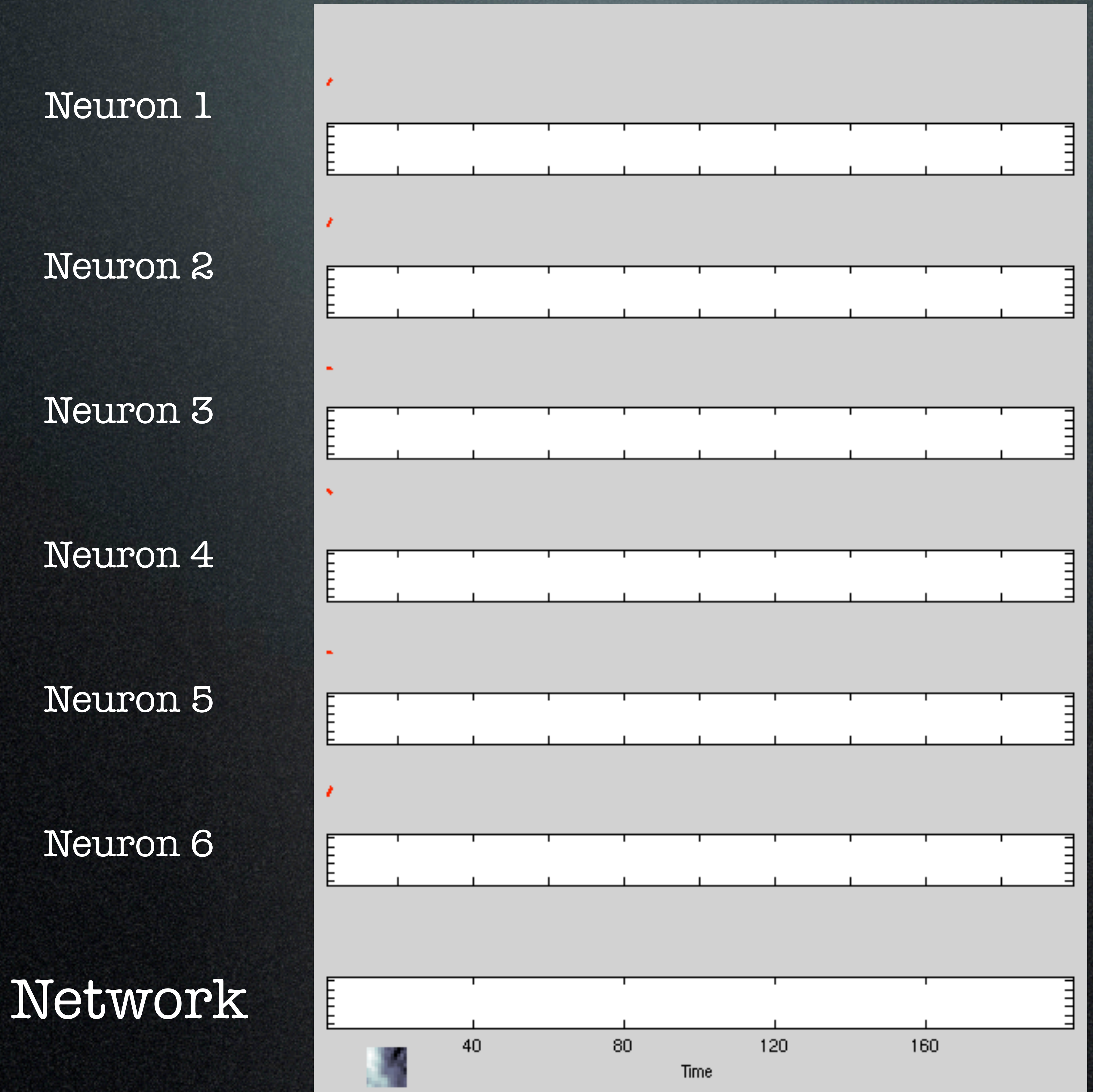
Neuron 3

Neuron 4

Neuron 5

Neuron 6

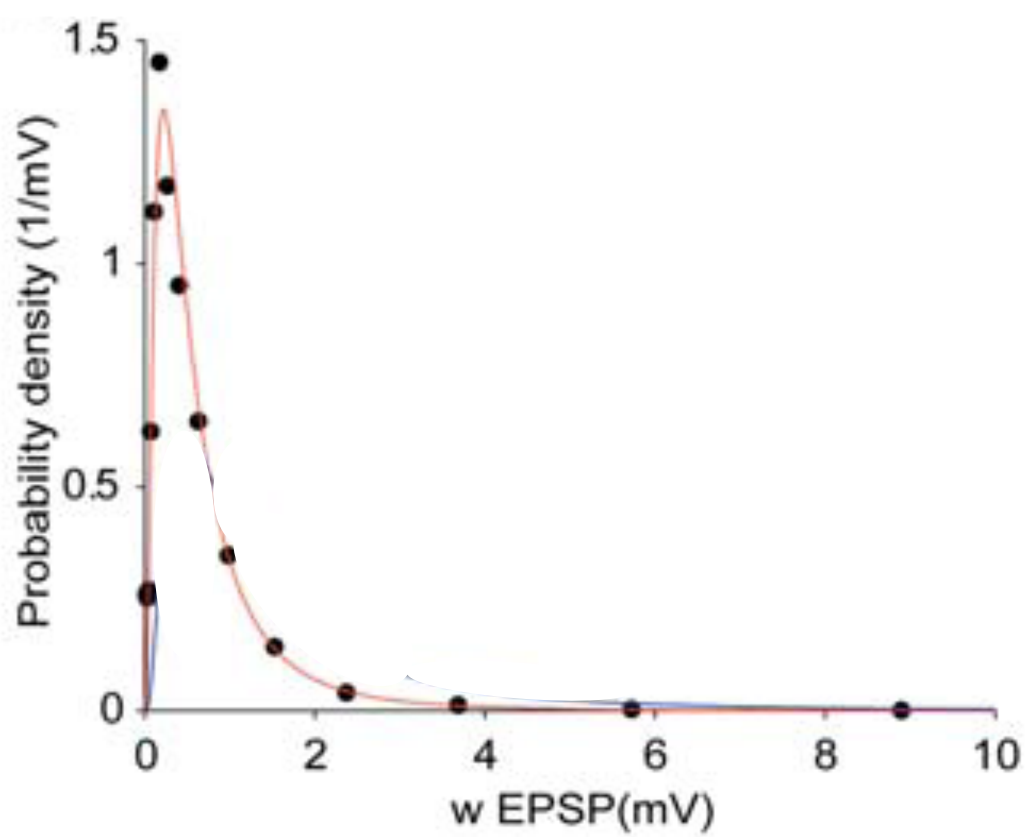
Network


Neuron 1
Neuron 2
Neuron 3
Neuron 4
Neuron 5
Neuron 6
Network
40
80
120
160
Time

Synaptic weight distribution matches Experiments

Synaptic weight distribution matches Experiments

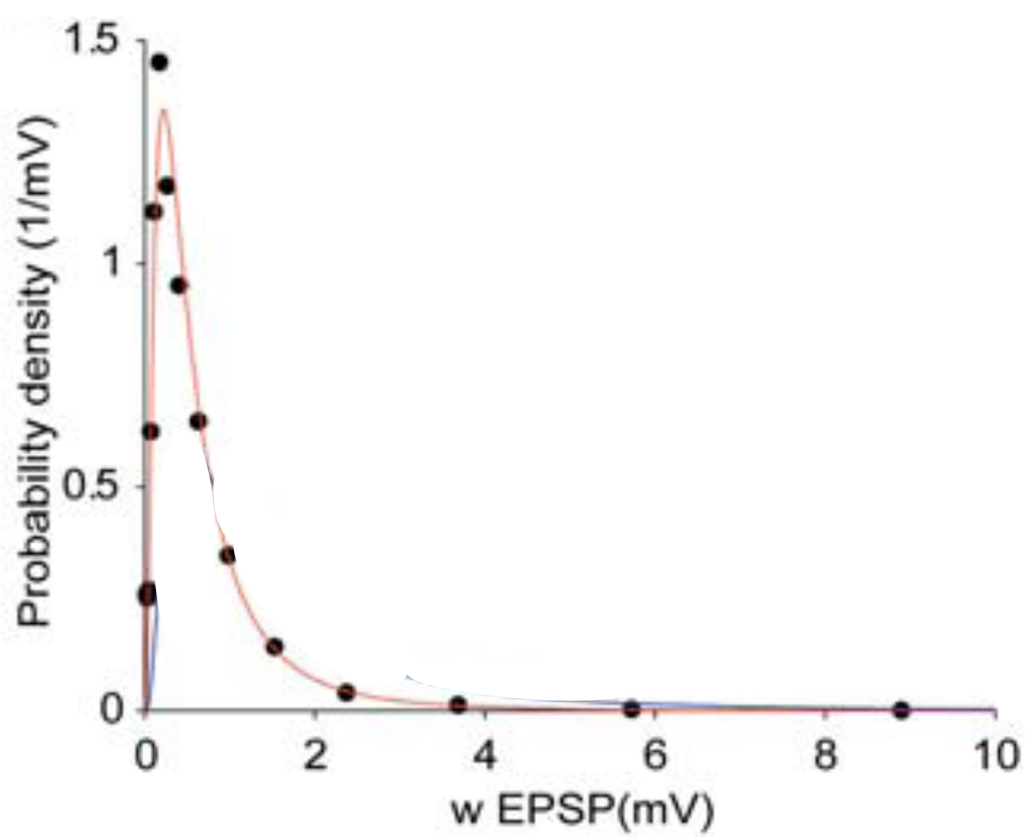
Experiment



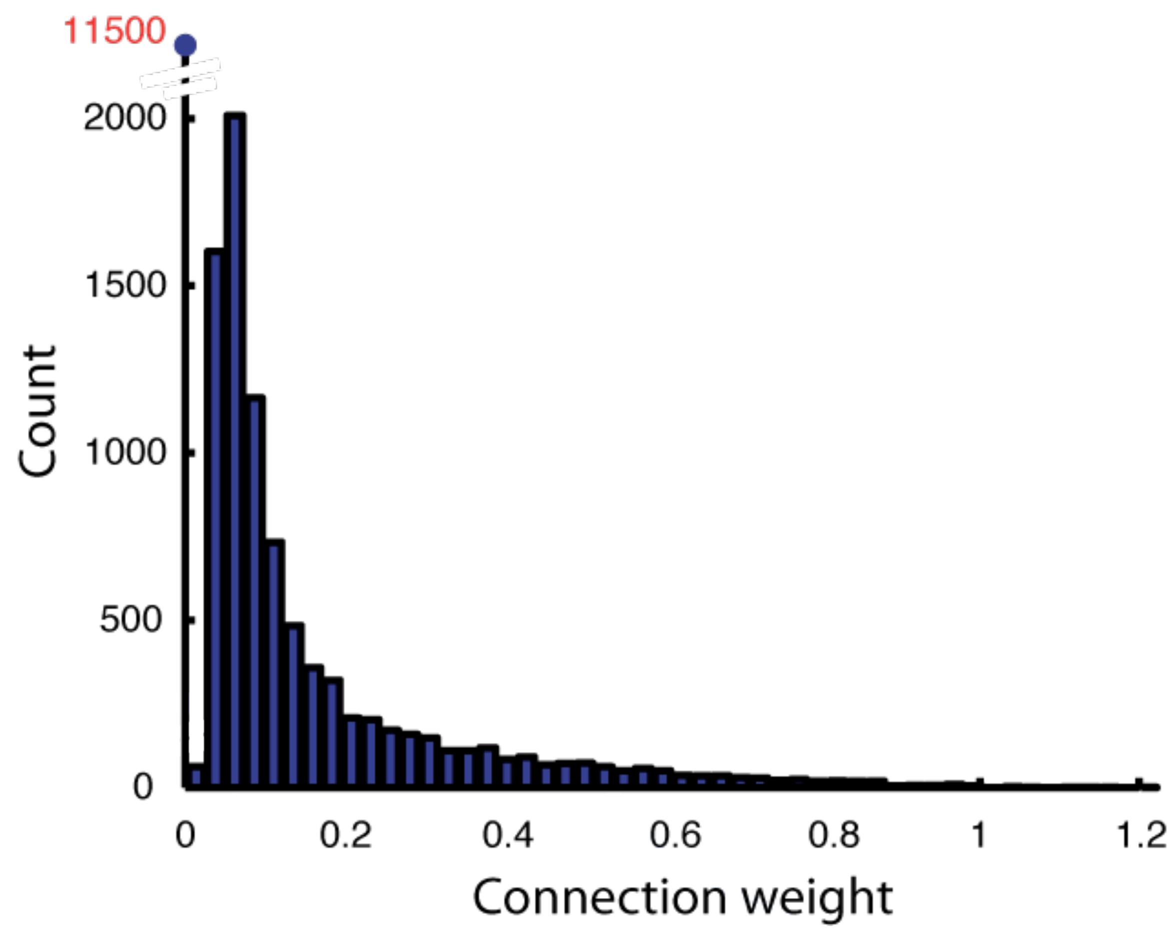
Song, Sjostrom, Reigl, Nelson, Chklovskii (2005)

Synaptic weight distribution matches Experiments

Experiment



Model



Song, Sjostrom, Reigl, Nelson, Chklovskii (2005)

Network Motifs match Experiments

Network Motifs match Experiments

2 Neuron motifs:

Network Motifs match Experiments

2 Neuron motifs:

Network Motifs match Experiments

2 Neuron motifs:

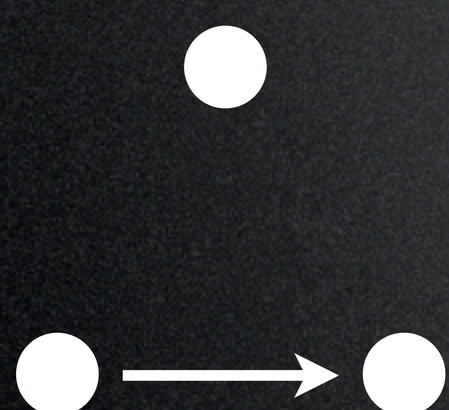
In cortex and in REFIRE network: $P_{recip} > p * p$

Network Motifs match Experiments

2 Neuron motifs:

In cortex and in REFIRE network: $P_{recip} > p * p$

3 Neuron motifs:

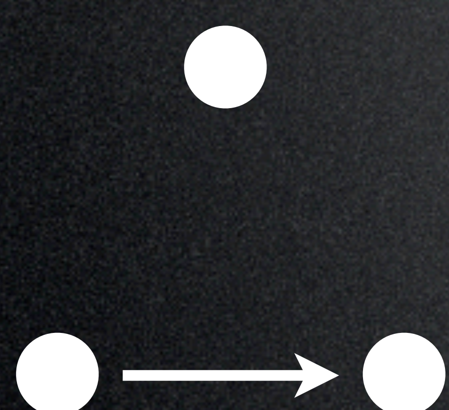


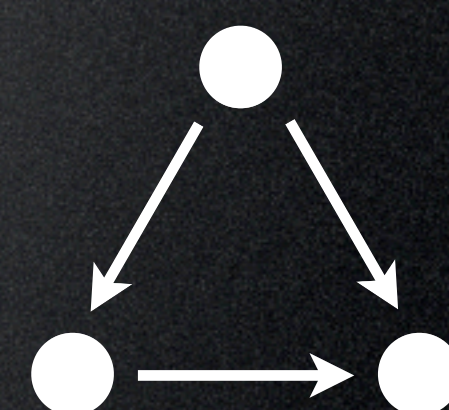
Network Motifs match Experiments

2 Neuron motifs:

In cortex and in REFIRE network: $P_{recip} > p * p$

3 Neuron motifs:



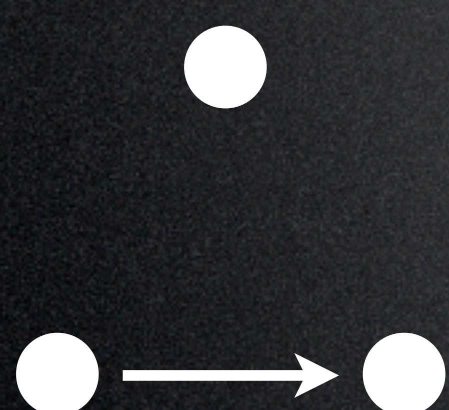


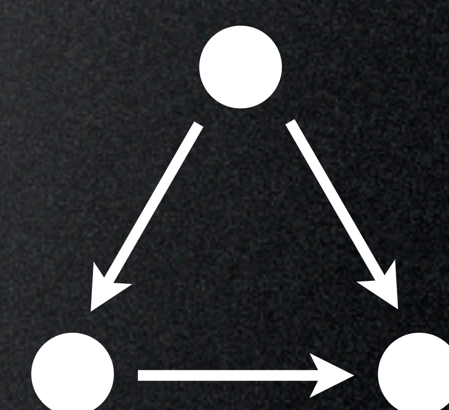
Network Motifs match Experiments

2 Neuron motifs:

In cortex and in REFIRE network: $P_{recip} > p * p$

3 Neuron motifs:





Over expression of reciprocal motifs

Summary

Summary

Can time-variant neuronal activity represent time invariant percepts?

Summary

Can time-variant neuronal activity represent time invariant percepts?

Yes. But not in an orthogonal basis!

Summary

Can time-variant neuronal activity represent time invariant percepts?

Yes. But not in an orthogonal basis!

We propose a specific form of a network that supports time invariant percepts with time-variant neuronal activity: REFIRE network

Summary

Can time-variant neuronal activity represent time invariant percepts?

Yes. But not in an orthogonal basis!

We propose a specific form of a network that supports time invariant percepts with time-variant neuronal activity: REFIRE network

This network qualitatively matches known statistical properties of cortical networks

Thanks

Thanks

Mitya Chklovskii

Thanks

Mitya Chklovskii

Shiv Vitaladevuni, Tao Hu, William Katz, Juan Nunez-Iglesias, Arjun Bharioke, Anatoli Grinspan and Lav Varshney

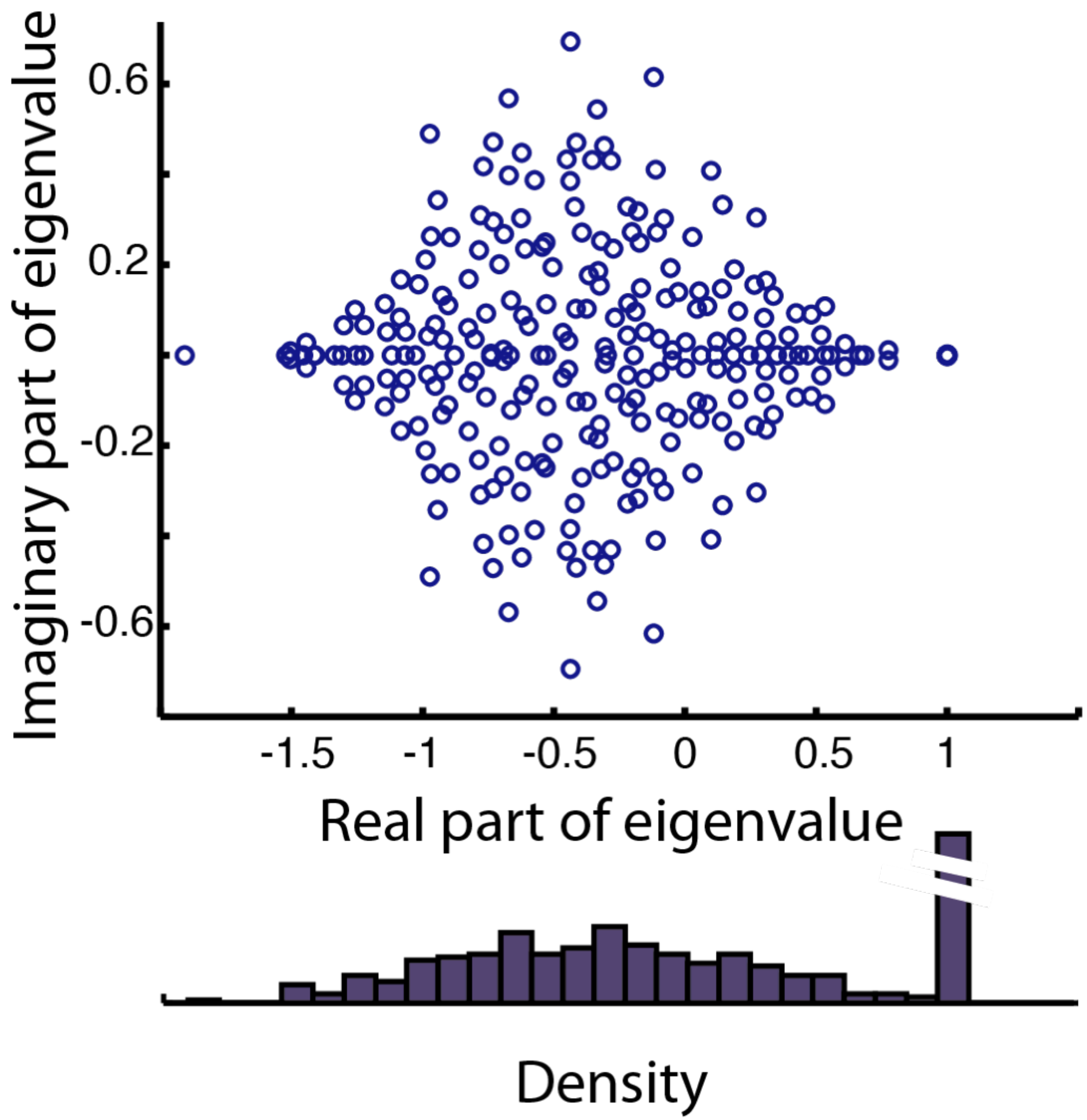
Frank Midgley

and thank you for your attention!

Poster: T9

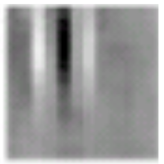
druckmanns@janelia.hhmi.org

Eigenvalues

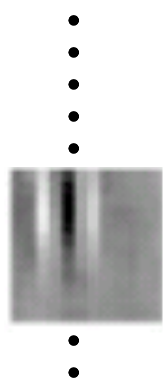


Similar Orientations, More Connections

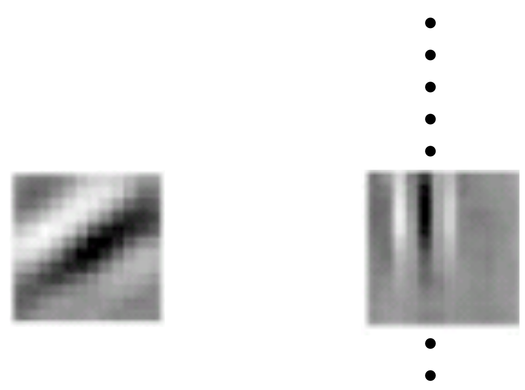
Similar Orientations, More Connections



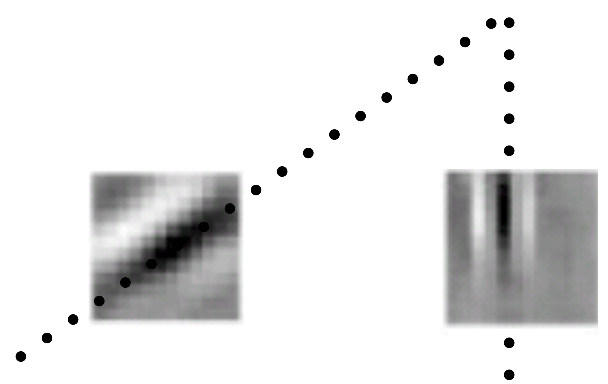
Similar Orientations, More Connections



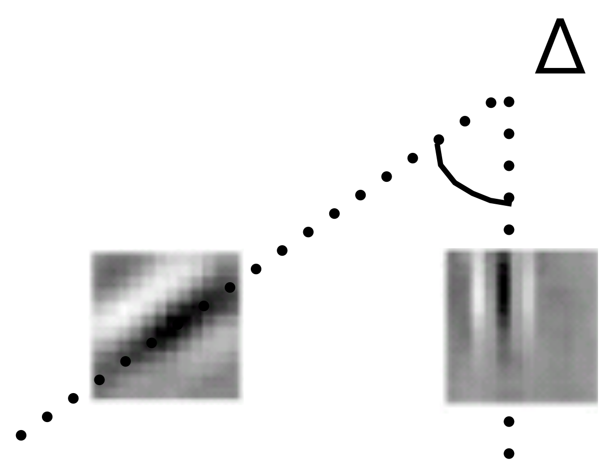
Similar Orientations, More Connections



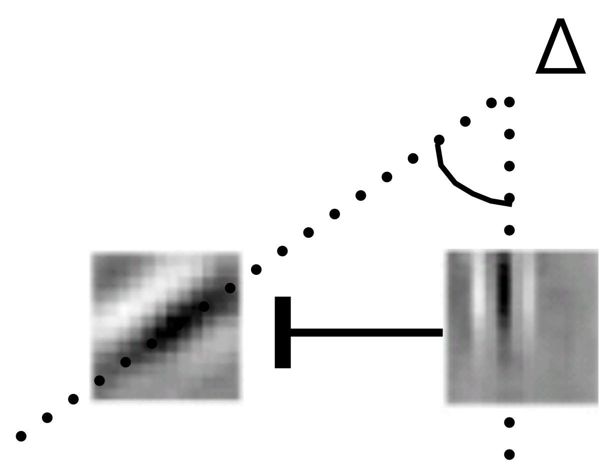
Similar Orientations, More Connections



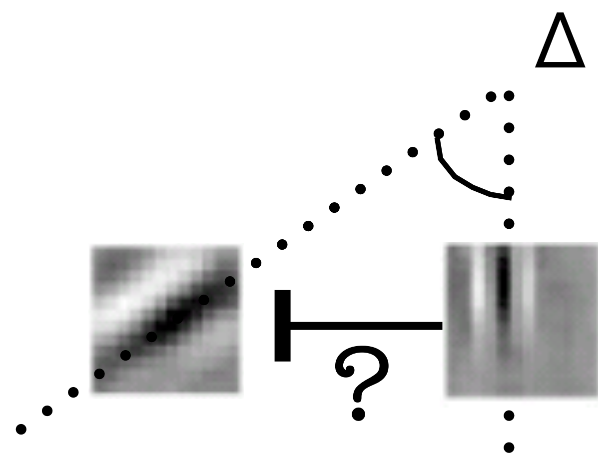
Similar Orientations, More Connections



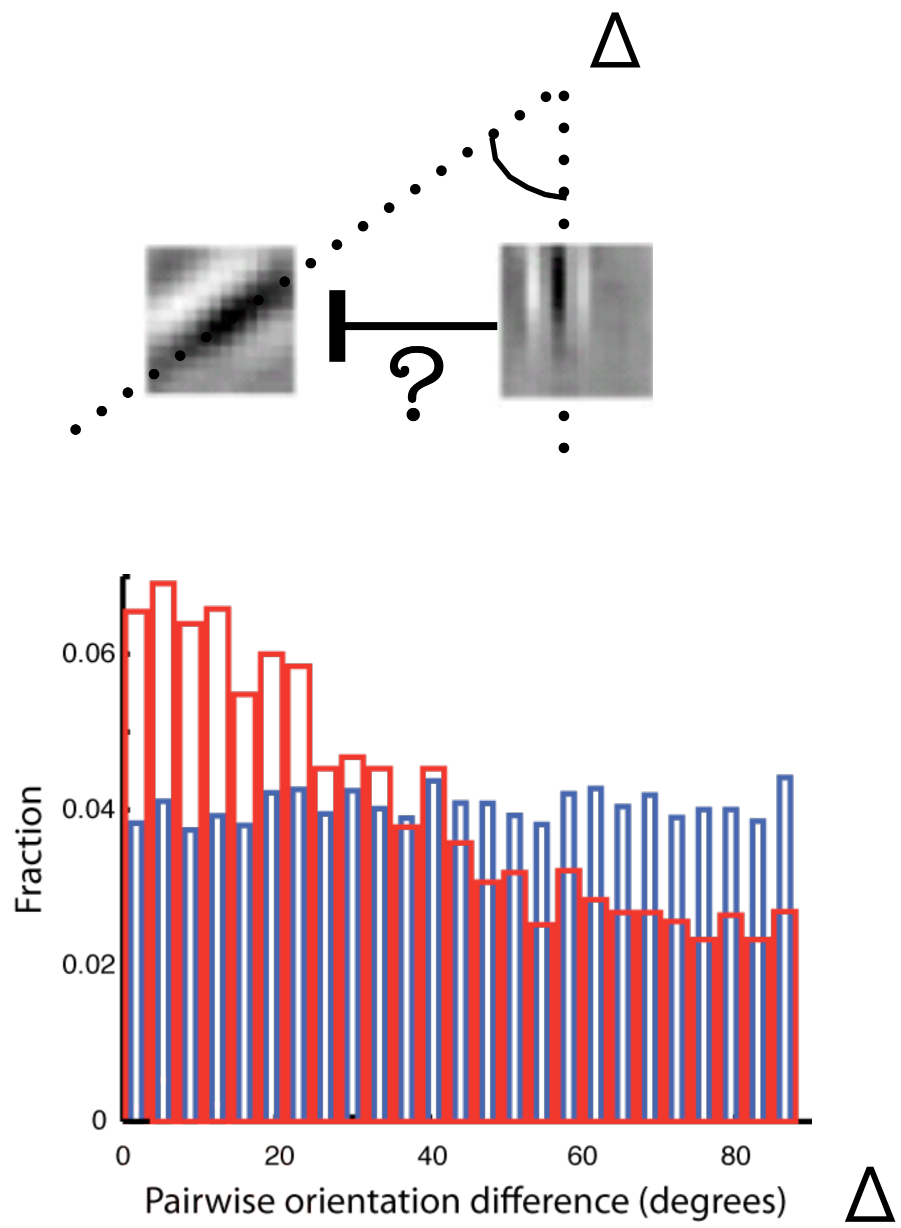
Similar Orientations, More Connections



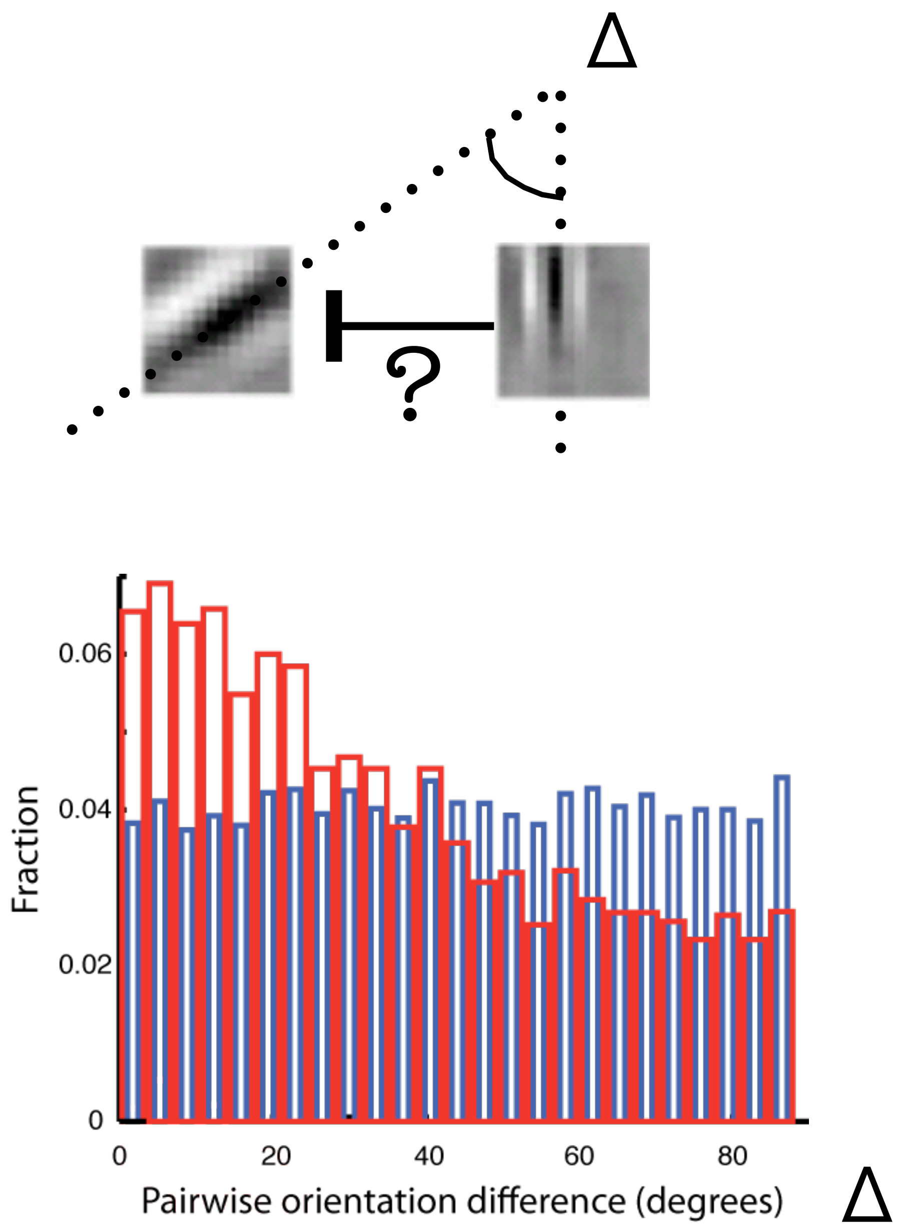
Similar Orientations, More Connections



Similar Orientations, More Connections

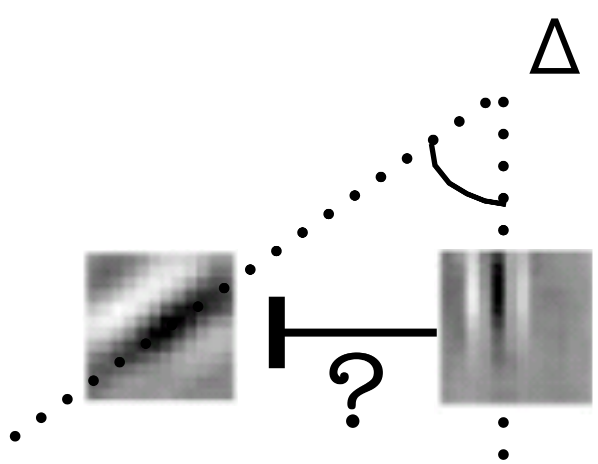


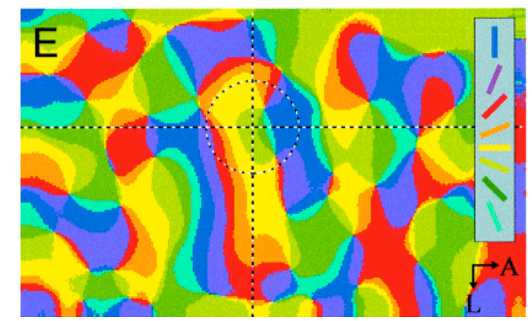
Similar Orientations, More Connections

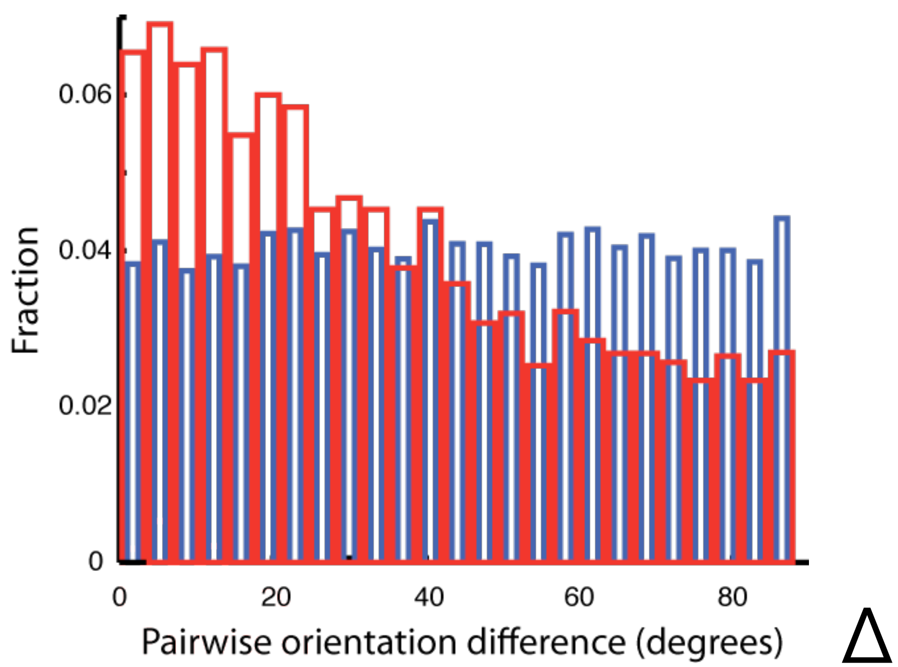


Youssef et al. (1999)

Similar Orientations, More Connections

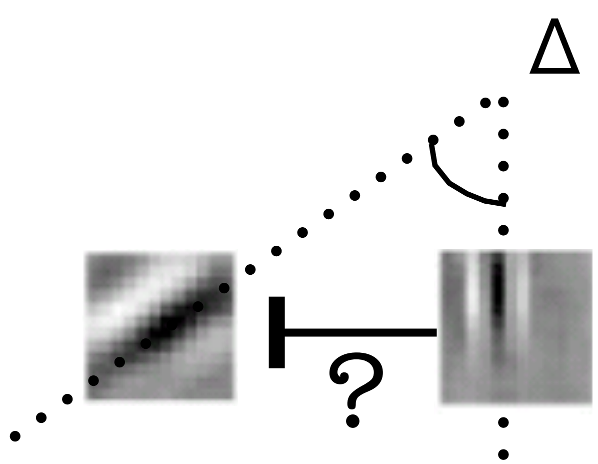


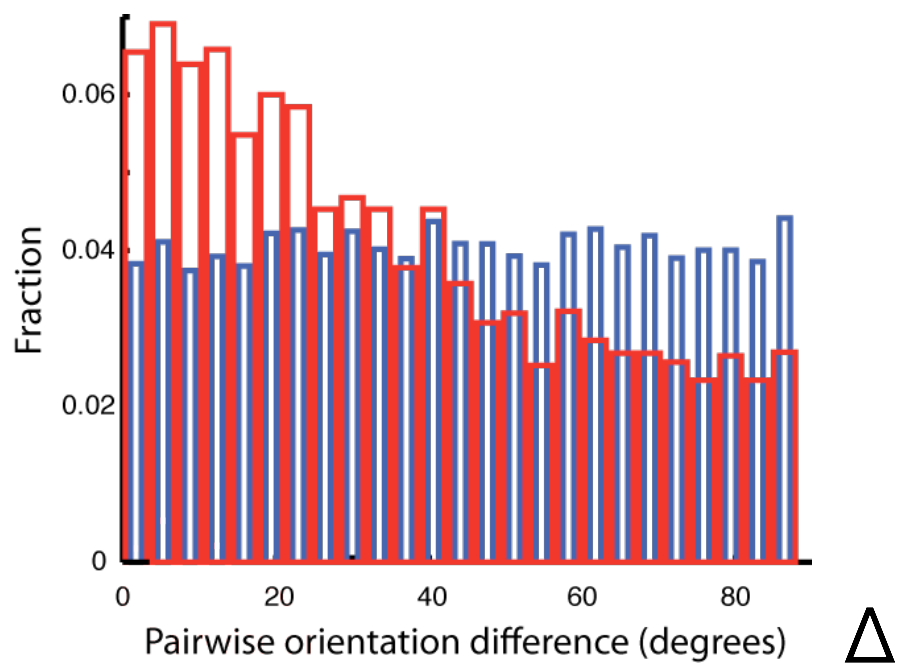


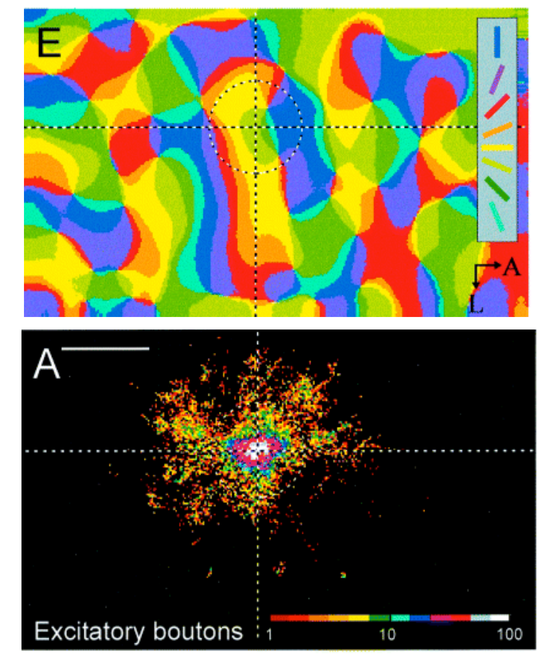


Youssef et al. (1999)

Similar Orientations, More Connections

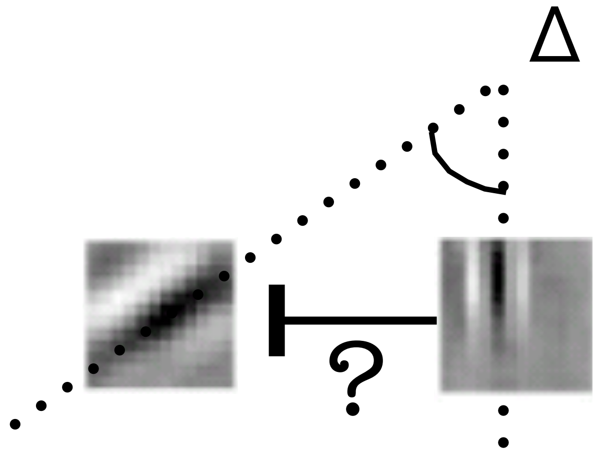


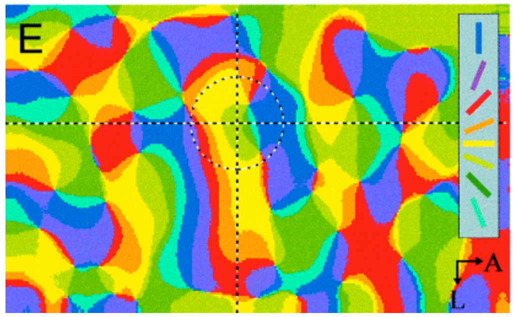


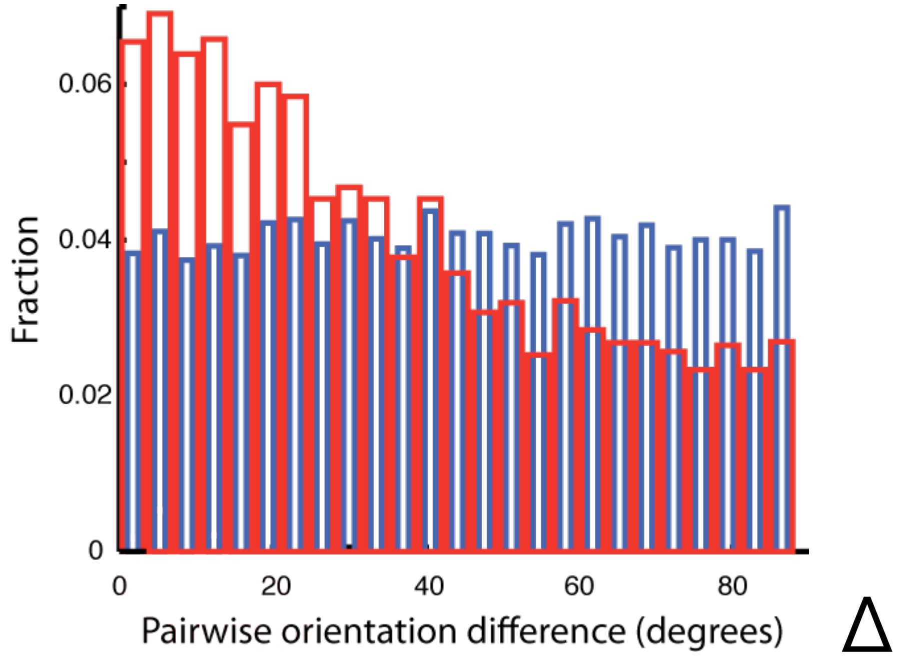


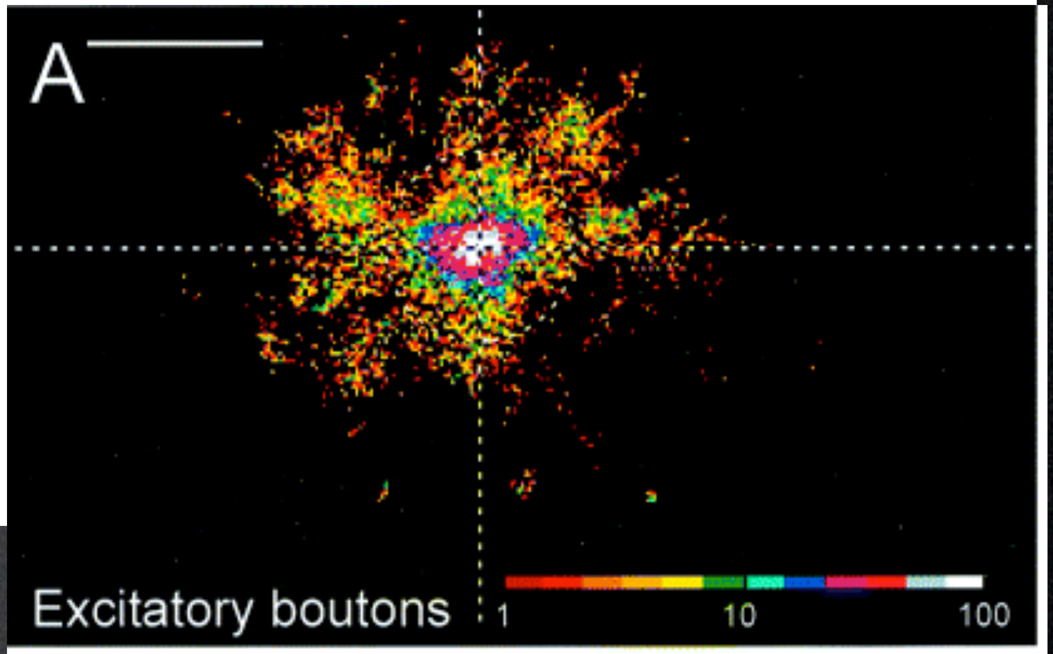
Youssef et al. (1999)

Similar Orientations, More Connections

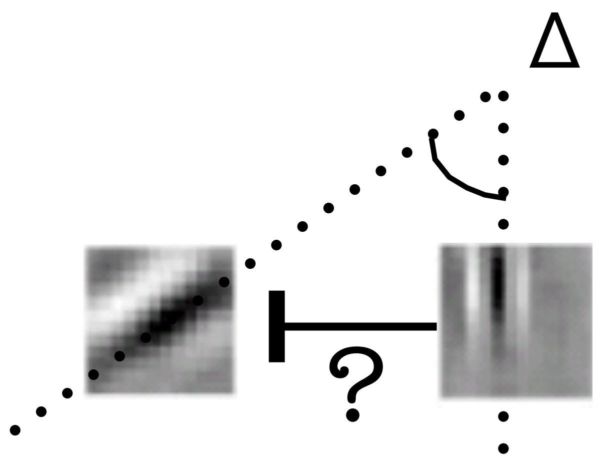


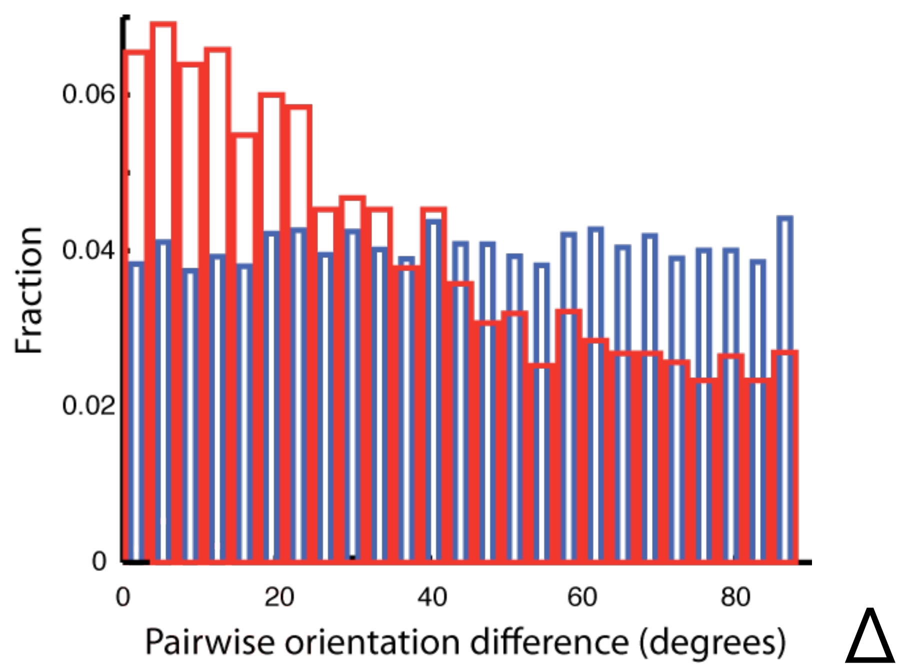


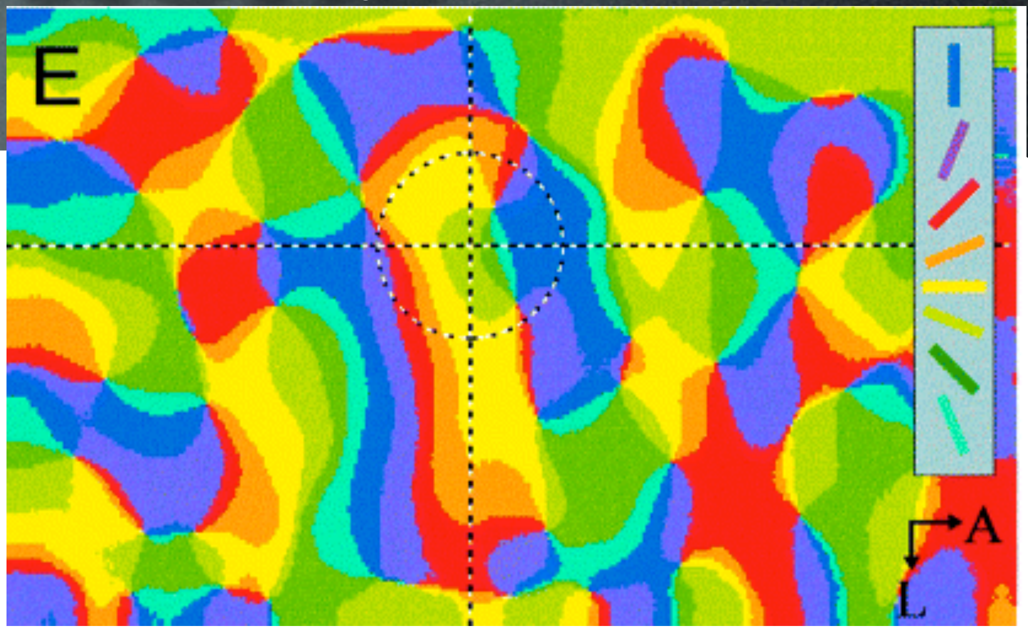


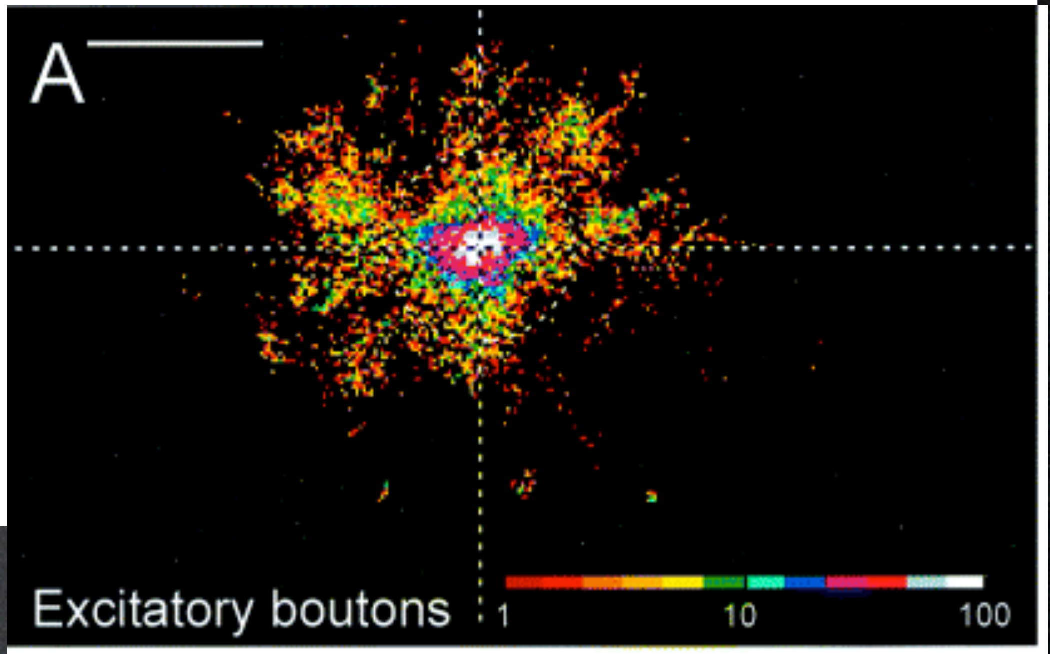


Similar Orientations, More Connections

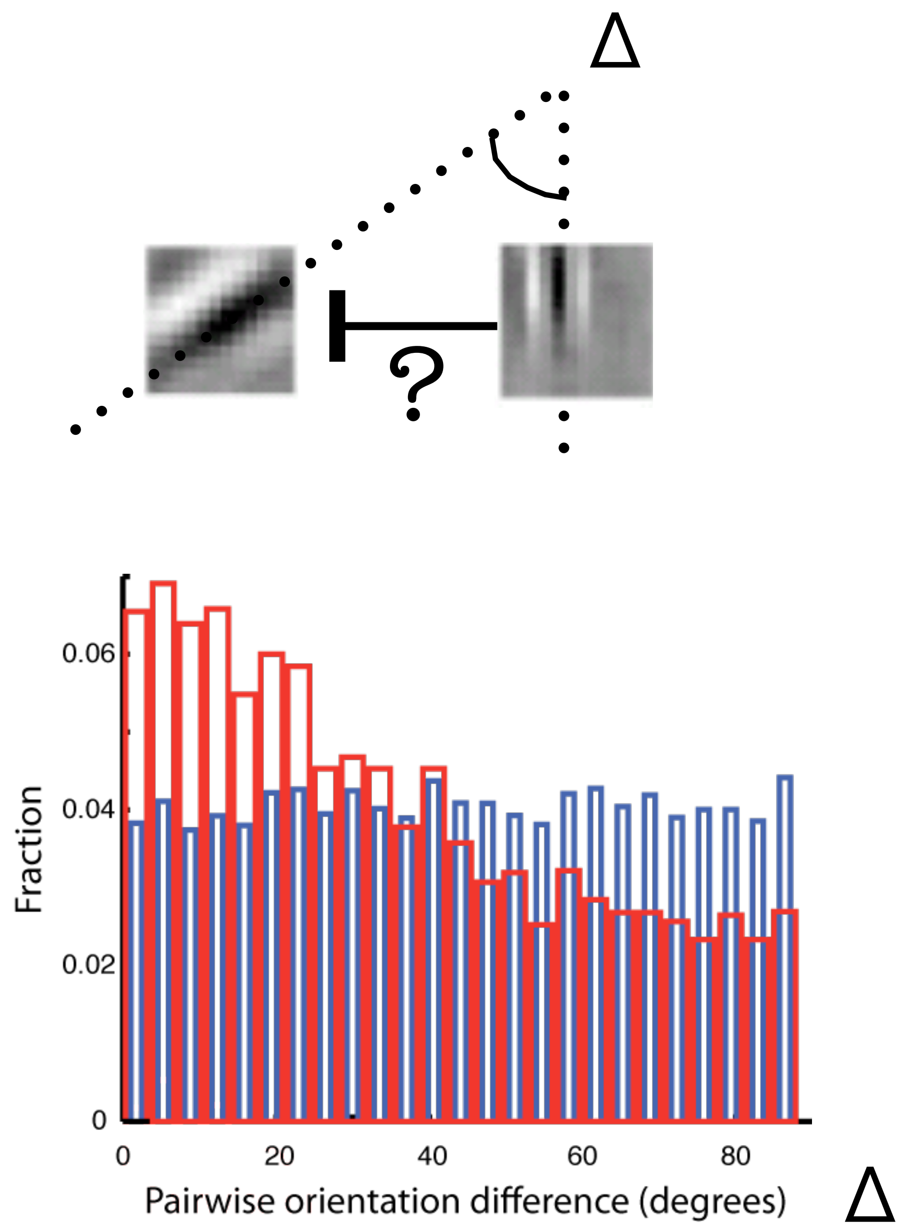


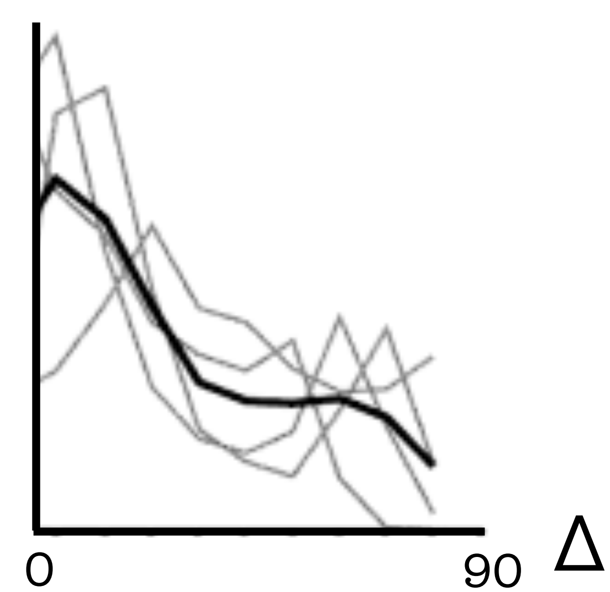






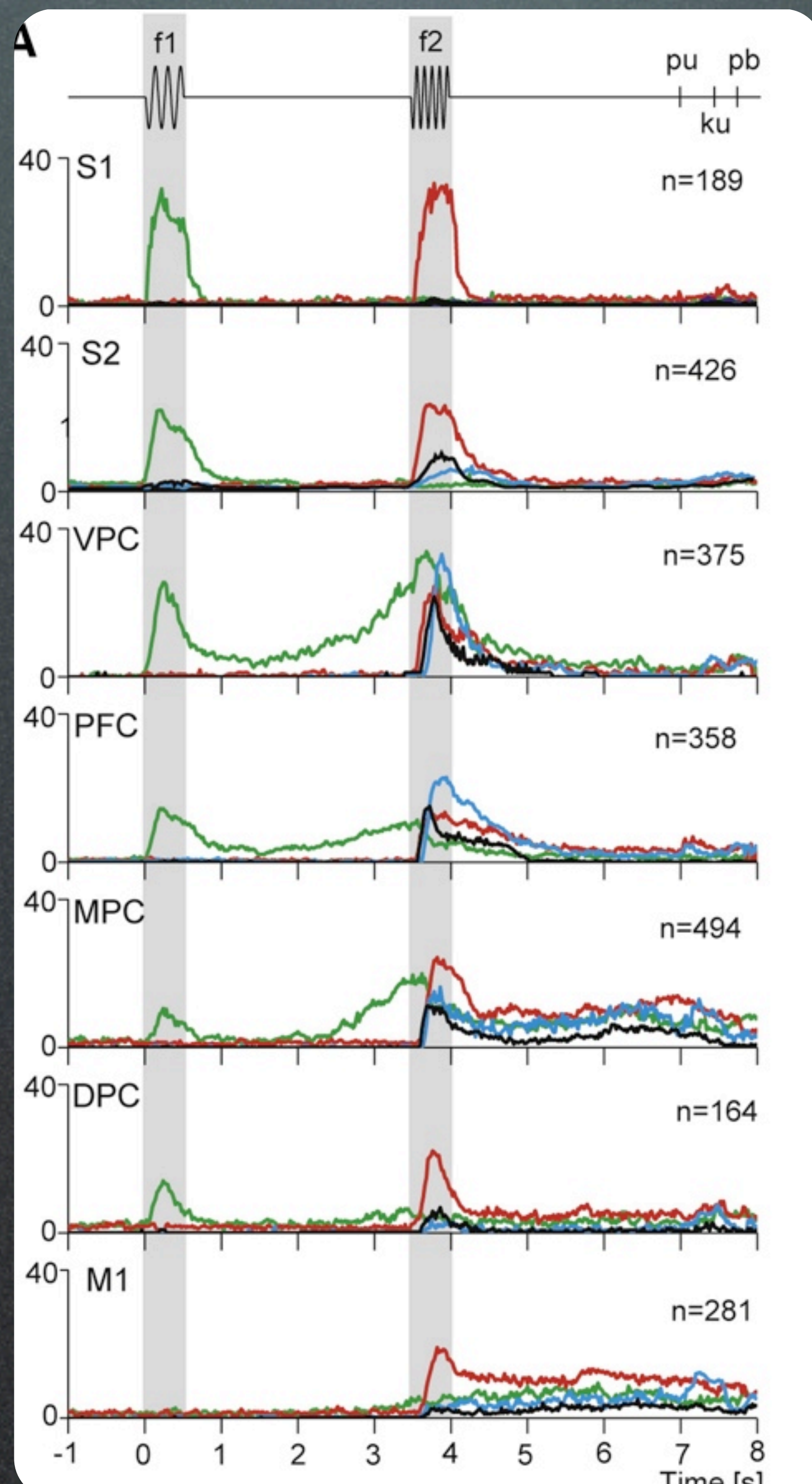
Similar Orientations, More Connections



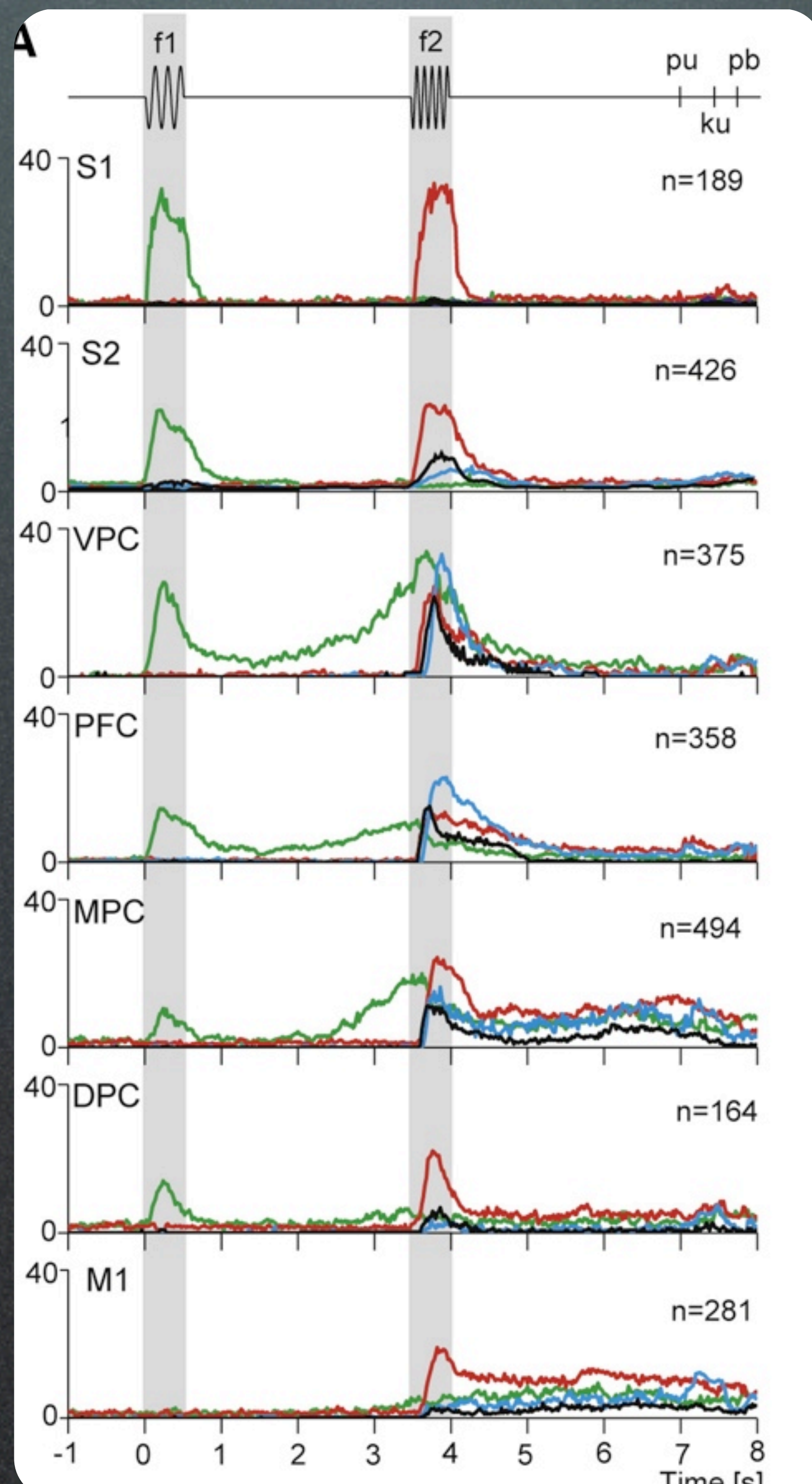


Youssef et al. (1999)

Persistence across Cortex



Persistence across Cortex



Hernandez, alii, et Romo (2010)