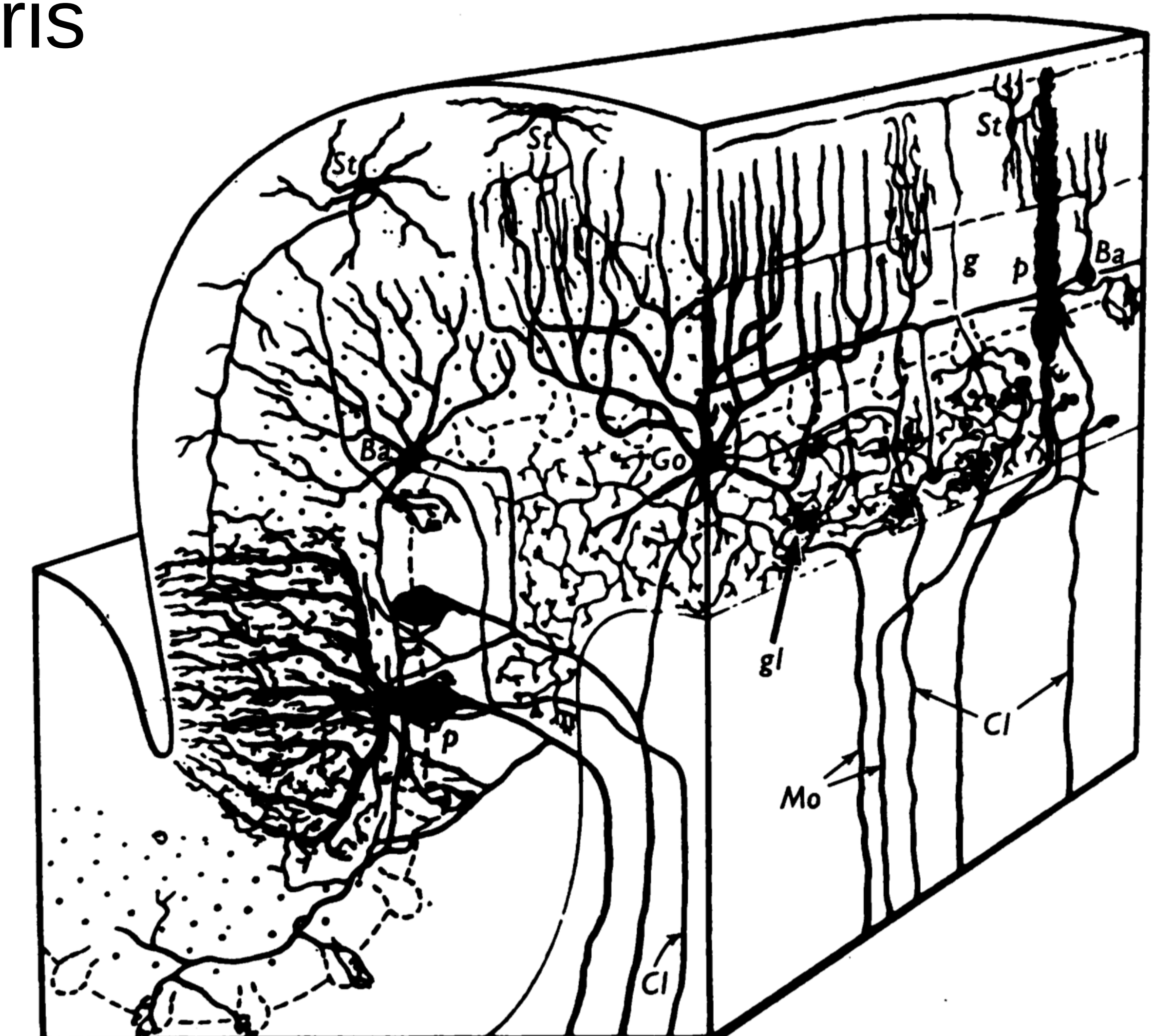
# Connections between dimensionality and network sparsity

Kameron Decker Harris
University of Washington
Computer Science, Biology

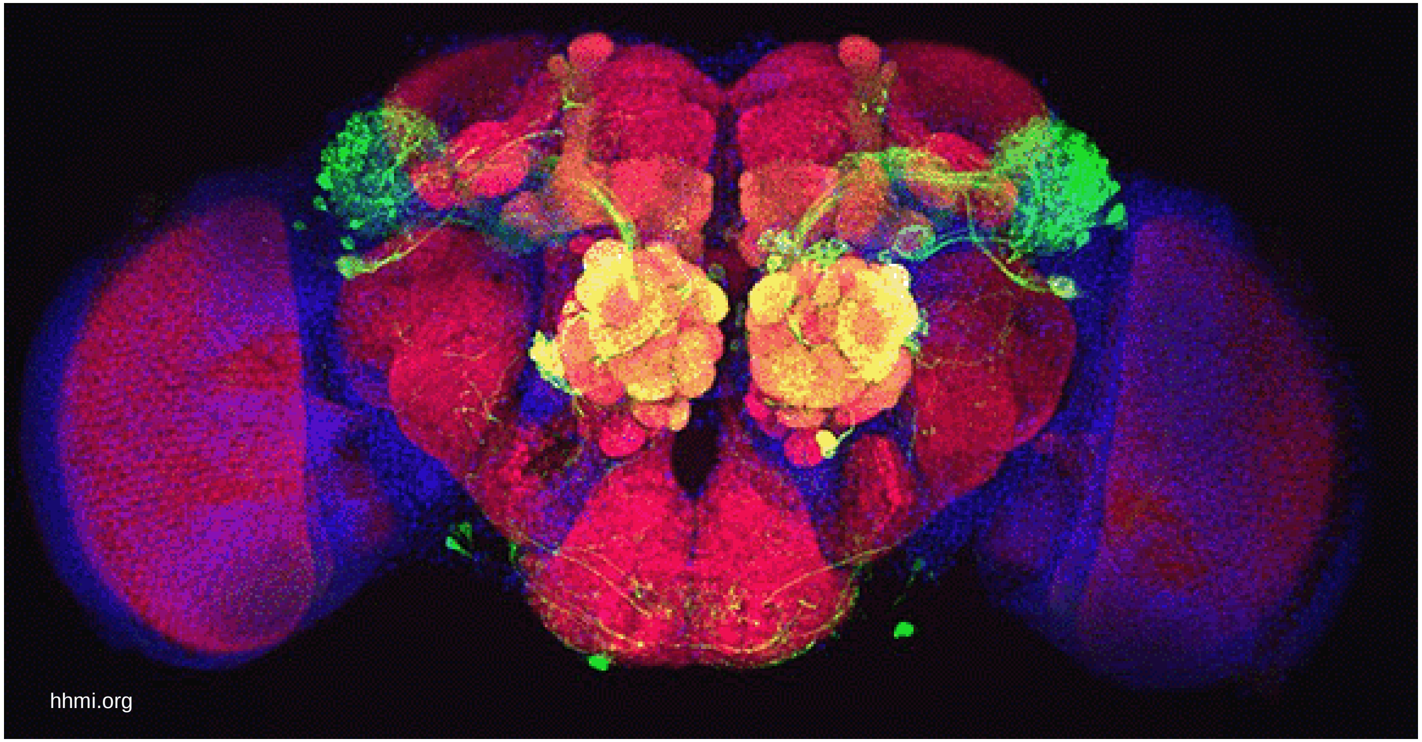
See our paper

"Optimal synaptic connectivity"
Litwin-Kumar, Harris, Axel
Sompolinsky, and Abbott.
Neuron (2017)

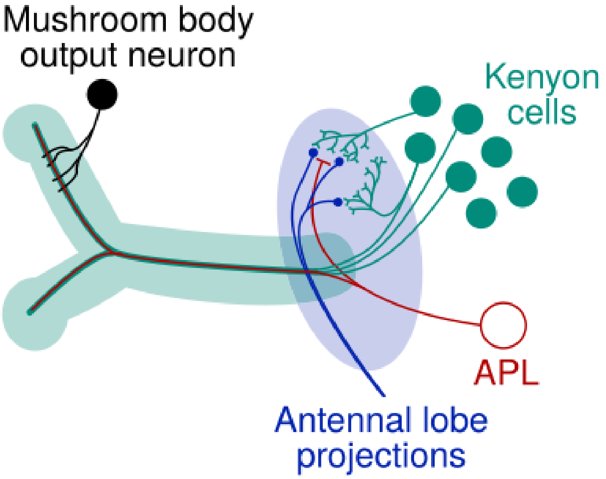



Cerebellum
from Eccles et al. (1967)

# Olfactory network of *Drosophila*

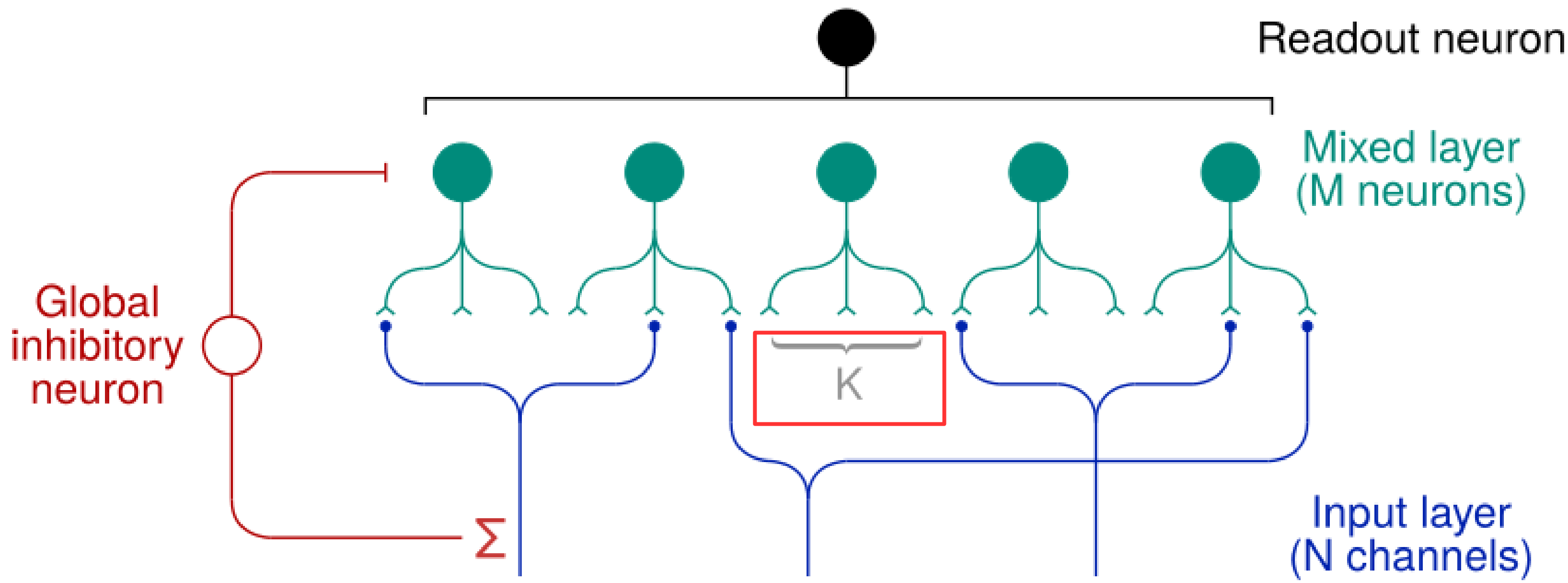



Output neuron decides:
good smell or bad smell?

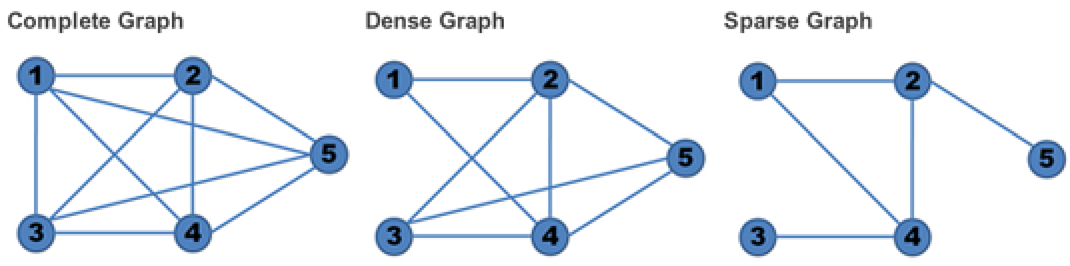



See related work to ours:
Cayco-Gajic, Clopath, Silver (2017)
Dasgupta, Stevens, Navlakha (2017)

# Common brain network structure: 2-layer sparse expansion



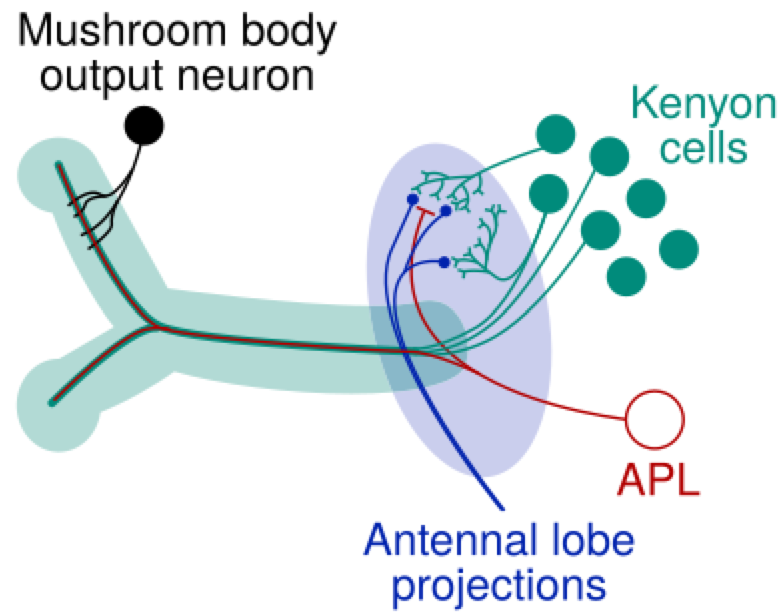

Parameters:
- K small
- M >> N, both large



# Why such *network sparsity*?

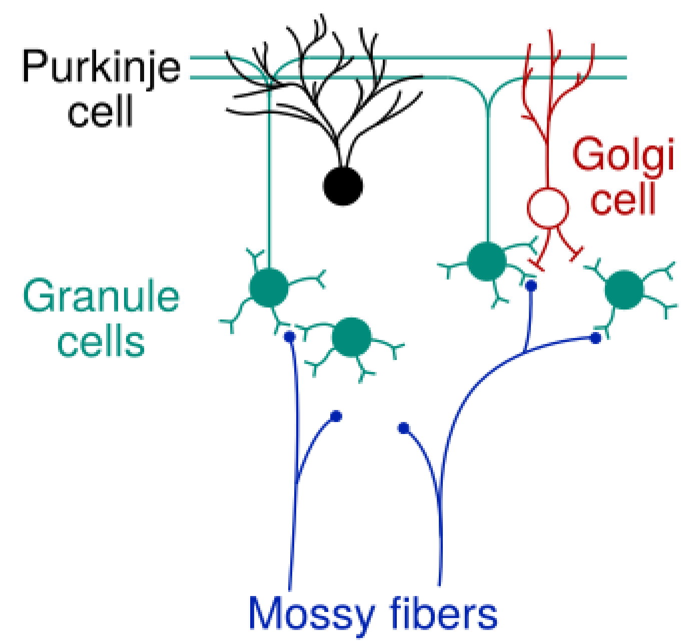
Mushroom body (olfaction)

- N = 50 and M = 2,000
- 40-fold expansion
- K = 7 inputs per mixed cell



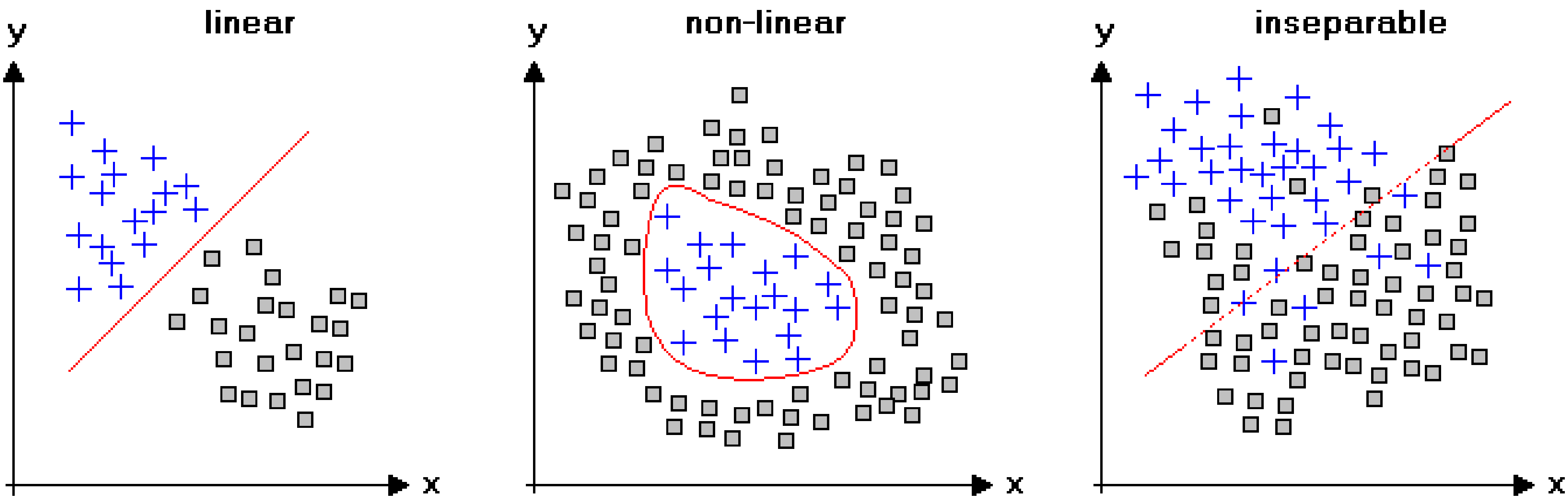

Cerebellum (motor control)

- N = 7,000 and M = 210,000
- 30-fold expansion
- K = 4 inputs per mixed cell

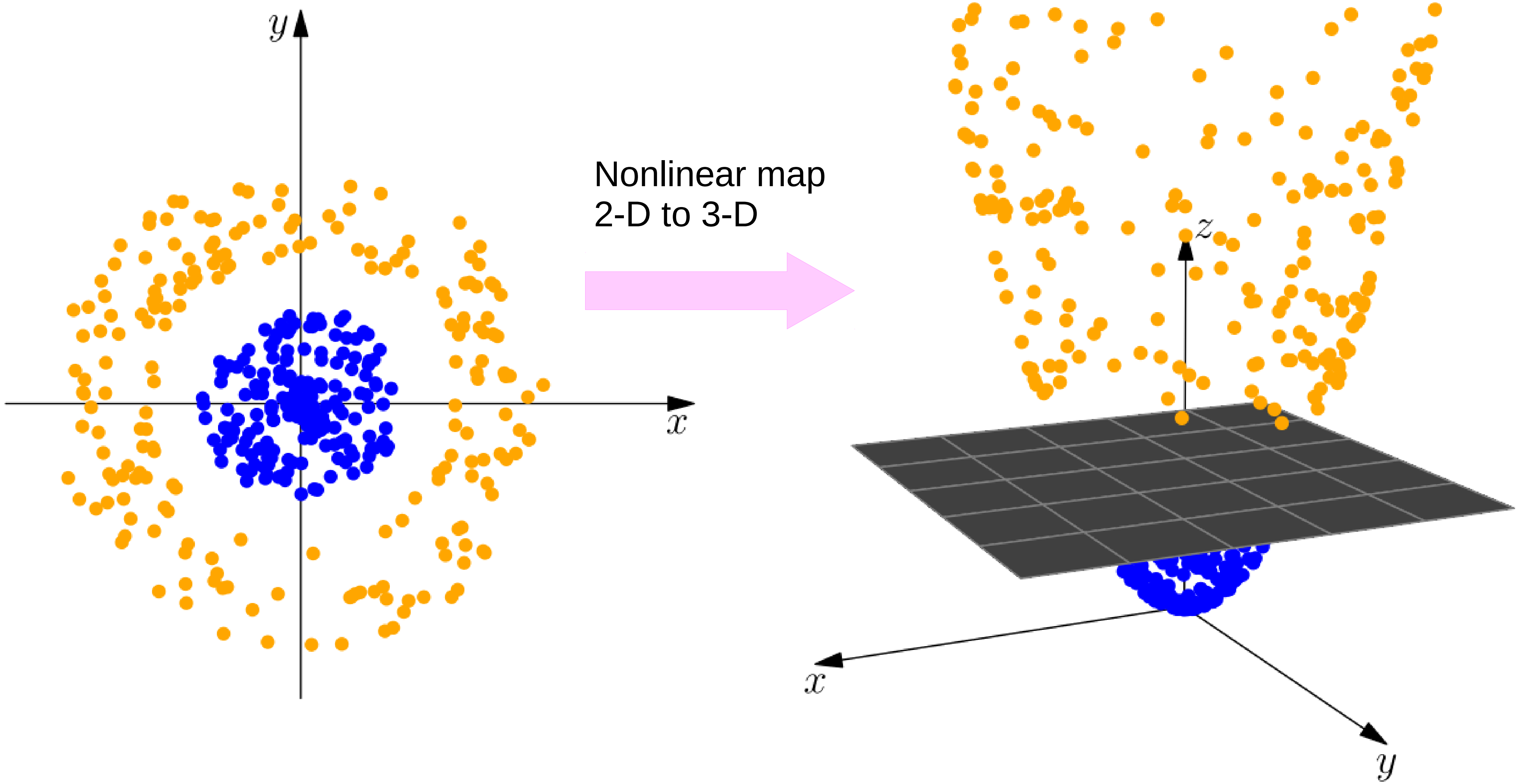



Some of the previous theories:
Marr (1969); Albus (1971); Hansel & van Vreeswijk (2012); Rigotti et al. (2013); Barak et al. (2013); Babadi & Sompolinsky (2014)

# Linear vs. nonlinear separability




H.Lohninger: Teach/Me Data Analysis, Springer-Verlag, Berlin-New York-Tokyo, 1999.




Expectation:

Expanding dimensionality improves pattern separation

https://www.researchgate.net/publication/262800142_Lecture_Notes_on_the_Kernel_Trick_I

# History of a dimensionality measure

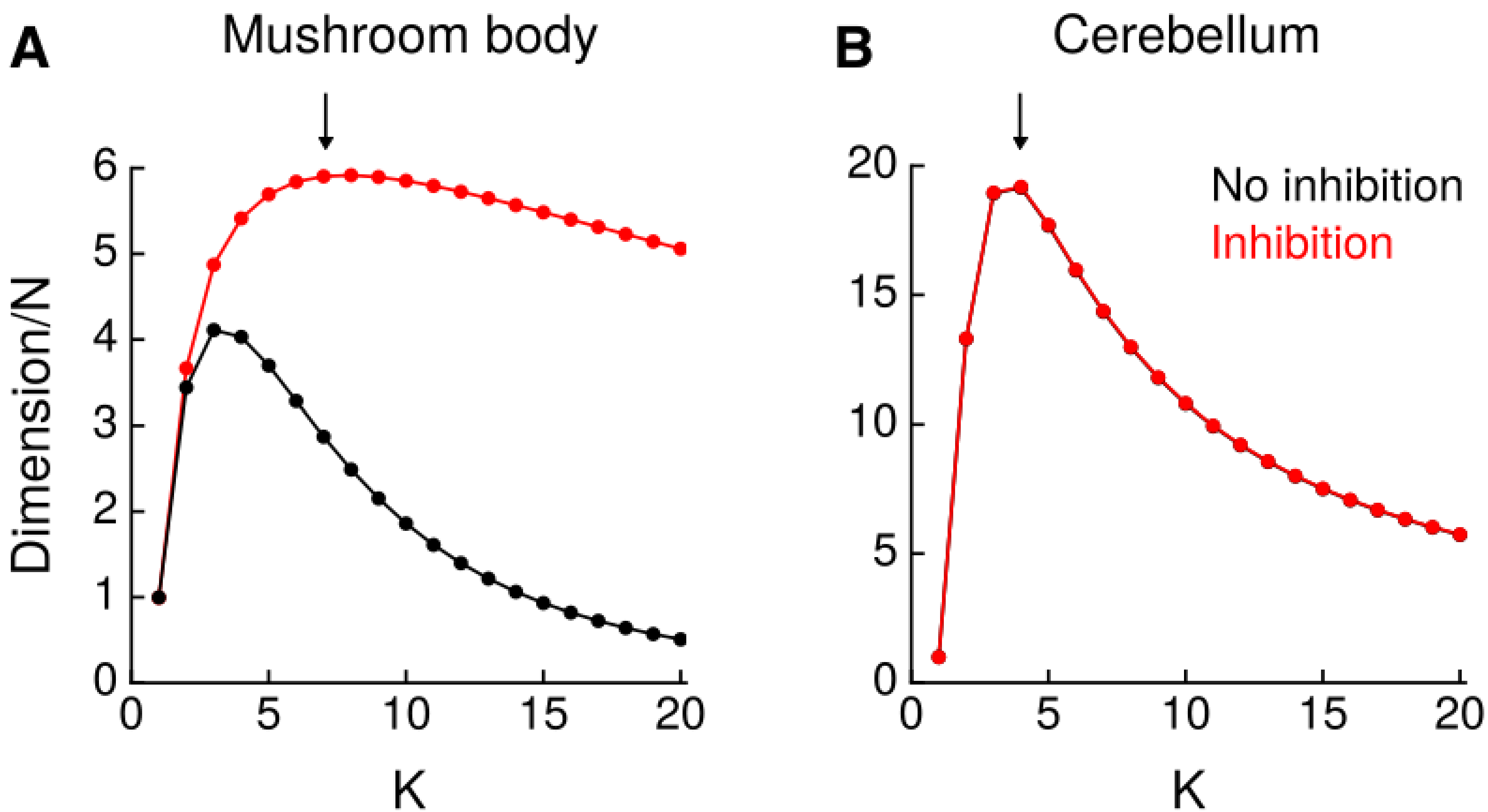
$$\dim(\mathbf{x}) = \frac{\left(\sum_{i=1}^{M} \lambda_i\right)^2}{\sum_{i=1}^{M} \lambda_i^2}$$

Eigenvalues
of mixed-layer
covariance

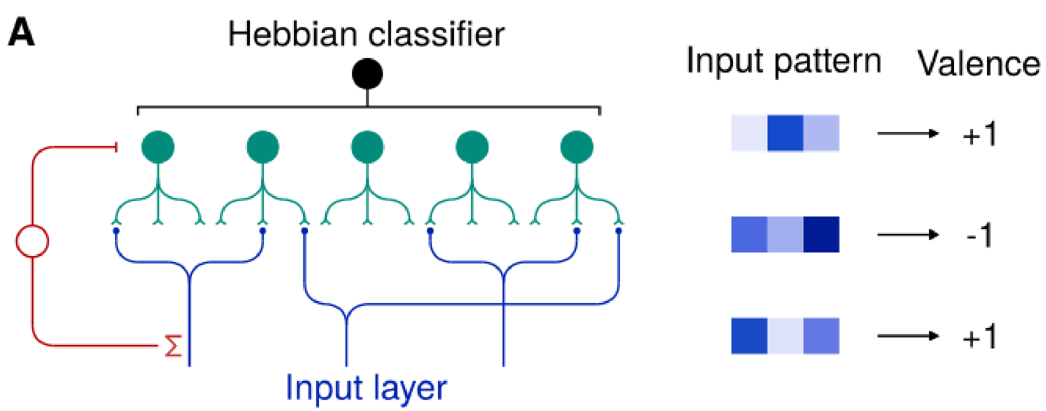
| | |
|---|---|
| Vocabulary “characteristic,” linguistics | Yule (1944) |
| Diversity measure, ecology | Fisher (1943), Simpson (1949) |
| Rényi entropy, mathematics | Rényi (1961) |
| Participation ratio/purity, physics | Bell & Dean (1970) |
| Dimensionality, neuroscience | Abbott, Rajan, Sompolinsky (2009) |

# Sparsity maximizes dimensionality

Arrows = avg degree observed in brains



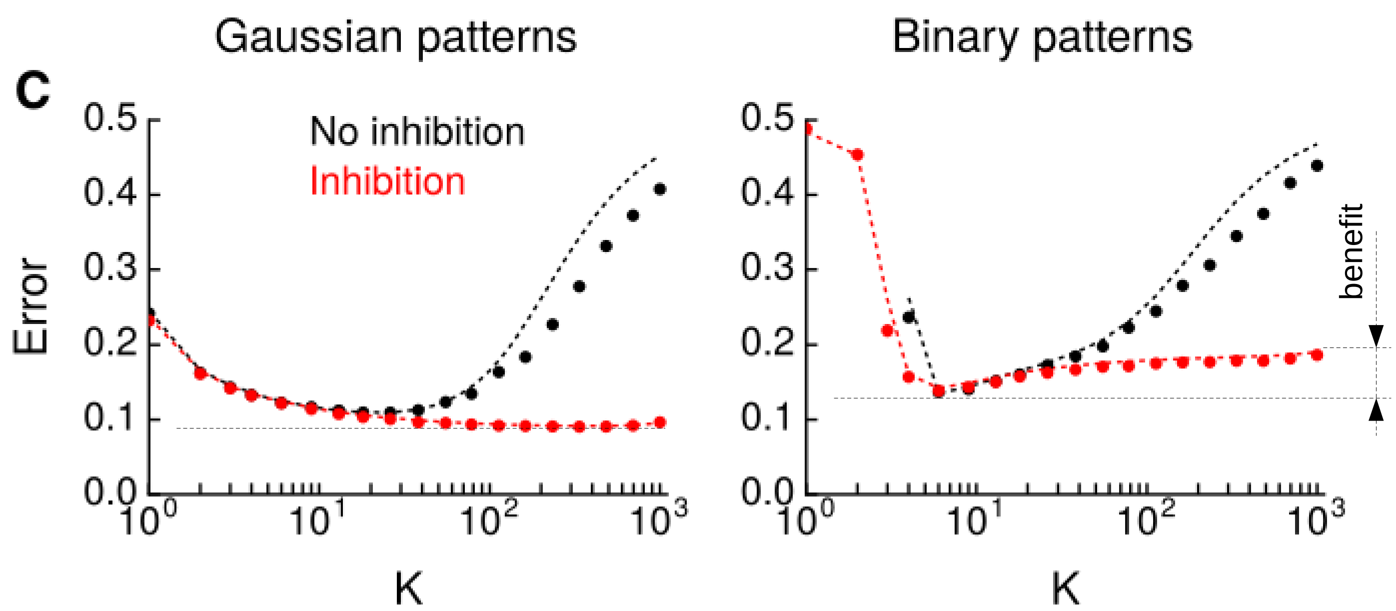
# Sparsity can improve classification




Classifier must also **denoise**

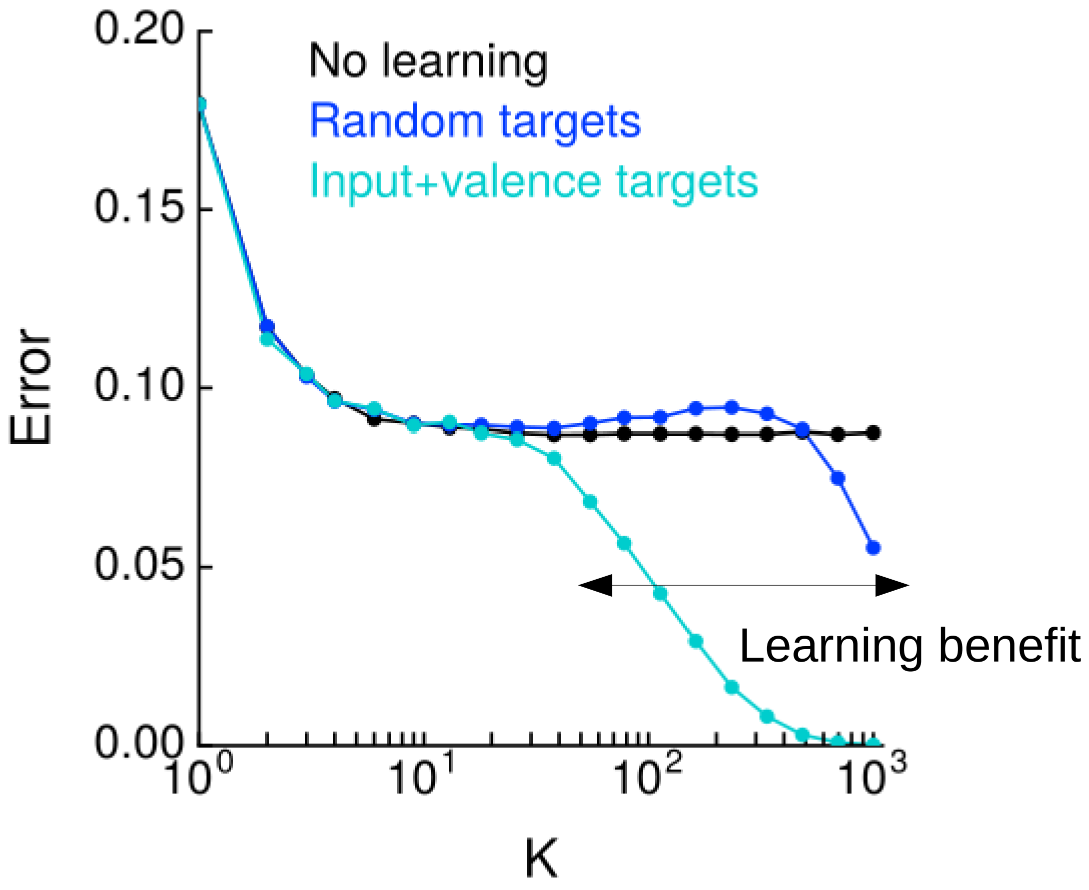
Uncorrelated random
- Input patterns
- Binary valence
- Binary noise

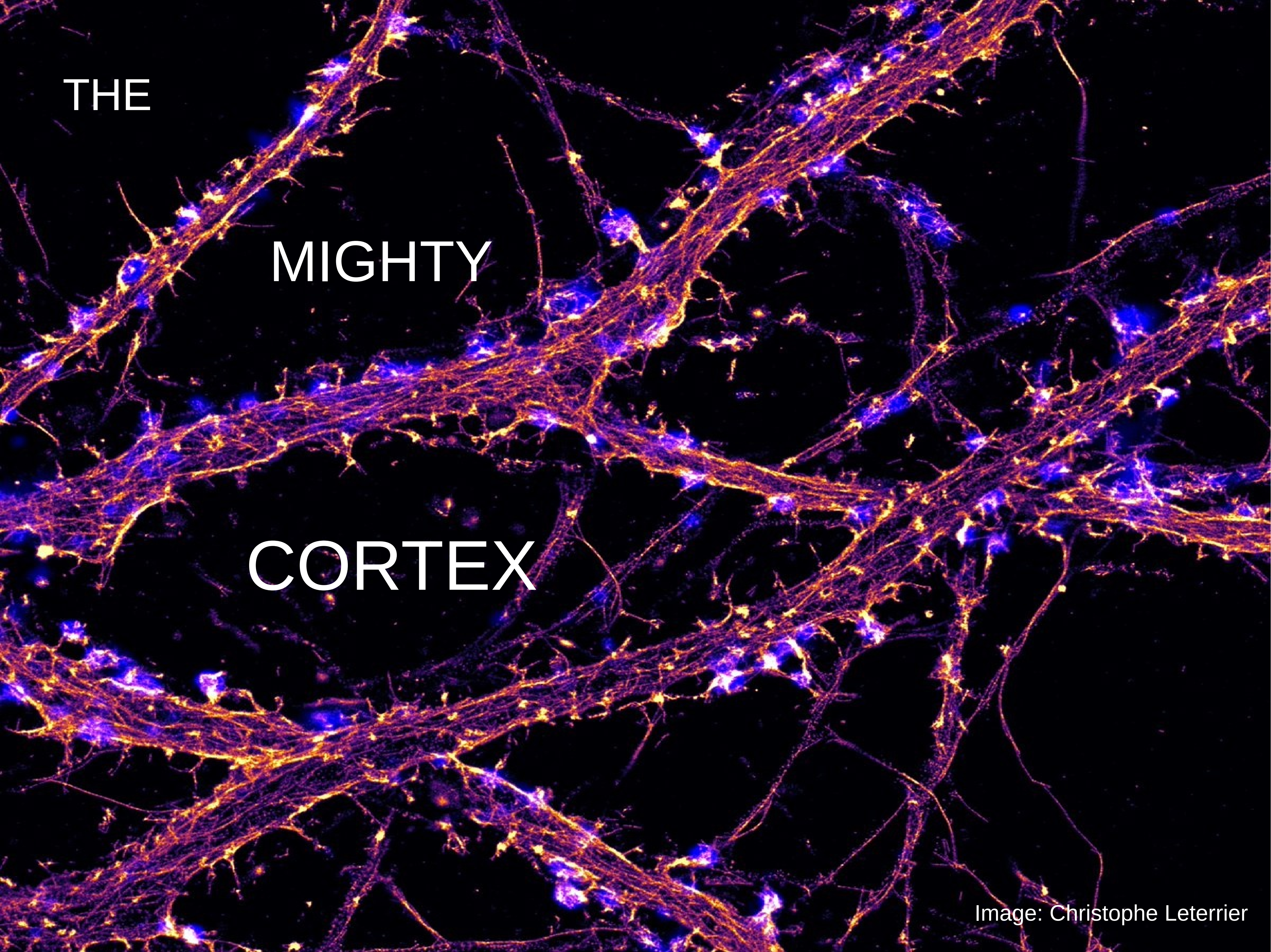
$$\mathrm{SNR} \approxeq \frac{\dim(\mathbf{m}) \cdot (1-\Delta)^2}{P}$$




Noise amplification

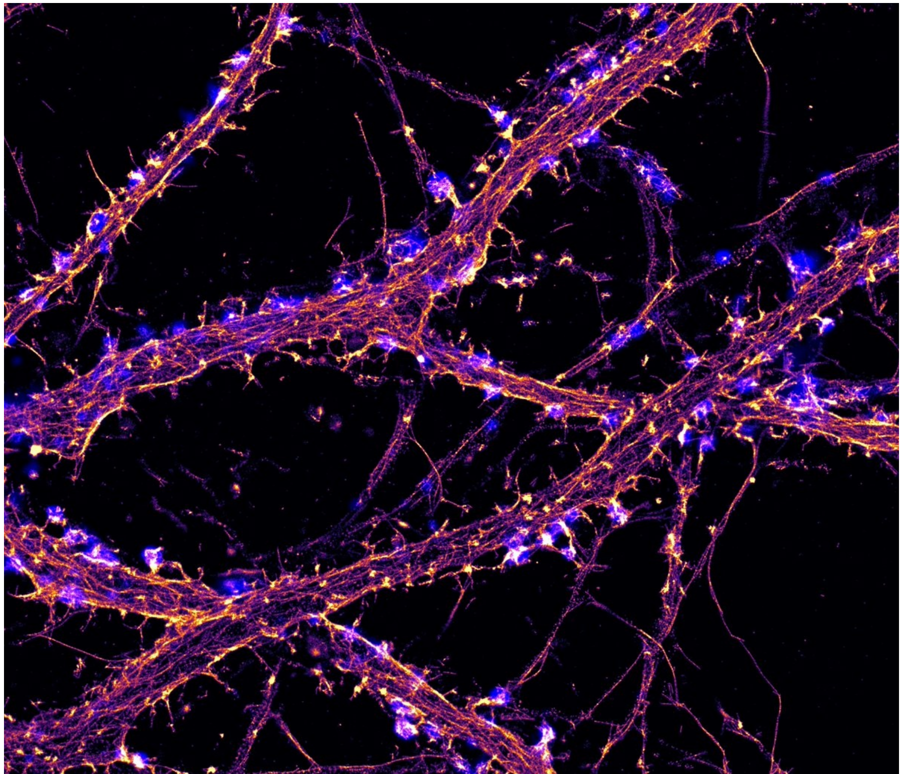
# Learning input-mixed weights most useful only in dense networks




THE
MIGHTY
CORTEX
Image: Christophe Leterrier

# Conclusions

- Sparsity optimal when **features are random**
- Dense connectivity if **features are learned**
- Results suggest principles of
  - Olfaction and cerebellum
  - Contrasting with cortex
- Coding sparsity important



# (Fast?)Food for thought

Review: “Randomness in neural networks” by Scardapane & Wang (2017)

- “Random features” show up many places
  - Rosenblatt’s perceptron (1958), expansion weights
  - Functional link networks & universal approximation

    Barron (1993), Pao, Park, Sobajic (1994), Igelnik & Pao (1995)
  - Radial basis functions    Broomhead & Lowe (1988)
  - Random features ~ kernels    Rahimi & Recht (2007)
  - Deep network kernels    Mairal et al (2014), Daniely, Frostig, Singer (2016)
- Meaning of dimensionality in statistical learning
  - Decay rate of eigenvalues in some basis (RKHS)
  - How smooth is your function?

# Crumbs...

- Remains to be seen how often sparse, random features yield benefits with real data
- Preliminary analysis with Merck dataset (chemical features) says yes:

```
linear regression          score
avg training score         0.675
avg testing score          0.613
2-layer dense network
avg training score         0.779
avg testing score          0.670
2-layer sparse network               rel. diff
avg training score         0.751        -3.5%
avg testing score          0.659        -1.7%
```

# Thank you!


- Funding:
  - UW NIH Big Data for Genomics and Neuroscience T32LM012419
  - Boeing Fellowship
  - Washington Research Foundation
- Collaborators:
  - Ashok Litwin-Kumar, Larry Abbott, Richard Axel (Columbia); Haim Sompolinsky (Harvard & Hebrew U.)