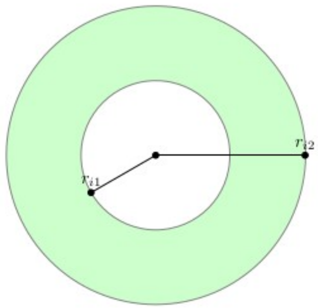
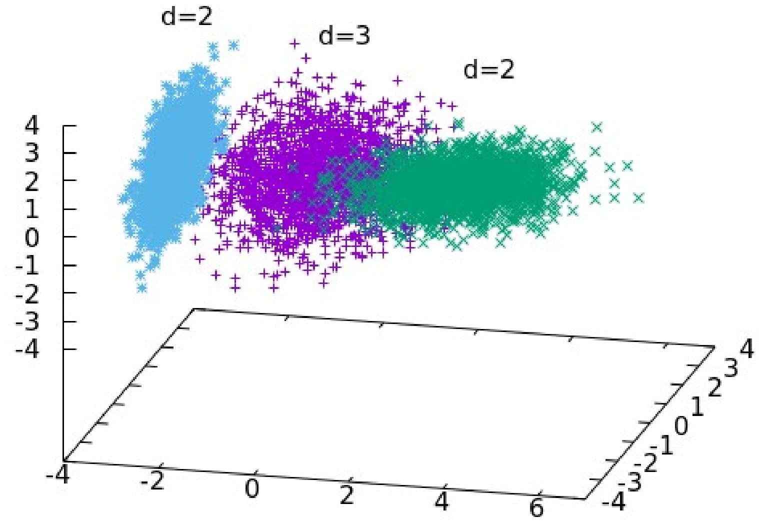
Intrinsic dimension estimation and neural activity

Overview

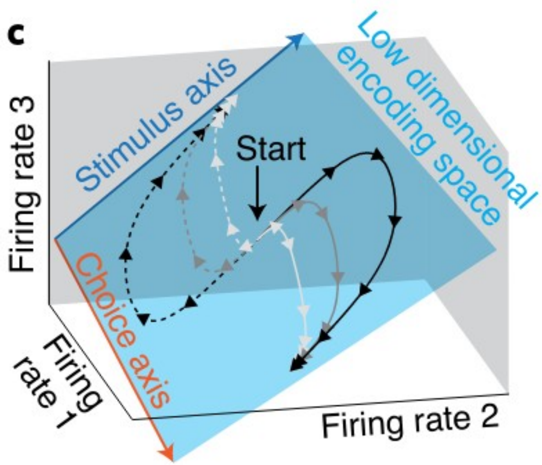
Intrinsic dimension estimation:
the two-NN approach



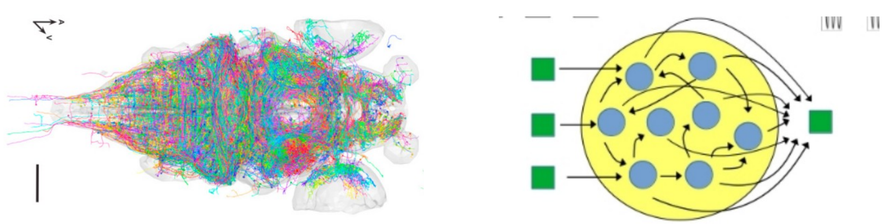
The case of variable ID



Low-dimensional neural codes

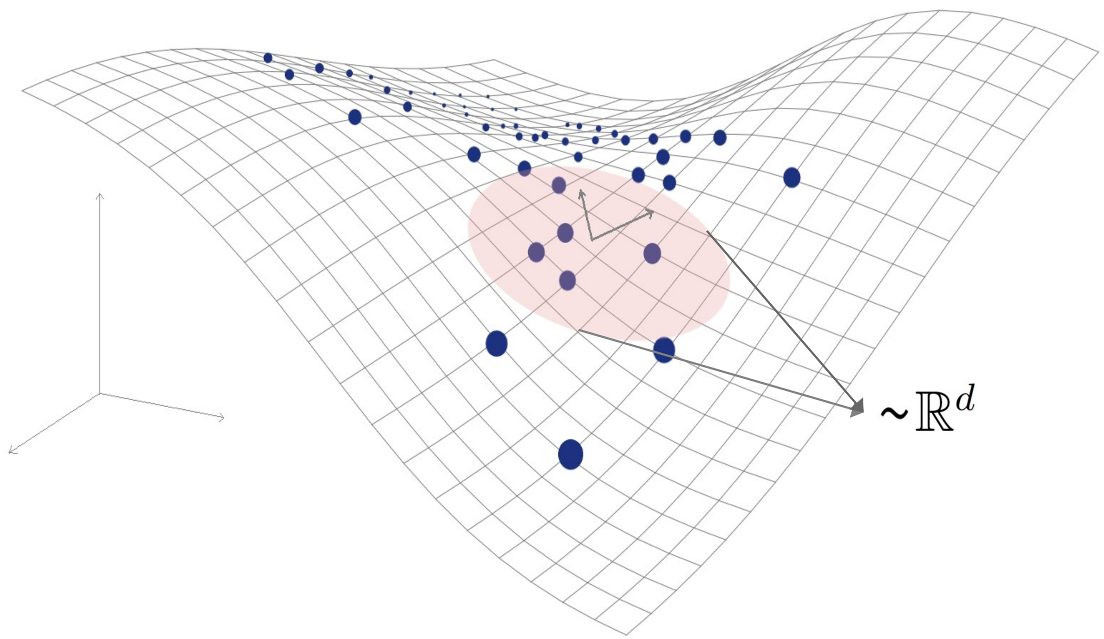


Some ideas for future work



Intrinsic dimension

- Data are defined in a space with D variables
- Generally, the data lie on hypersurface of lower dimension d << D
- This dimension is called the ***intrinsic dimension***



ID estimation: projective approach

Project *D*-dimensional data into lower dimension *d* : $\Pi^d : \mathbf{x}_i \in \mathbb{R}^D \mapsto \mathbf{y}_i \in \mathbb{R}^d$

- Try different *d* and evaluate for each a "loss function" $\mathcal{L}(\Pi^d)$
- $\mathcal{L}(\Pi^d)$ measures the "data loss" occurring in the projection. Examples:

$$\mathcal{L}(\Pi^d) = \sum_i ||\mathbf{x}_i - \mathbf{y}_i||^2$$ preservation of original distance relations

$$\mathcal{L}(\Pi^d) = \sum_i \mathbf{x}_i\mathbf{x}_i^T - \mathbf{y}_i\mathbf{y}_i^T$$ preservation of original covariance matrix

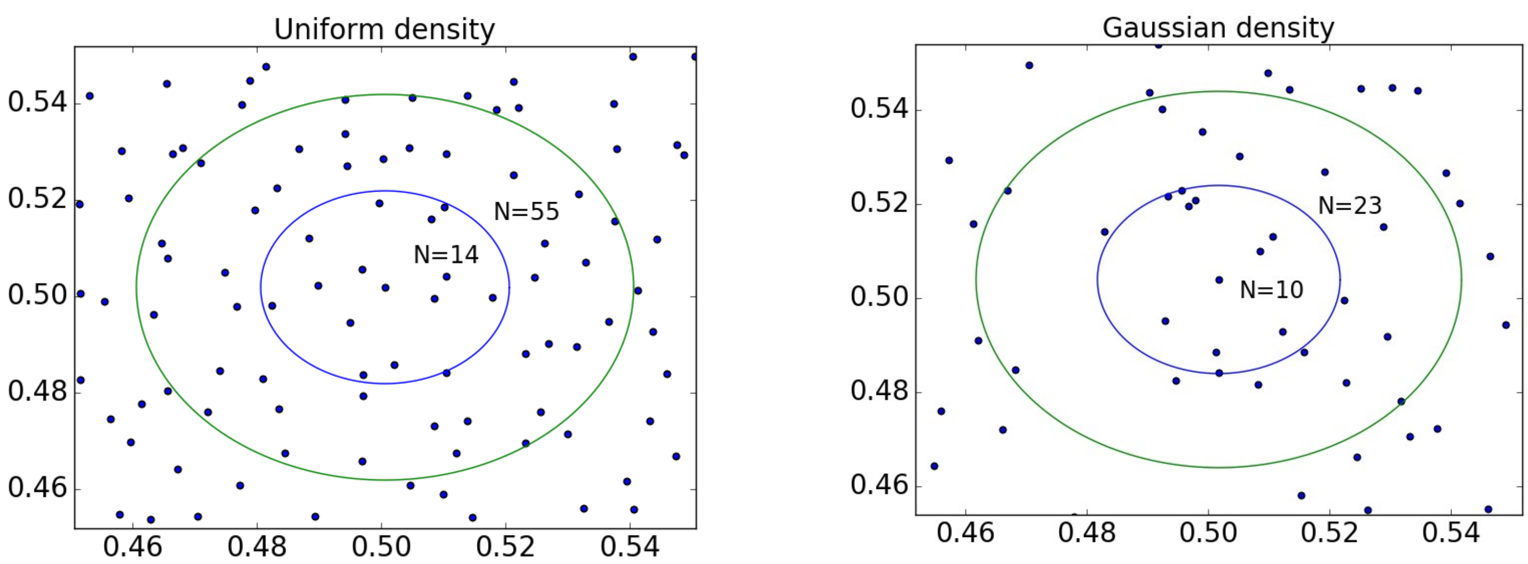
- **ID estimate based on tradeoff between dimensionality reduction and data loss**
- Problem (1): Computationally burdensome (search for optimal projection for each *d*)
- Problem (2): robust ID estimates only if $\mathcal{L}(\Pi^d)$ has large gap as a function of *d*

if no gap, the estimation can be rather arbitrary

ID estimation: scaling approach

- Data are sampled from a distribution *ρ(X)*
- **distances follow scaling laws that depend on *ρ, d***
- Example: "correlation dimension" (Grassberger & Procaccia, PRL 50, 1983):

  The number of nearest neighbors at distance <ε from point *i* scales as $N_i(\varepsilon) \sim \rho\varepsilon^d$
  estimate *d* with simple linear fit.
- ρ should not have large variations, or the scaling is violated (and estimation fails)



ID estimation: TWO-NN

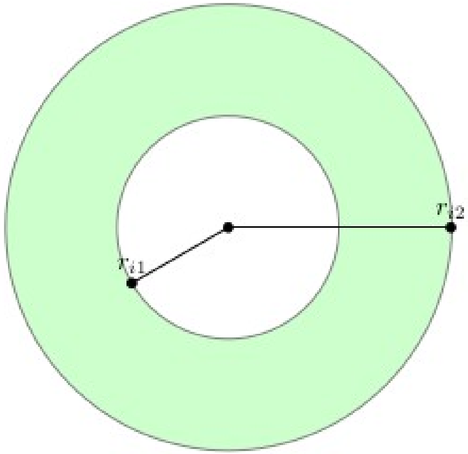
E Facco, M D'Errico, A Rodriguez, A Laio, Scientific Reports 7, 12140. (2017)

- TWO-NN: estimating the ID in case of (strongly) variable density

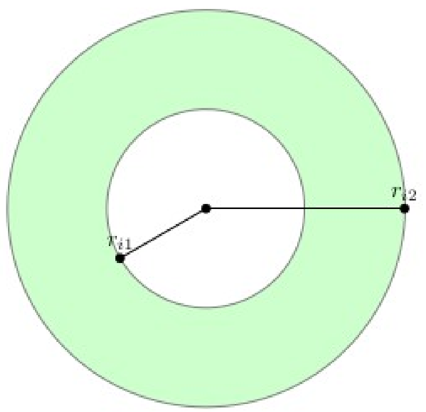
  Two weak assumptions:

  - **H1)** the data points x_i are **independent samples** from a distribution $\rho(x)$.
  - **H2) local uniformity:** $\rho(x) \sim$ const. in the region containing the first 2 neighbors of x_i

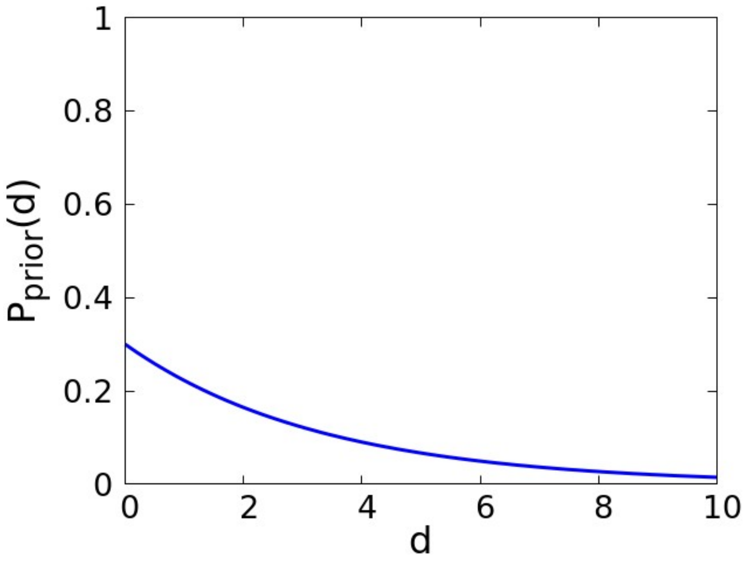
- r_{i1}, r_{i2} distances of 1st and 2nd neighbor of point i
- $\mu_i = r_{i2}/r_{i1}$ follows **Pareto distribution: $P(\mu) = d\mu^{-d}$**
- **the distribution of μ depends only on *d***



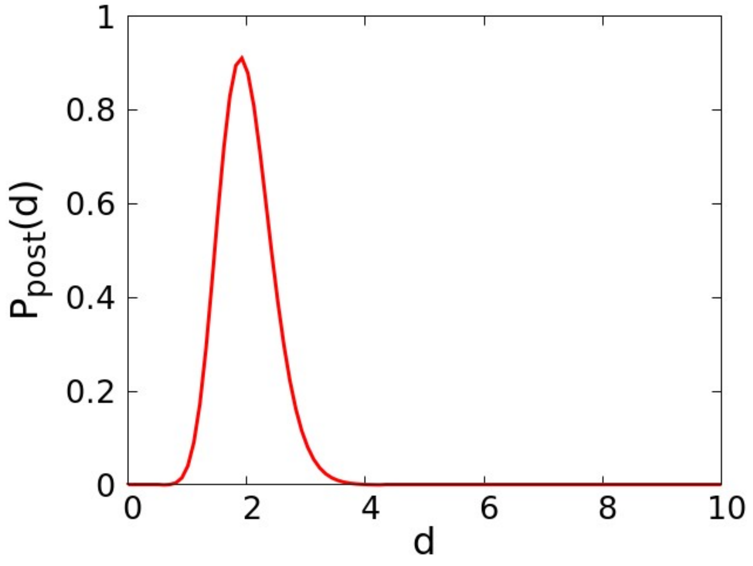
ID estimation: TWO-NN



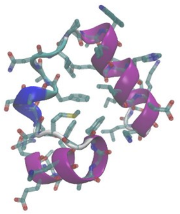
- assume $P_{prior}(d) \sim Gamma(a,b)$
- $\mu_i = r_{i2}/r_{i1}$ follows a Pareto distribution: $L(\mu|d) \sim Par(d)$
- given the $\{\mu_i\}$, $P_{post}(d) \sim Gamma(a+N, b+N\langle\log \mu\rangle)$
- **d estimate**: posterior average $\langle d\rangle_{post} = (a+N)/(b+N\langle\log \mu\rangle) \sim 1/\langle\log \mu\rangle$

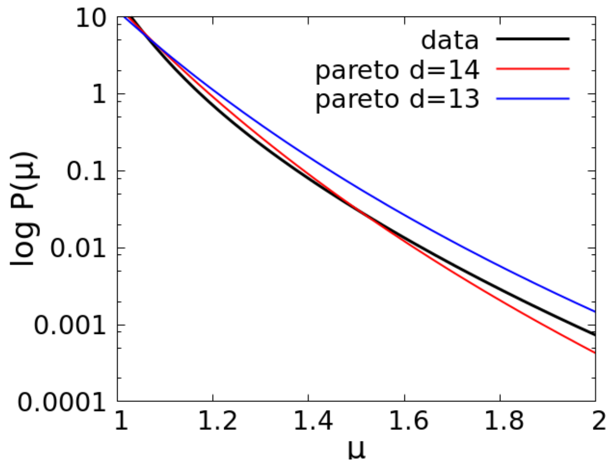


data $\{\mu_i\}$



The problem of multiple IDs



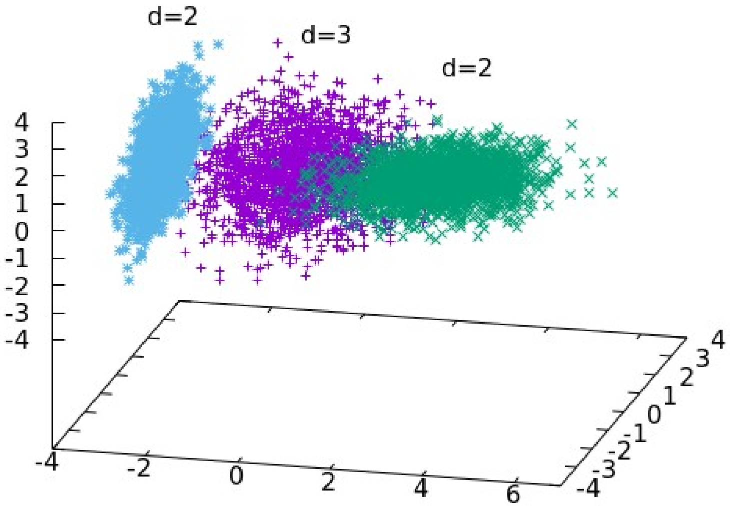


N ~ 32,000 configurations
M=32 coordinates

the empirical μ do not well fit a Pareto

What if *d* is not uniform?

the data may lie on several manifolds, each with different intrinsic dimension



Extending TWO-NN to multiple IDs

[M. Allegra, E. Facco, F. Denti, A. Laio and A. Mira, Scientific Reports 10,16449 (2020)

- Assume the points lie on K «manifolds» with different IDs **d**=$d_1,\ldots,d_K$

- The distribution of {μ} is simply a **mixture of Pareto distributions** (with mixing parameters **p**=$p_1,\ldots,p_K$)

$$\mu \sim \sum_k p_k Par(d_k) \qquad \Rightarrow \qquad \mathcal{L}(\boldsymbol{\mu}|\mathbf{d},\mathbf{p}) = \prod_{i=1}^{N} \sum_{i=1}^{K} p_k d_k \mu_i^{-d_k-1}$$

$$P_{prior}(d,p) \sim \prod_{k=1}^{K} Gamma(a_k,b_k) \cdot Dir(c_1,\ldots,c_K)$$

- Introduce latent variables **Z**=$Z_1,\ldots,Z_N$, **manifold membership** of each point

$$\mathcal{L}(\boldsymbol{\mu}|\mathbf{d},\mathbf{p},\mathbf{Z}) = \prod_{i=1}^{N} p_{Z_i} d_{Z_i} \mu_i^{-d_{Z_i}-1} \qquad\qquad P_{prior}(Z=k|\mathbf{p}) \sim p_k$$

- Estimate jointly **d,p,Z** by Gibbs Sampling of the posterior distribution

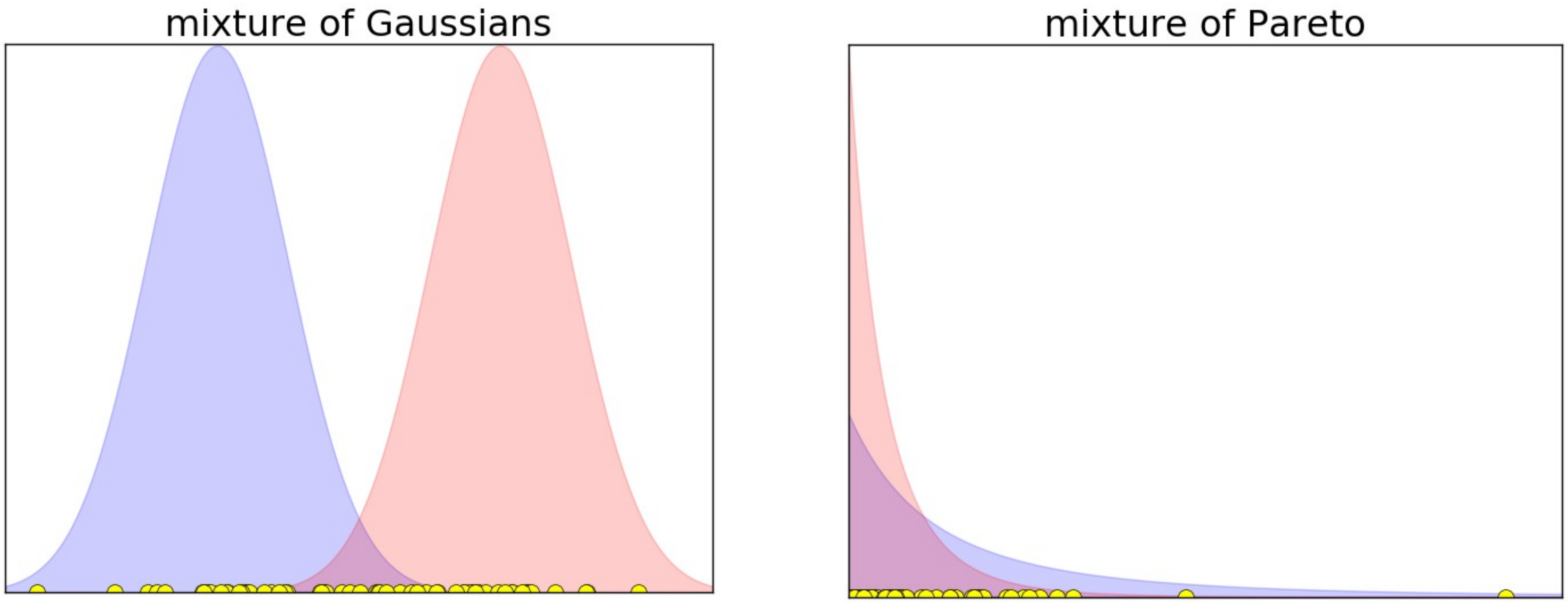
Extending TWO-NN to multiple IDs

Problem is correctly estimating Z

Z are easy to assign only if mixture components are largely non-overlapping

But Pareto distributions with different *d* are highly overlapping!

The Z assigment is not reliable



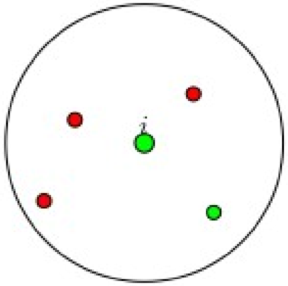
Extending TWO-NN to multiple IDs

Let the neighborhood of point *i* be defined by its first *q* neighbors

n_i^{in} # neighbors with same Z as *i*

n_i^{out} # neighbors with different Z

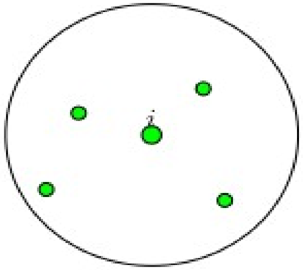
We get non-uniform neighborhoods: $n_i^{out} > n_i^{in}$



One more assumption: **neighborhoods must be approximately uniform**

Enforce with **additional term in the likelihood:**

$$\mathcal{L}(n^{in}|\boldsymbol{Z}) \propto \prod_i \xi^{n_i^{in}} (1-\xi)^{n_i^{out}} \qquad 0.5 < \xi < 1$$



ξ : propensity of neighbors to be in the same manifold

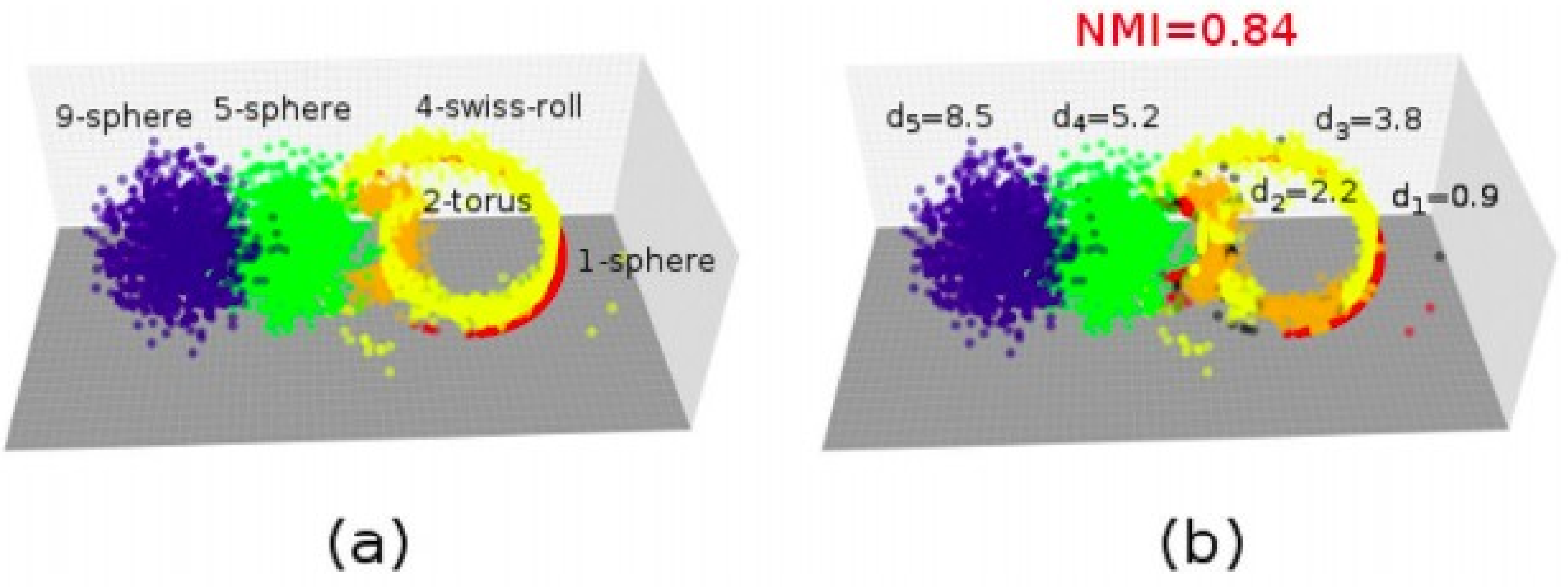
Heterogeneous ID algorithm (Hidalgo)

[M. Allegra, E. Facco, F. Denti, A. Laio and A. Mira, Scientific Reports 10,16449 (2020)

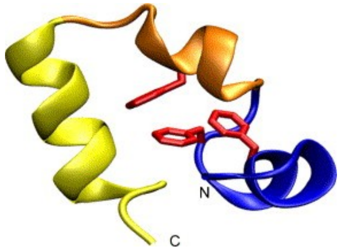
Find regions (manifolds) of different ID in the data

Works also for nonlinear and topologically complex manifolds

E.g. circle in d=1, swiss roll in d=4, torus d=2, sphere d=5, sphere d=9



Real example: phase space of folding protein



- simulation of unfolding/refolding protein
- for each of the N ~ 32,000 configurations, D=32 coordinates

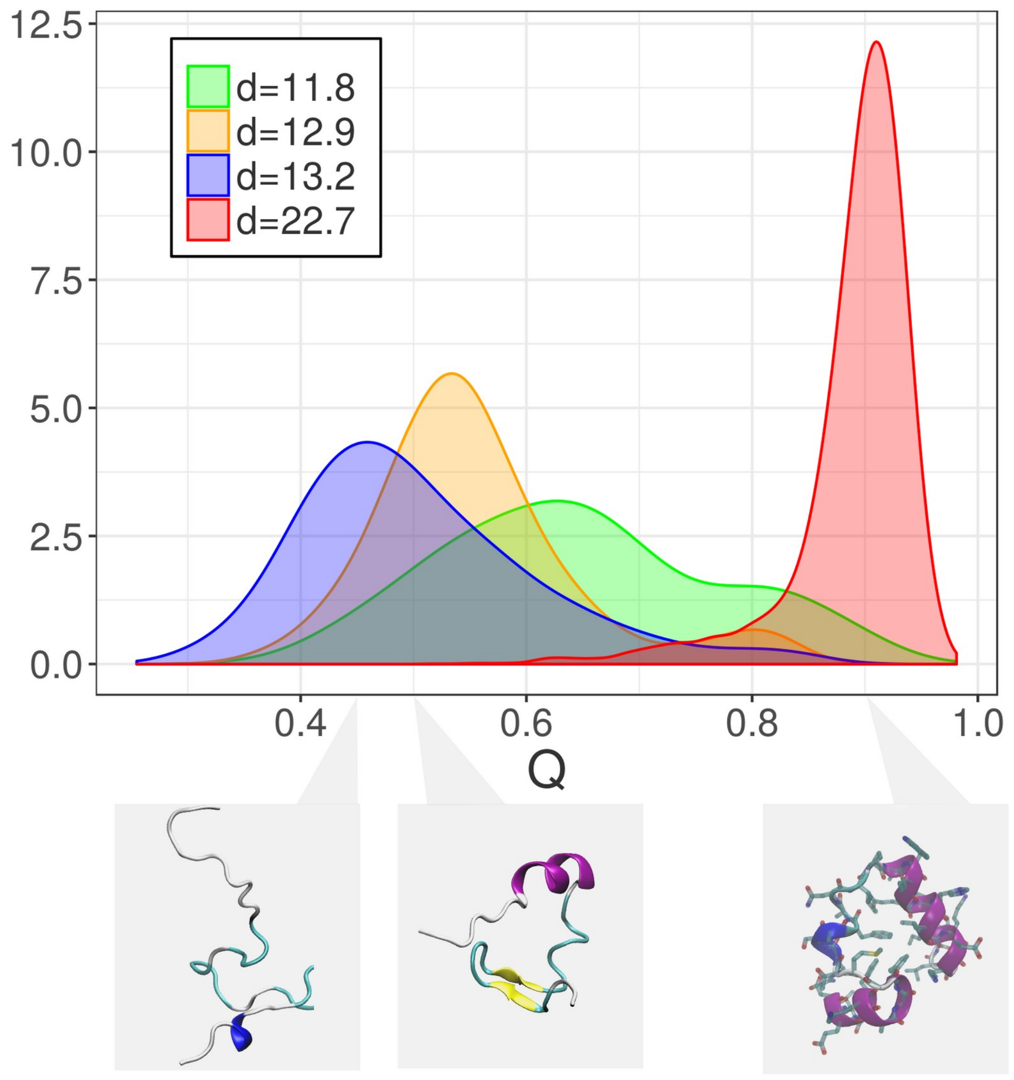
We find four manifolds,

- three with low dimensions d=11.8,d=12.9, d=13.2
- one with high dimension d=22.9

Which configurations are assigned to the different manifolds?

- Consider **q=fraction of native contacts (=degree of folding)**

Example: phase space of folding protein



- **Folded configurations are in the high-dimensional manifolds**
- The local ID is able to discriminate between folded and unfolded configurations

The effective # of phase space directions the system can explore varies in the two states

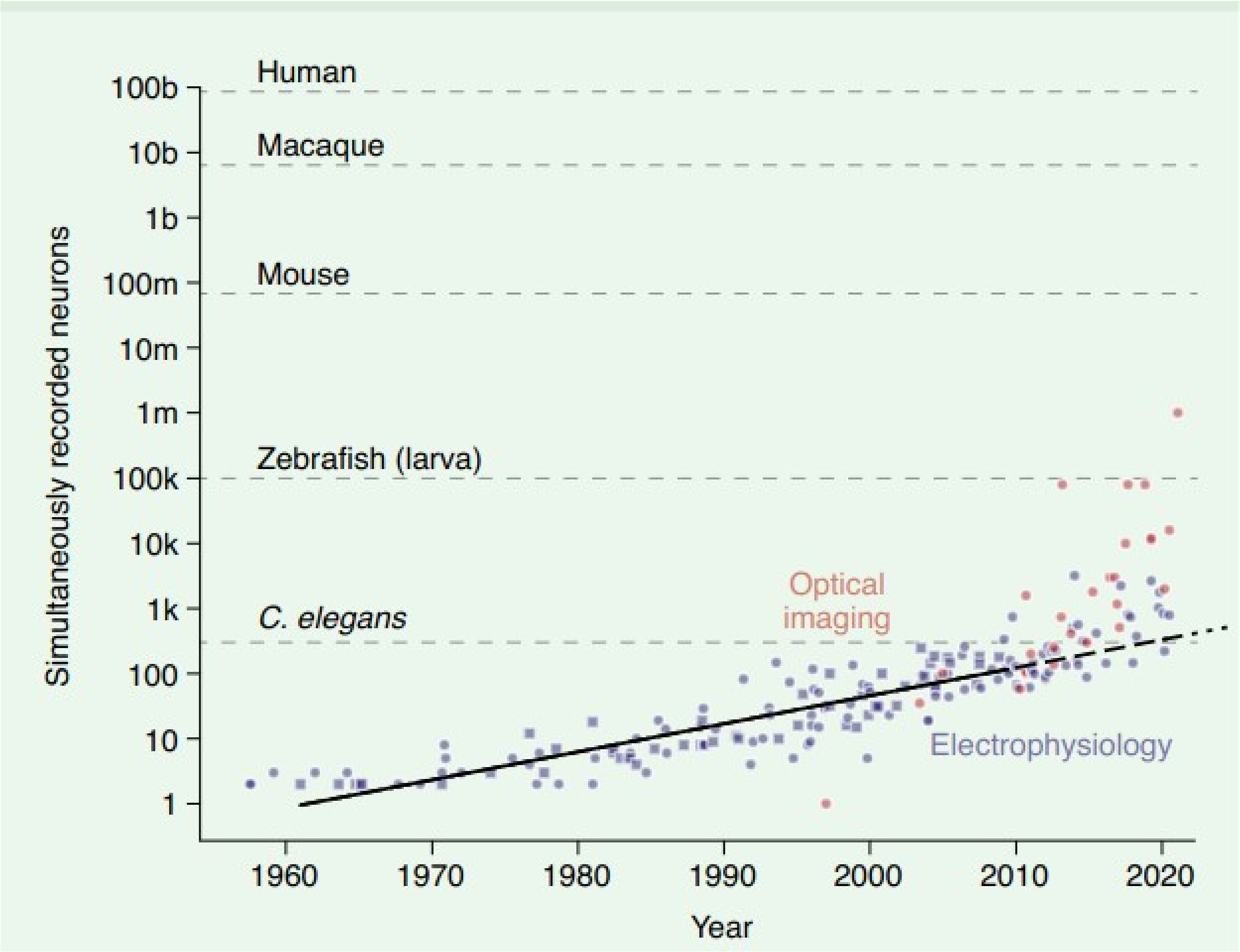
Conclusions first part

- We developed a Bayesian approach for intrinsic dimension (ID) estimation
- Our method rests on quite weak assumptions (local uniformity of density and dimension)
- We find regions of different local ID in the data
- In real data, we find large variations of the ID, highlighting relevant structure in the data
- An R implementation is freely available in the R package intRinsic

The advent of large-scale neural recordings

Moore-like increase in the amount of simultaneously recorded neurons

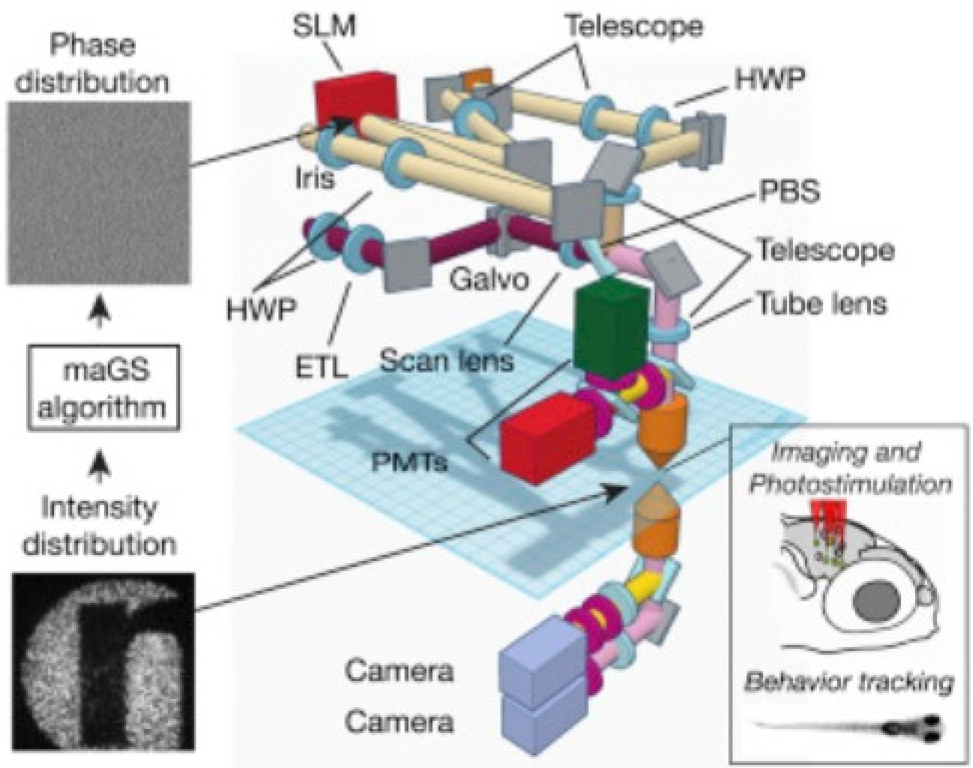
Urai, A. E., Doiron, B., Leifer, A. M., & Churchland, A. K. (2022). Large-scale neural recordings call for new insights to link brain and behavior. Nature neuroscience, 1-9.



The advent of large-scale neural recordings

Optical recordings through fluorescence imaging

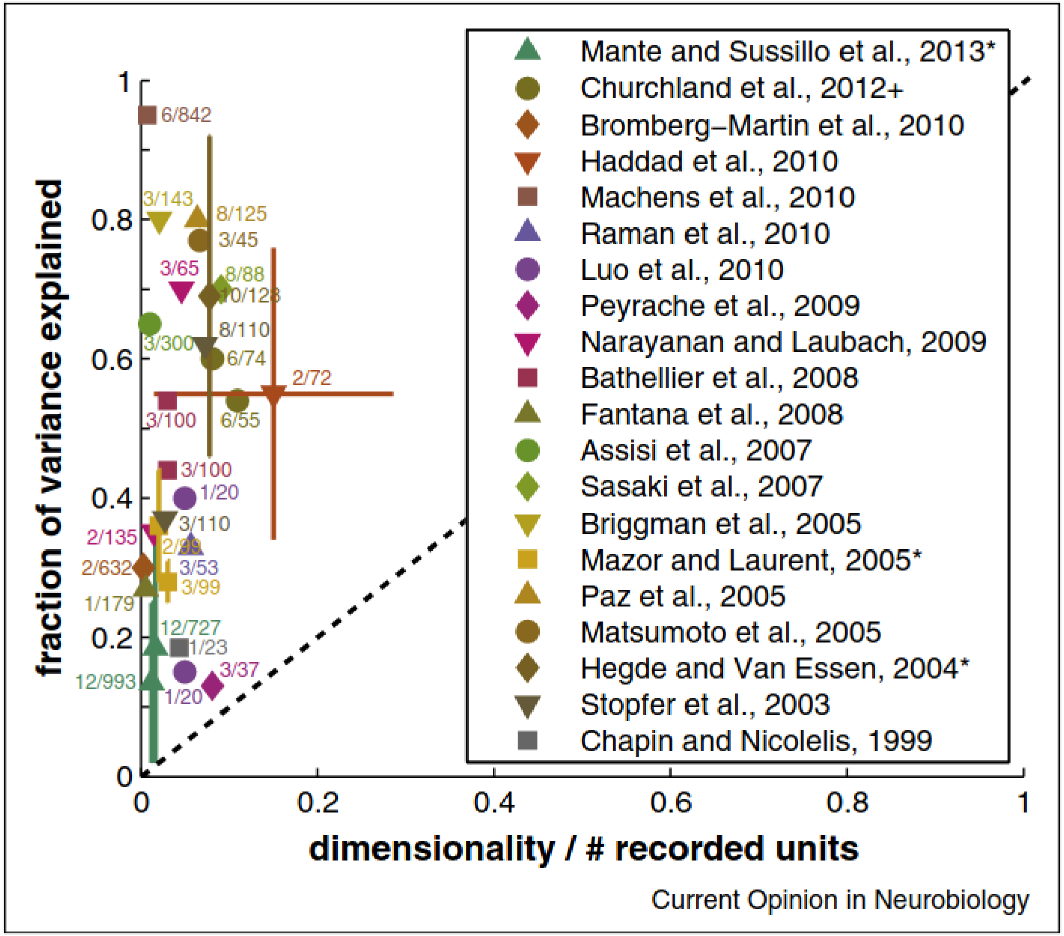
Dal Maschio, M., Donovan, J. C., Helmbrecht, T. O., & Baier, H. (2017). Linking neurons to network function and behavior by two-photon holographic optogenetics and volumetric imaging. Neuron, 94(4), 774-789.



Low dimensionality of neural codes

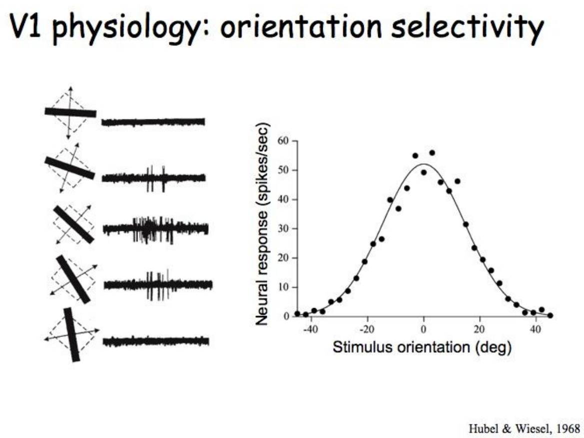
In many experiments the intrinsic dimension is much smaller than the number of recorded neurons

Gao, P., & Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. Current opinion in neurobiology, 32, 148-155.



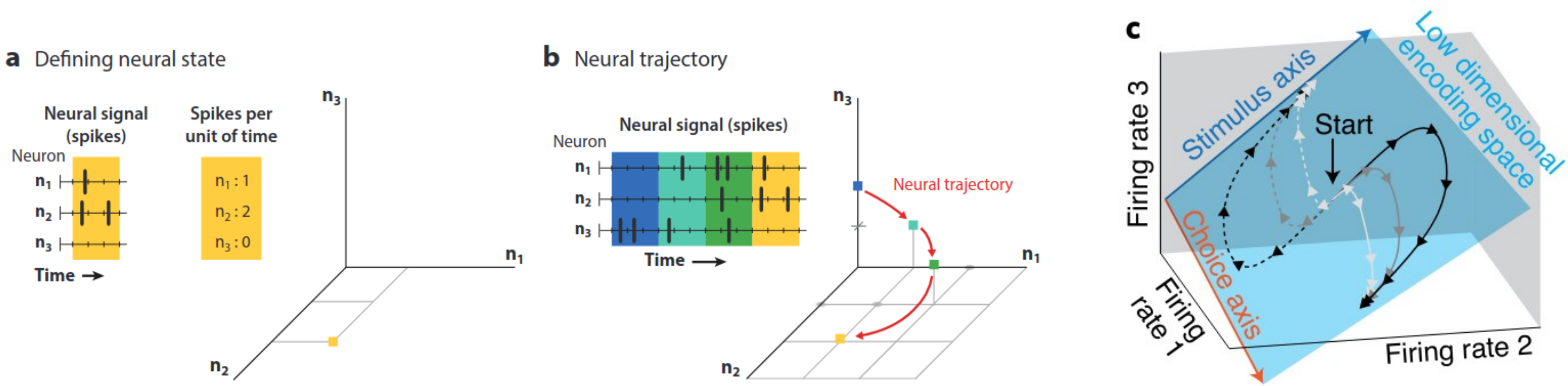
Computation through neural population dynamics

Neuron code: single neurons represent and process specific features ('selectivity')



Information processing is performed through trajectories in neural space

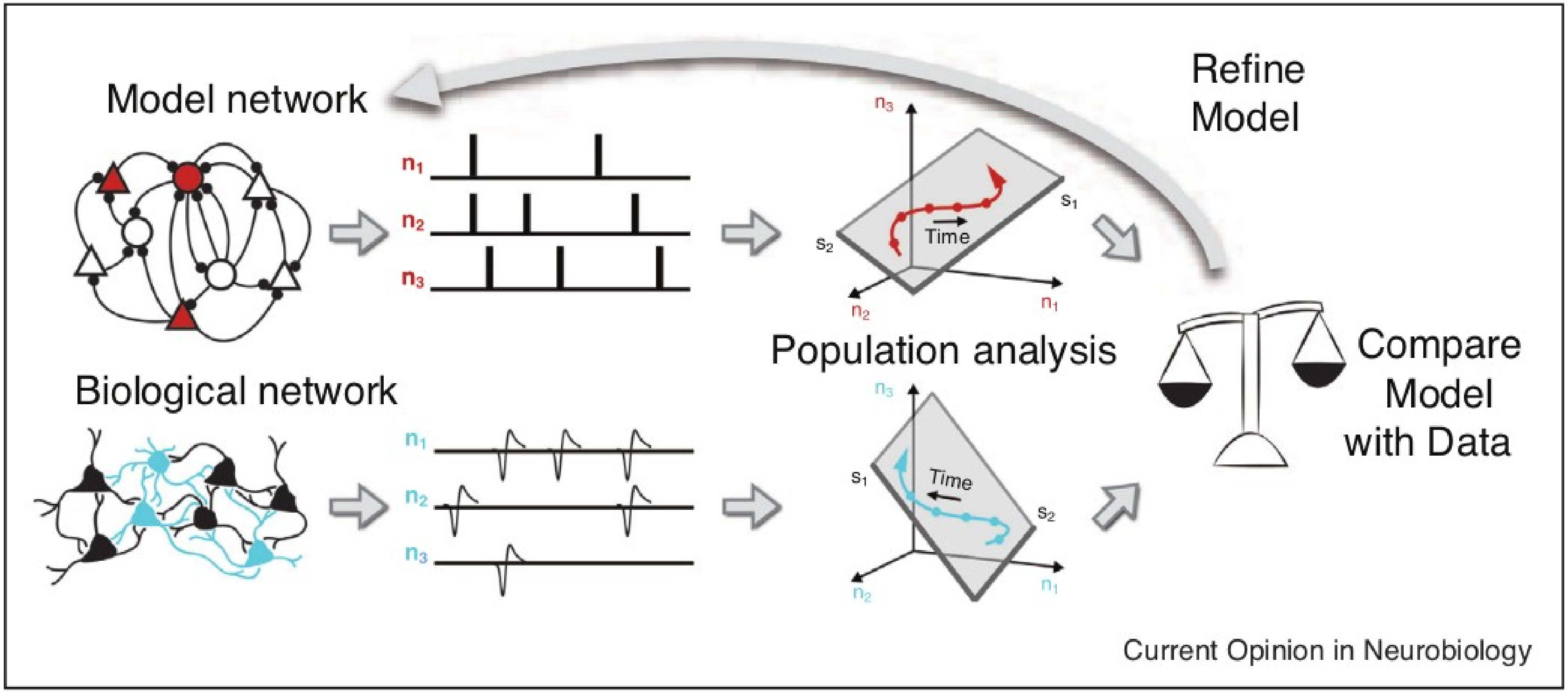
Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. Annual Review of Neuroscience, 43, 249-275.



Computation through neural population dynamics

The dimensionality depends on the computational structure of the task and can be predicted through artificial recurrent neural networks

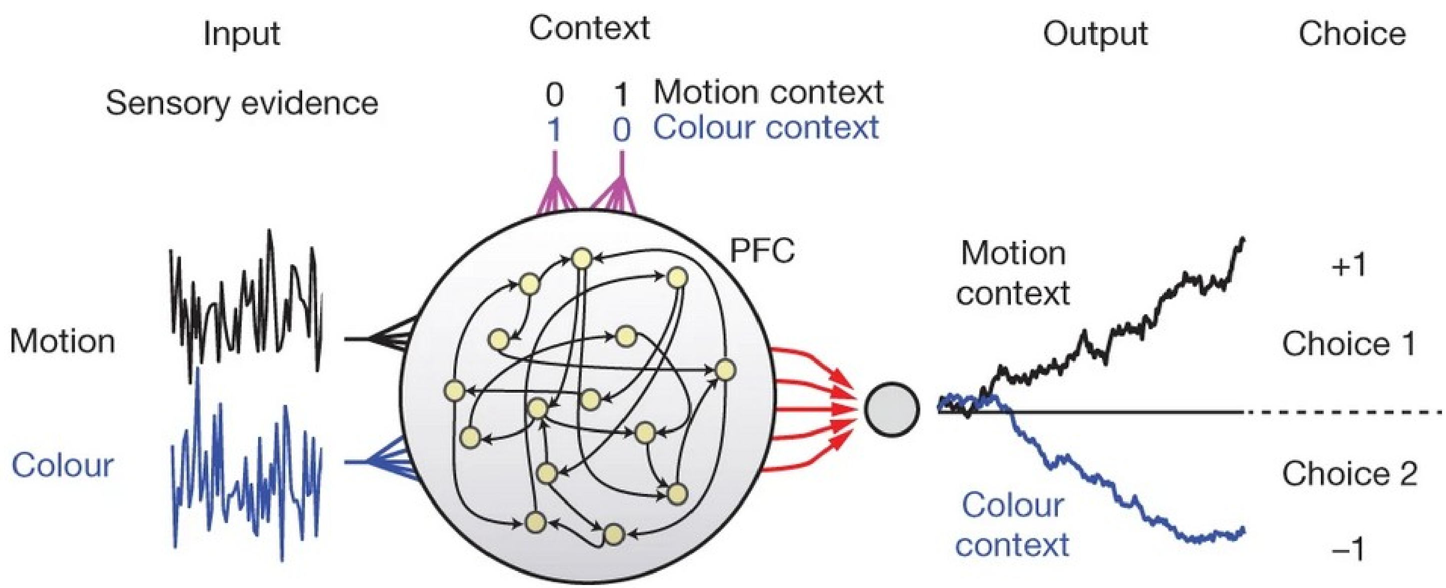
Williamson, R. C., Doiron, B., Smith, M. A., & Byron, M. Y. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. Current opinion in neurobiology, 55, 40-47..



Convergence between biological and artificial network at the level of low-dimensional dynamics

Artificial neural network models

Maass, W., Natschlager, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. Neural Comput. 14, 2531–2560 (2002).



Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. nature, 503(7474), 78-84.

[The artificial neural network models can be thought of as models of a sufficiently downstream areas where decision making occurs]

What determines the dimensionality of neural code?

1) The minimum dimensionality is determined by task computational structure

Intrinsic connectivity and input patterns determine the dimensionality in recurrent neural networks

Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2020). Complementary roles of dimensionality and population structure in neural computations. Biorxiv.
Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F., & Ostojic, S. (2021). Shaping dynamics with multiple populations in low-rank recurrent networks. Neural computation, 33(6), 1572-1615.

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j) + I_i^{FF}(t) + \eta_i(t).$$

dynamics

$$z_k = \frac{1}{N}\sum_{j=1}^{N} w_j^{(k)}\phi(x_j)$$

readout

$$\boldsymbol{J} = \boldsymbol{m}^{(1)}\boldsymbol{n}^{(1)T} + \ldots + \boldsymbol{m}^{(R)}\boldsymbol{n}^{(R)T}.$$

low rank connectivity

$$I_i^{FF}(t) = \sum_{s=1}^{N_{in}} I_i^{(s)} u_s(t).$$

$$\boldsymbol{x}(t) = \sum_{r=1}^{R} \kappa_r(t)\boldsymbol{m}^{(r)} + \sum_{l=1}^{N_{in}} u_l(t)\boldsymbol{I}^{(l)}.$$

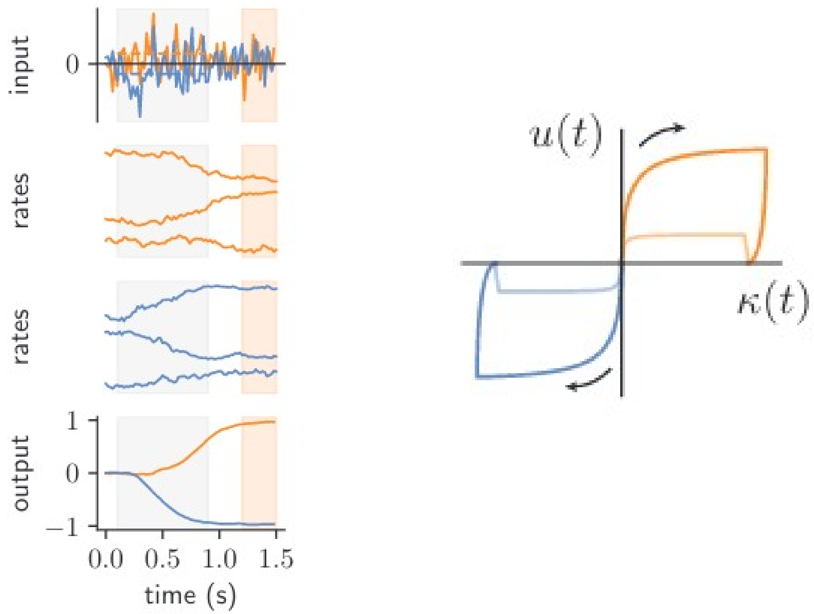
d=R+N$_{in}$

What determines the dimensionality of neural code?

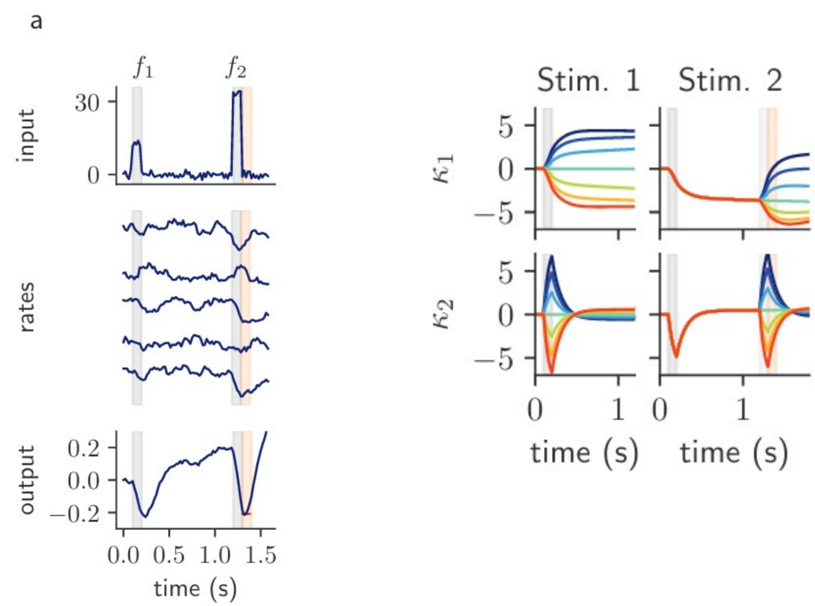
1) The minimum dimensionality is determined by task computational structure

The dimensionality depends on task complexity

Perceptual decision making d=2



Parametric working memory d=3



Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2020). Complementary roles of dimensionality and population structure in neural computations. Biorxiv.

What determines the dimensionality of neural code?

2) Dimensionality is related to readout simplicity

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature, 497(7451), 585-590.

In principle a space of dimension can store $N_c = 2^d$ binary inputs

M inputs could be stored in $d \sim \log_2 M$

However, if $d \sim M$ each input becomes easily readable

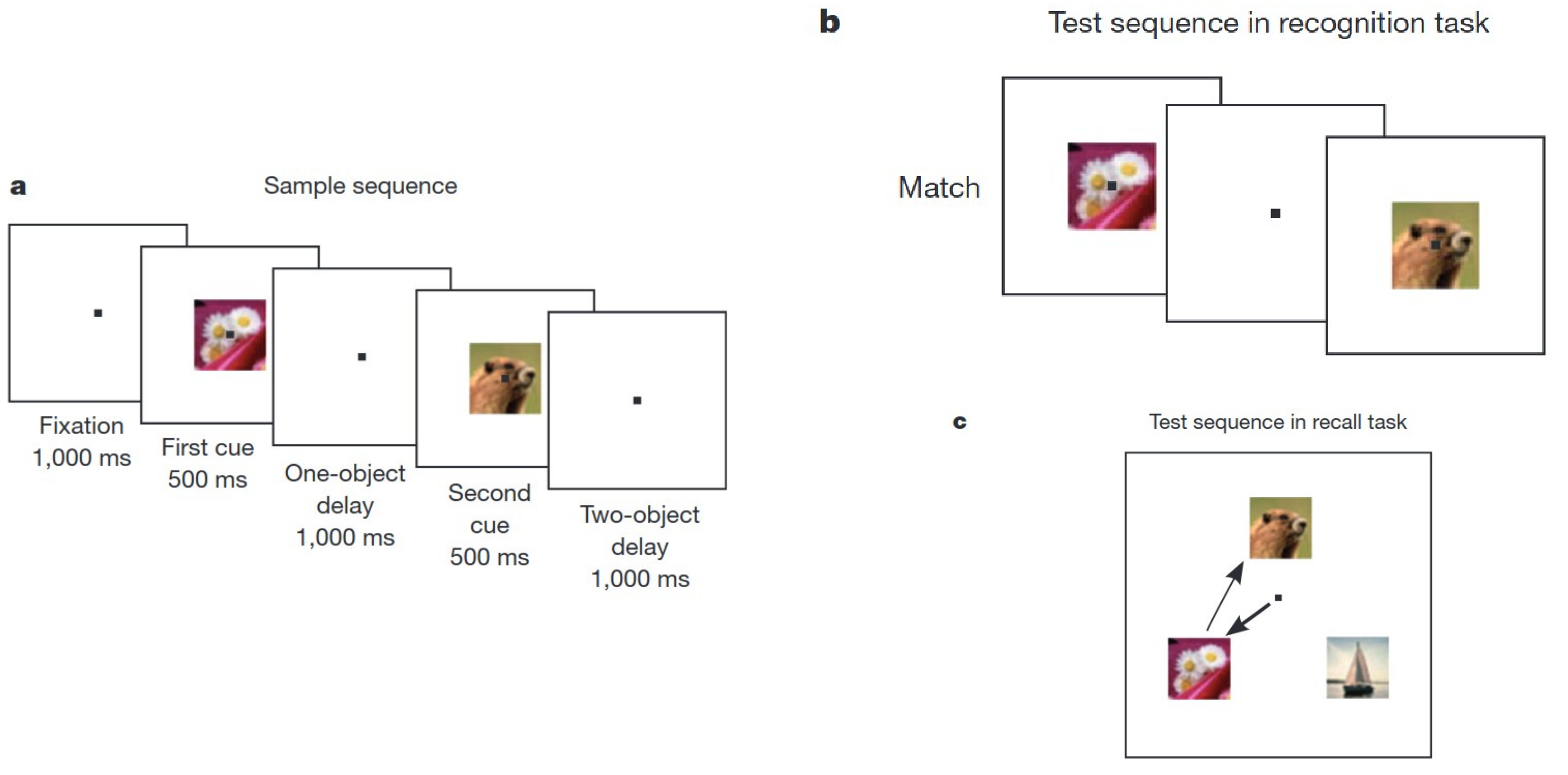
Even if final readout is low-dimensional, partial steps may require higher dimensionality for easy readout by downstream neurons

What determines the dimensionality of neural code?

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature, 497(7451), 585-590.

Delayed match-to-sample task with recall/readout

Conditions: Stim1 (4) x Stim(3) x task type(2)



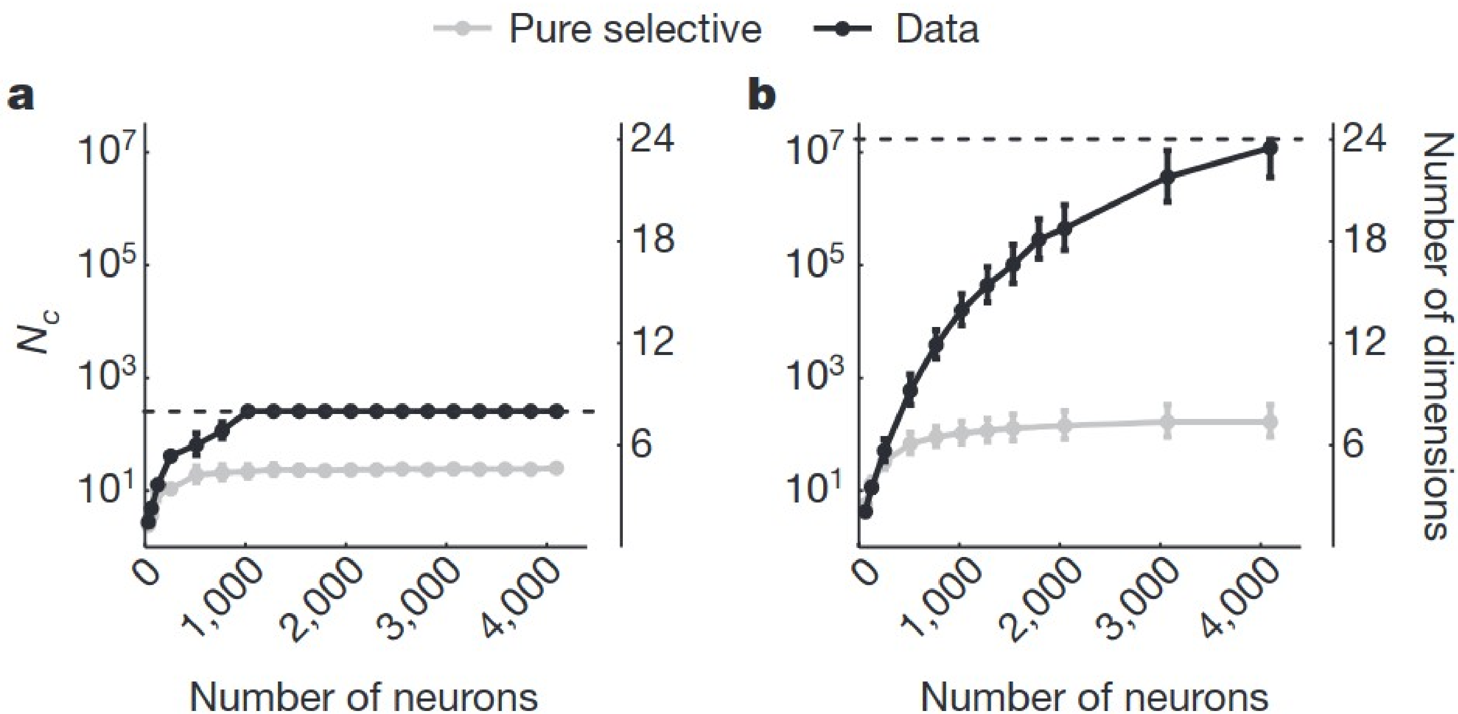
What determines the dimensionality of neural code?

Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature, 497(7451), 585-590.

Conditions: Stim1 (4) x Stim(3) x task type(2)

In principle d=5 is sufficient to encode condition
d=8 if linear responses to each stimulus *(linear selectivity)*

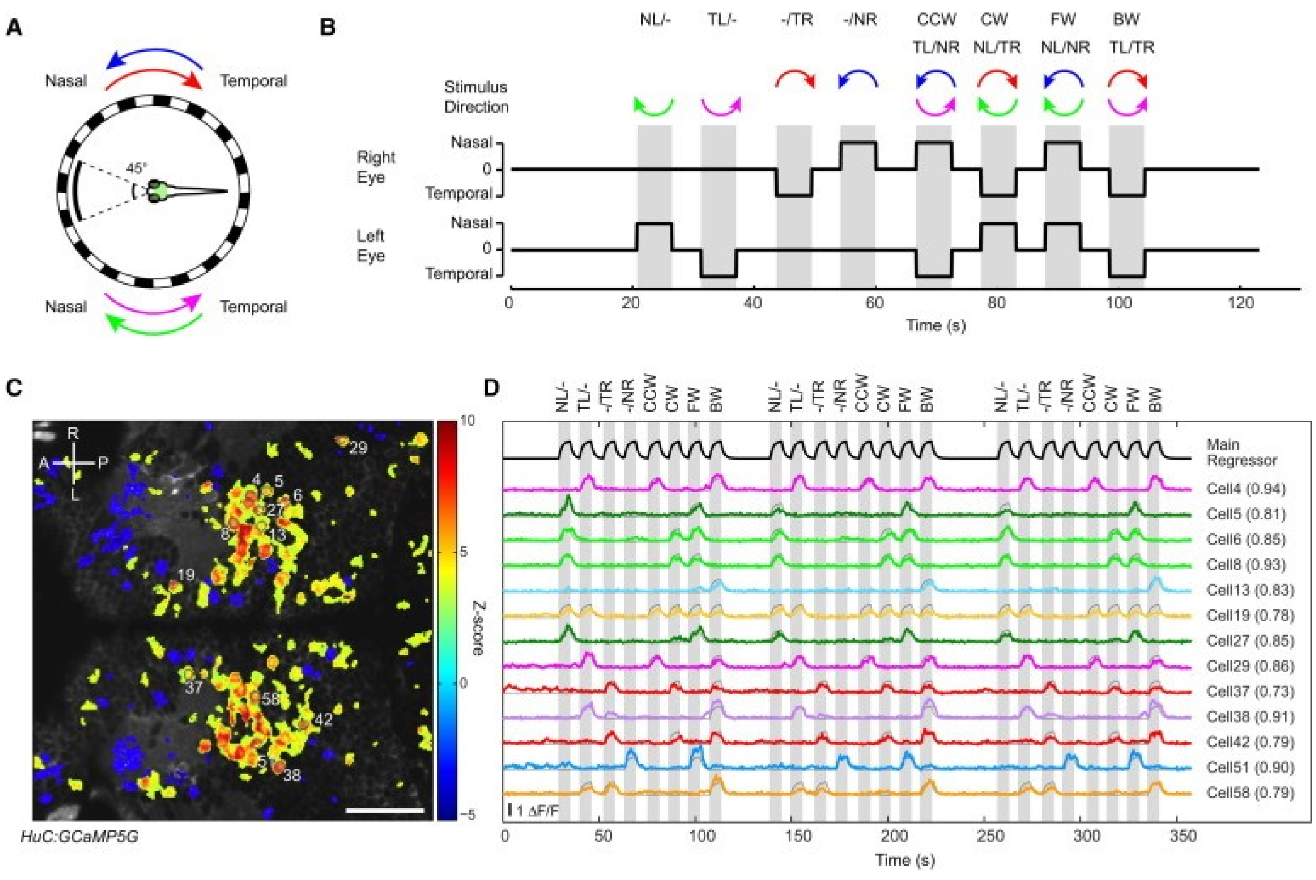
neural activity has dimension d=24



Dimension of neuron activity in zebrafish

Dimensionality is related to readout

Kubo, F., Hablitzel, B., Dal Maschio, M., Driever, W., Baier, H., & Arrenberg, A. B. (2014). Functional architecture of an optic flow-responsive area that drives horizontal eye movements in zebrafish. Neuron, 81(6), 1344-1359.



Second part: program sketch

- Collect whole-brain neural activity of D. rerio during visuomotor task
- Compute ID of neural activity and its spatial variations across regions
- Compare temporal ID variations across task/rest and investigate whether ID modulates task performance
- Replicate task in recurrent neural network
- Compare ID in biological network and in artificial network
- ...

Acknowledgements

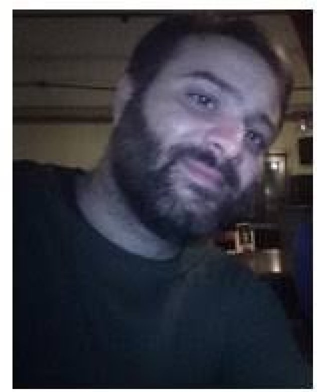
Alex Rodriguez

Alessandro Laio

Elena Facco

Antonietta Mira

Marco Dal Maschio



Pietro Rotondo

Acknowledgements

PADOVA neuroscience CENTER

Dipartimento di Fisica e Astronomia Galileo Galilei

1222·2022 800 ANNI

Università degli Studi di Padova

Sebastian Goldt

SISSA DATA SCIENCE
Machine Learning for the Natural Sciences

ID estimation: scale dependence