Sensory input to cortex encoded on low-dimensional periphery-correlated subspaces

Andrea K Barreiro[1], Cheng Ly[2], Prashant Raju[3], Shree Hari Gautam[3], Woodrow L Shew[3]

[1]Department of Mathematics, Southern Methodist University, Dallas, TX 75275, USA

[2]Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA 23284, USA

[3]UA Integrative Systems Neuroscience, Department of Physics, University of Arkansas, Fayetteville, AR 72701, USA

Abstract: As information about the world is conveyed from the sensory periphery to central neural circuits, it mixes with complex ongoing cortical activity. How do neural populations keep track of sensory signals, separating them from noisy ongoing activity? Here we show that sensory signals are encoded more reliably, with less noise in certain low-dimensional subspaces. These coding subspaces are defined by correlations between simultaneously recorded neural activity in primary sensory cortex and upstream sensory brain regions; the most correlated dimensions were best for decoding. We analytically predicted and experimentally confirmed that coding subspaces can be further improved when defined based on populations with lower noise correlations between cortex and upstream regions. We show that this principle generalizes across diverse sensory stimuli in the olfactory system and the visual system of awake mice. Our results suggest the cortex may multiplex different functions by executing them in different low dimensional subspaces.

Significance statement: Neuroscience is a cottage industry; while one study aims to understand how neurons encode function X, a separate study seeks neurons that encode for function Y, etc. However, the brain does multiple things simultaneously. The cottage industry approach leaves unclear how the brain performs functions X and Y simultaneously, sometimes using the same neurons. Here our results suggest that the same neural circuit can perform different functions by performing them in different "subspaces". We show that the brain may take advantage of correlated interactions between sensory cortex and upstream regions to define these subspaces.

Introduction

Neurons in primary sensory cortices are involved in diverse aspects of brain function; their activity is not limited to encoding sensory signals[1–4]. It is becoming increasingly clear that primary sensory cortex is a multiplexed mess of cross talk and multipurpose signals. For example, neuronal activity in primary visual cortex (V1) does not just encode physical features of visual stimuli, but is also related to locomotion[5,6], whisking and pupil diameter[7], forepaw manipulations[8], decision making[8–10], and learned consequences (rewards) of the visual stimuli[11]. Similarly, neurons in primary olfactory cortex (piriform cortex, PC) go beyond odor coding, exhibiting activity related to spatial navigation[12], thirst[13], decision making[10], and working memory[14], and can drive distinct behaviors[15]. In general, involvement in these diverse 'non-sensory' functions will vary across repeated trials of a sensory stimulus. Thus, it is not surprising that the responses of single cortical neurons to a repeated sensory stimulus vary greatly from trial to trial, often making the stimulus identity impossible to decode accurately with a single neuron. How does the brain reliably keep track of sensory signals when they are mixed into the complex, multipurpose dynamics of the cortex?

Here we propose a population-level solution to this problem. We start from the fact that at the sensory periphery, neuronal activity is purely sensory and not mixed with other functions. As the signal climbs the sensory hierarchy from the periphery to cortex, it becomes increasingly mixed with non-sensory signals due to increasingly recurrent interactions with other brain regions[16,17]. It stands to reason that sensory signals in thalamic nuclei or olfactory bulb could be less noisy (closer to purely sensory) than sensory signals in cortex. Consistent with this, dorsal lateral geniculate nucleus (LGN), which provides input to V1, exhibits response to visual stimuli that has lower dimensionality than V1[18] and is less affected by locomotion than response in V1[6]. Similarly, LGN firing is modulated more by sensory input and less by behavioral context compared to V1[16]. Likewise, olfactory bulb (OB), which innervates PC, is often less noisy than PC. For example OB has more neurons that are clearly responsive to olfactory stimuli compared with PC[19]. Thus, we hypothesized that certain coding subspaces in cortex – those that share variability with subspaces in upstream sensory regions (thalamus or OB) – may contain sensory signals with less noise.

What do we mean by a coding subspace? Considering a population, rather than single neurons[20,21], we treat the single-trial response of N cortical neurons as a vector in an N-dimensional space; the sixth component the vector is the response of the sixth neuron, and so on. The responses to many repeated trials of two different stimuli can be represented as two clouds of points in N-dimensional cortical space, one point for each trial, one cloud for each stimulus type. The spread of each cloud of points reflects the trial-to-trial variability (the non-sensory signals discussed above) and overlap of the two

clouds makes decoding the stimuli difficult. However, if the response variability due to non-sensory noise lies along different directions than the variability due to switching the sensory signal, then decoding can be greatly enhanced by projecting the N-dimensional response onto a coding subspace, i.e. a lower dimensional subspace that excludes some noise. Our hypothesis here is that such coding subspaces can be found by considering signal correlations and noise correlations between the cortical population and upstream populations.

Recent studies have adopted conceptually related approaches demonstrating that high dimensional neural circuits may manage multiple operations by performing them in different subspaces. For instance, neurons in mouse auditory cortex 'rotate' sensory representations from a sensory subspace to a memory subspace over time[22]. Neurons in rat posterior parietal cortex use different subspaces to represent decision and movement[23]. Monkeys making a choice about motion and color of a visual stimulus, exhibited neurons in PFC that used three different subspaces to encode color, motion, and choice[24]. Similarly, working memory and movement planning are separated into different subspaces within a population of lPFC neurons in monkeys[25]. In anesthetized monkeys, signals are transmitted between V1 and V2 in a 'communication subspace'[26]. The outputs of mouse cerebellar neurons were shown to represent quiescent and active behavioral states in orthogonal subspaces[27]. Computational models together with human brain imaging suggests that orthogonal subspaces are used to represent different task variables in an image classification task[28]. Our work here extends these ideas, establishing sensory subspaces in cortex and in the sensory brain regions that provide input to cortex and new ways to find these subspaces.

Projecting high-dimensional activity into a lower dimensional coding subspace is a type of dimensionality reduction. More generally, dimensionality reduction has long been recognized and used to improve decoding of sensory signals with supervised pattern classification techniques like linear discriminant analysis (LDA) [29,30]. However, LDA and similar techniques are biologically implausible coding strategies because these algorithms require information about which stimuli caused which response; the brain is not privy to stimulus identities. Without information about stimulus identities, the brain would have to blindly try many (infinite) possible low dimensional projections to find the optimal one, which is implausible. Here we identify a more biologically plausible strategy; we show that low dimensional coding subspaces can be found without knowledge of stimuli identities by considering correlations between cortex and upstream brain areas that provide input to cortex. Using canonical correlation analysis (CCA, see refs [21,31] for an introduction), we define subspaces in cortex and subspaces in LGN or OB in which responses to stimuli are most correlated across the brain regions. We show that these cross-population correlated subspaces can effectively separate signal from noise, often approaching theoretical limits of optimal decoders (like LDA). We developed an analytical approach to better understand these coding subspaces and successfully predicted improved coding subspaces among neurons with low cross-population noise correlations.

Results

We analyzed extracellular recordings of spiking activity of single neurons in the visual and olfactory systems of awake mice (Fig 1, originally reported in [16,19]) and analytically studied a simplified theoretical two-population model. We first present results from the experimental data. Relatively recent advances in recording technologies allowed us to test our hypothesis which requires simultaneous recordings from large populations in primary sensory cortices and upstream brain regions that provide input to cortex – V1 and dLGN for the visual system, PC and OB for the olfactory system. In the Bolding-Franks data set[32] (n=8 mice), the number of recorded neurons was 27±6 (mean±SD) in OB and 48±11 in PC. We analyzed responses to six different odorants (fixed concentration) with 15 trials each (Fig 1B), considering orthonasal response (following inhale) separately from retronasal response (following exhale) as shown in Fig 1C. In the Allen data set (n=9 mice), the number of recorded neurons was 61±27 in LGN and 245±48 in V1. We analyzed responses to static gratings with 6 different orientations and 5 different spatial frequencies with 48 trials each (Fig 1I). For both the olfactory and visual responses, we defined response as the spike count in the 250 ms period following stimulus onset (Fig 1C,J). As a baseline comparison for our main results below, we first assessed how well single neuron responses decoded pairs of stimulus types (22 pairs for the Bolding-Franks data (Fig 1b), 25 pairs for the Allen data (Fig 1I)). Decoding accuracy D was quantified as the fraction of correctly classified trials using the optimal threshold. Generally, OB neurons outperformed PC neurons (Fig 1D), but PC, dLGN, and V1 neurons decoded poorly, not much above chance-level D=0.5, when averaged over all neurons (Fig 1K).

In what follows, we describe higher dimensional population coding strategies to improve upon this somewhat poor decoding at the single neuron level. Before doing this, it is prudent to first ask what is the effective dimensionality of the recorded responses and how does it depend on how many trials or types of stimuli were presented? Since many neurons have covarying responses and covarying ongoing activity, a population of N neurons is likely to occupy a manifold with less than N dimensions[7,33–35]. We assessed this using the participation ratio (PR) which quantifies how many dimensions are needed to capture approximately 80% of variability based on the eigenvalues of the covariance matrix[33]. Generally, such assessments of dimensionality could depend on the number of neurons, the number of stimulus trials and the number of types of stimuli for evoked activity, or the number of samples for ongoing activity. For a fixed number of trials or samples, we found that dimensionality (PR) grows with the number N of neurons, but is much less than N (one example mouse shown in Fig 1E and another in Fig 1L). For fixed N=18, dimensionality tends to increase rapidly and then plateau below 10 for increasing number of trials (including all stimulus types) or increasing ongoing samples (Fig 1F,M and Fig S1). However, for fixed N=18 and fixed number of trials (30), response dimensionality is relatively independent of how many types of stimuli were presented (Fig 1G,N and Fig S1).

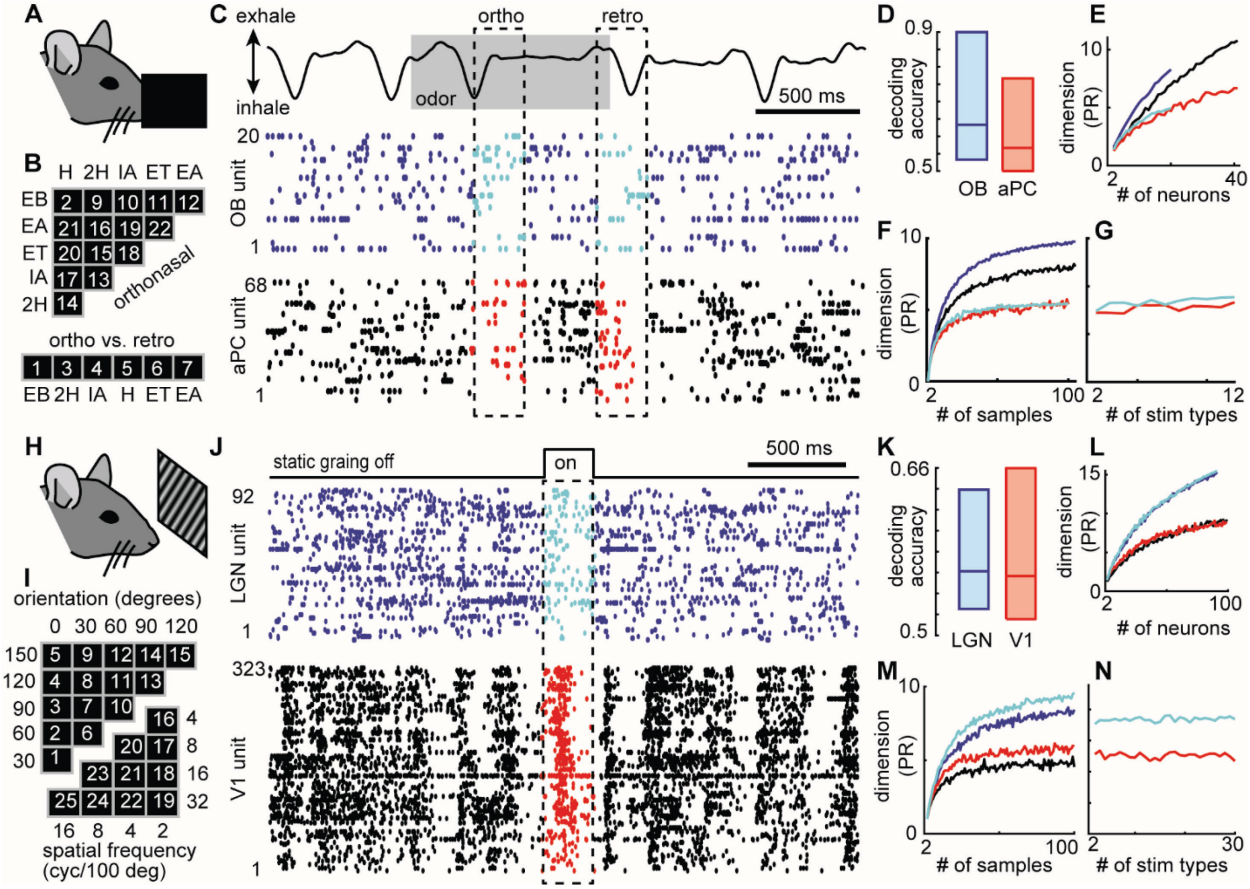

Figure 1. Noisy multi-region response to olfactory and visual sensory input. A,B) *Awake, freely breathing mice were presented with 6 different odors, 15 times each, through a controlled air flow nose-cone (reported first in*[19]*).* ***C)*** *For each OB and PC neuron, we quantified response as spike count during 250 ms (dashed box on raster plots) following peak inhale (orthonasal) or peak exhale (retronasal).* ***D)*** *Single neuron decoding accuracy was typically worse in PC than in OB. Box plots span quartile range. Line indicates median.* ***E)*** *Dimensionality (participation ratio, PR) of ongoing activity (blue OB, black PC, 30 samples) and stimulus evoked activity (cyan OB, red PC, 30 trials) increases with number of neurons, but remains much lower than the number of neurons in the population.* ***F)*** *Dimensionality increases with the number of samples or trials, but saturates near between 5 and 10 (population size fixed at 18, included all types of stimuli).* ***G)*** *Dimensionality is weakly dependent on the number of types of stimuli (population size fixed at 18, number of samples/trials fixed at 30).* ***H,I)*** *Awake mice viewed static gratings with 6 different orientations, 5 different spatial frequencies, 42 reps each, for 250 ms (reported first in*[16]*).* ***J-N)*** *Same as panels C-G, but response time window began at onset of visual stimulus.*

A central idea underpinning our approach is that the brain can improve decoding accuracy of single neurons by projecting neural activity onto a subspace which excludes some of the “noise” that compromises decoding. We hypothesized that we could identify such decoding subspaces based on inter-regional correlations between cortex and upstream extracortical regions (dLGN or OB). To demonstrate how this might work, we first present a simple, instructive case based on simulated data that allows visualization: two neurons in cortex and two neurons in the upstream region (Fig 2). In the presented example, the simulated responses are drawn from a multivariate Gaussian distribution (Method) with parameters chosen such that the two cortical neurons have strong noise

correlations and a small difference in mean response for the two stimuli. The two extracortical neurons have noisy overlapping responses to the two stimuli. (In this example, there are no cross-population noise correlations, which is important for our approach, as we discuss further below.) All four of these neurons are rather poor decoders at the single neuron level, but decoding improves dramatically when projected onto a particular subspace (green line in Fig 2). The optimal subspace can easily be found using LDA (dashed line in Fig 2 is the LDA classification boundary), but LDA requires knowledge of the stimulus identities, which is "cheating"; the brain does not have access to stimulus identities before they are decoded. The optimal subspace can also be found, without knowledge of stimulus identity, by performing CCA, which is the key to the findings in this paper.

The first CCA component (green line in Fig 2) is, by definition, the direction along which the cortical fluctuations are most correlated with the extra-cortical fluctuations. In this example, the first CCA component is also aligned with the direction along which signal varies most, thus identifying the optimal decoding subspace (orthogonal to the LDA classification boundary). Note that, in the cortical population, the first CCA component is also orthogonal to the direction along which noise fluctuations have the greatest variance (black line in Fig 2, determined by principal component analysis (PCA) on the trials for a single stimulus type), thus separating signal from noise. A more extensive analytical study of this 2x2 case with a wide variety of possible correlations among the 4 neurons and signal-to-noise scenarios is described further below and in Supplementary Materials.

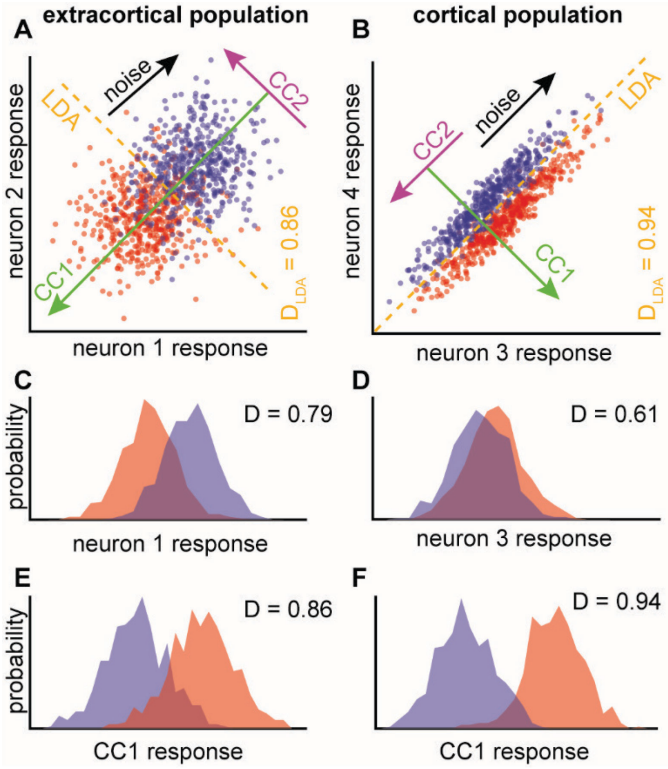

Figure 2. How CCA finds optimal coding subspace: simulated 2x2 dimensional case. A) *Each point represents the response of neurons 1 and 2 to one of two stimuli (red and blue). An optimal linear decoder (LDA) achieves 86% decoding accuracy (D). CCA on the two populations identifies two directions (CC1 and CC2). PCA finds the principal direction of variability due to noise (black).* ***B)*** *Same as panel A, but for neurons 3 and 4 in the cortical population. Optimal LDA decoder achieves 94% accuracy.* ***C,D)*** *Response distributions for the two stimulus types overlap substantially for neurons 1 and 3, resulting in suboptimal decoding accuracy (79% for neuron 1, 61% for neuron 3).* ***E,F)*** *When projected onto CC1, the response distributions better separate the two stimulus types, achieving optimal decoding accuracy (same as LDA).*

Next, we extend our approach to higher dimensions (more neurons) and apply it to the data described in Fig 1. For the case with N cortical neurons and N neurons in the extra-cortical region, can we project the recorded responses onto a coding subspace that improves decoding? Again, we turn to CCA, which identifies N canonical components, i.e. N directions in cortical space, and another N directions in the extra-cortical space[21,31,36]. The canonical components are ranked according to how strongly activity is correlated across the two regions when projected onto each CCA direction. We performed CCA on stimulus-evoked responses recorded from N=18 neurons from cortex and the same number from the extracortical region. (We chose 18 because that was the smallest number of neurons recorded from a single brain region in all the recordings we analyzed). We performed CCA on the responses to one pair of stimulus types at a time, thus allowing for the likely possibility that different pairs of stimulus types may be encoded in different subspaces. We considered 25 pairs of visual stimulus types (Fig 1B) and 21 pairs of olfactory stimulus types (Fig 1I). We first projected the 18-dimensional activity onto 1-dimensional subspaces defined by single CCA components, like the example in Fig 2, but with an 18-to-1 dimensionality reduction instead of the 2-to-1 reduction in Fig 2. For example, the response distributions for two gratings with different spatial frequencies (fixed orientation) are shown for 18 V1 neurons and 18 LGN neurons, before and after projection onto the first canonical component (Fig 3A). After projecting onto the same 1-dimensional subspace, all neurons necessarily have the same response distribution (up to a change in sign) and therefore all neurons have the same decoding accuracy after projection onto CC1. Considering all experiments and all units, for both the visual system and the olfactory system, we found that projection onto a single CC resulted in significantly improved decoding, on average (Fig 3C, Kolmogorov-Smirnov test, $p<10^{-5}$). This improvement was greatest, on average, for the first few CC's, i.e. in the subspaces with the greatest inter-regional correlation (Fig 3C insets).

In Fig 3A, the decoding accuracy of all 36 neurons improved in the CC1 subspace. However, in other experiments sometimes a different CC was better than CC1 or in some cases the best single unit was a very good decoder before any projection and was not improved by projecting onto any CC (Fig 3B,C). Next, we asked which CC's beat the best single unit and in how many experiments this occurred. In the visual system, at least one CC beat the best single unit in 60% of experiments in LGN and 37% of experiments in V1 (total of 225 experiments = 9 mice x 25 stimulus pairs). In the olfactory system, there were fewer experiments in which a single CC beat the best unit (14% in OB and 36% in PC). This was because a small subset of the single units were very good decoders in these experiments, especially in OB. For the visual system, CC2 and CC3 beat the best single unit most often. For the olfactory system, CC1-CC7 beat the best single unit more than higher CCs. In Fig 3D summarizes which CC's beat the best single unit and how often this occurred for all experiments.

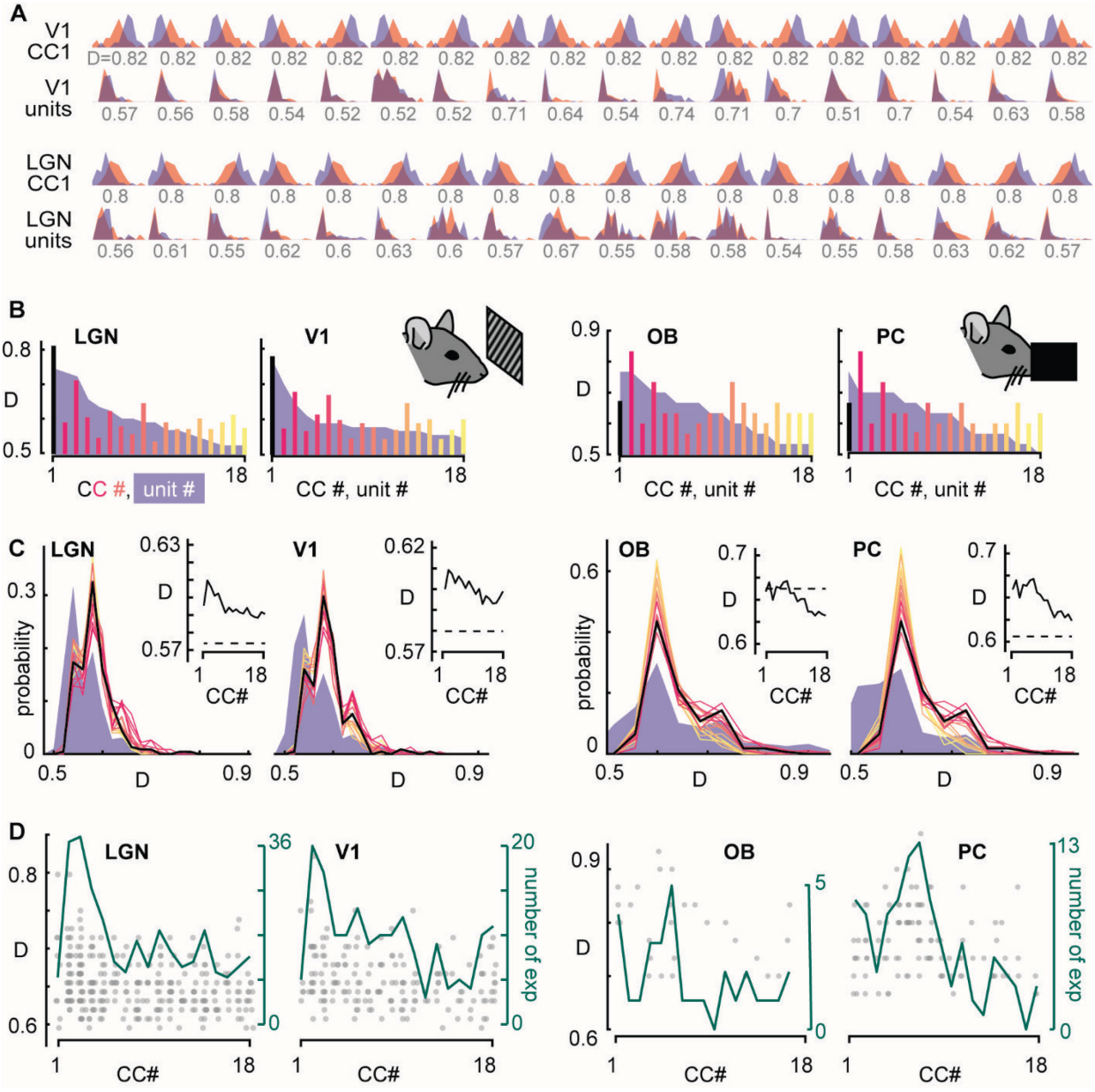

Figure 3. Decoding accuracy improved in 1-dimensional subspaces defined by inter-regional correlations. A) *Example case, showing response distributions to two different visual stimuli (red, blue) for 18 LGN units and 18 V1 units. Projection of responses onto CC1 better separates responses, improves decoding accuracy (D).* ***B)*** *In this example (left, same as panel A) only CC1 (black) beats the best single neuron (peak of blue shading). Example (right) from olfactory system in which only CC2 decoding beats the best single unit.* ***C)*** *Distributions of decoding accuracy (D) for all experiments, all units show that, on average, decoding is improved most for first few CC's. Color code indicates CC# as in panel B. Inset shows that mean of distributions decreases with CC#.* ***D)*** *Each gray point represents one experiment for which decoding accuracy (left axis) of the CC beat the best single unit. Green line represents the number of experiments (right axis) for which the CC beat the best single unit. For the visual system (left two panels), CC2-CC4 exceeded the best single unit most often. For OB, the best single unit was often a very good decoder; the CC's beat the best unit in relatively few experiments. For PC, CC6 and CC7 beat the best single unit most often.*

The results in Fig 3B suggest that more than one CC may contain useful information for decoding sensory input. For instance, the PC example in Fig 3B shows that CC2 and CC4 were both decent decoding dimensions. We next asked whether greater improvements in decoding can be found by projecting onto subspaces with more than one dimension. The example in Fig 4A shows response distributions for 36 neurons before and after projection onto a 4-dimensional subspace, defined by the four CC's which individually had the top four decoding values when considered as single dimensions. Note that when projecting different single neuron responses into a subspace with more than one dimension, the response distributions are not identical; some neurons improve more than others. This often resulted in further improvements in decoding compared to what we found for 1-dimensional subspaces (Fig 4B). We systematically considered subspaces ranging from 1 up to 17 dimensions. (We also considered 18 dimensions, but this involves no dimensionality reduction since we worked with 18 neurons and is therefore not a subspace and is equivalent to the original data). For each n-dimensional subspace, we defined the subspace as that spanned by the n CC's with the n best decoding accuracies when considered individually. Considering distributions of all neurons, all stimulus pairs, and all animals (Fig 4C), we found that all subspaces (with $n<18$) resulted in improved decoding, on average. Comparing to the best original single units, we found that nearly all cases (99% in LGN, 91% in V1, 81% in OB, 95% in PC, across all stimulus pairs and animals) had at least one single unit that increased its decoding accuracy above that of the original best unit after projection onto the coding subspace. Moreover, compared to high dimensional subspaces, low dimensional subspaces usually had more units that beat the best original units (Fig 4B, C, D).

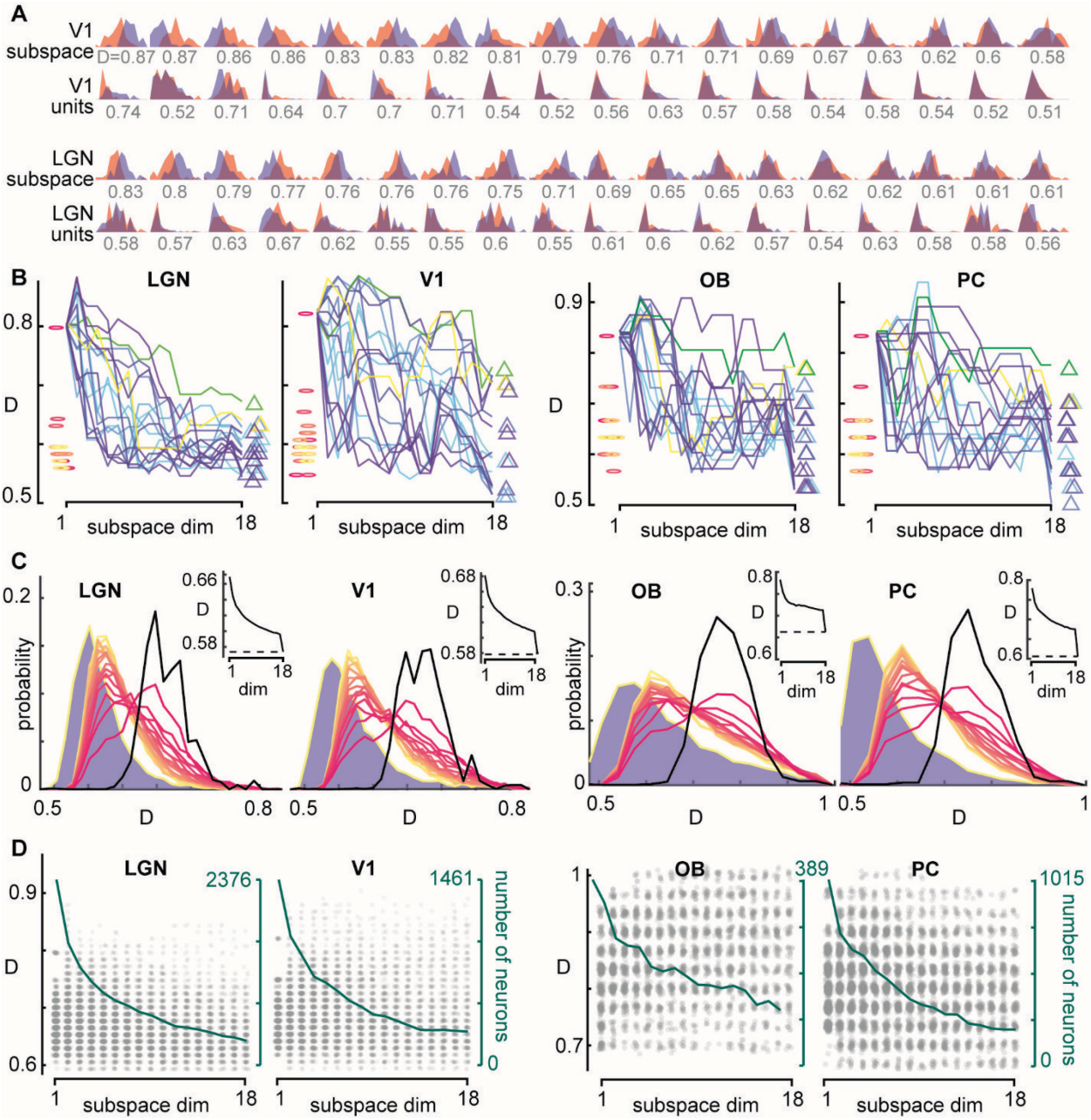

Figure 4. Decoding accuracy improved further for higher dimensional subspaces. A) *Example case, showing response distributions to two different visual stimuli (red, blue) for 18 LGN units and 18 V1 units. Projection of responses onto a 4 dimensional subspace better separates responses, improves decoding accuracy compared to single neurons before projection. (D).* ***B)*** *Summary of subspaces with different dimension for one visual system experiment (left) and one olfactory system experiment (right). Each line represents one neuron. An n-dimensional subspace was defined by the n CCA components with the n highest D considered singly. Pink markers on left side indicate D for one dimensional CC projections. Triangles on right side indicate original single unit D. Note that 2-5 dimensional subspaces beat many single units and the best singe CC dimension. Color of lines are arbitrary, except green and yellow indicate best and second best original single units.* ***C)*** *Distributions of decoding accuracy (D) for all experiments, all units show that, on average, decoding is improved most for low dimensional subspaces (color: black-1D, red-yellow 2D-18D). Inset shows that mean of distributions decreases with subspace*

dimensionality. ***D)*** *Each gray point represents one neuron for which decoding accuracy (left axis) after projection onto subspace, beat the best original single unit. Green line represents the number of neurons (right axis) for which the CC beat the best single unit. For the visual system (left two panels), CC2-CC4 exceeded the best single unit most often. For OB, the best single unit was often a very good decoder; the CC's beat the best unit in relatively few experiments. For PC, CC6 and CC7 beat the best single unit most often.*

Improvements in decoding accuracy after projection onto the subspaces defined by CCA varied; some experiments and some neurons had greater improvements than others. What factors determine whether the subspace defined by CCA is a good coding subspace or not? We next sought to answer this question in a more controlled and analytically tractable scenario with just 2 neurons in each of two populations. We assumed that each neuron had responses to two different stimuli that were drawn from multivariate gaussian distributions. We further assumed that the responses of the 4 neurons are governed by 14 parameters: 8 mean responses (2 stimuli x 4 means), 4 variances, 1 within-population covariance, and 1 cross-population covariance (Fig 5A). Variances and covariances were assumed to be the same for the two stimuli. We considered 1000 different configurations of these 14 parameters, drawn randomly (Methods). We excluded parameter sets for which the best single neuron was already an excellent decoder (D>0.99), because these neurons cannot be improved, leaving 846 configurations. Fig 5B shows the distribution of D for single neurons before any dimensionality reduction. As expected, after projecting onto the best of the two possible CC's decoding is substantially improved (Fig 5B). To better understand these improvements, we next analytically computed the optimal decoding for each parameter set and compared to the decoding after projection onto the CC's (Fig 5C). We note that here we developed analytical methods to compute the CC directions; our results depend only on the 14 parameters discussed above and are independent of the number of samples (Methods).

We found that in 75% of parameter sets, the best CC reached the optimal decoding level (within 3%). In most cases (63%), the best CC was the first CC (the most correlated between the two populations). However, in many cases, projection onto the best CC still fell well below the optimal D. What causes this suboptimality? And why is the first CC not always the best? For both these questions, we found that stimulus-independent covariance across the two populations was responsible, i.e. cross-population noise correlations. Keeping all other parameters fixed, when we set the cross-population covariance to zero, projection onto the first CC reached the optimal decoding level for all 846 parameter sets (Fig 5D). Similarly, for non-zero cross-population covariance, we found that the deficit between optimal decoding and the best CC decoding was significantly correlated with the magnitude of cross-population noise correlation (Fig 5E, Spearman $\rho = 0.38$, $p<10^{-29}$).

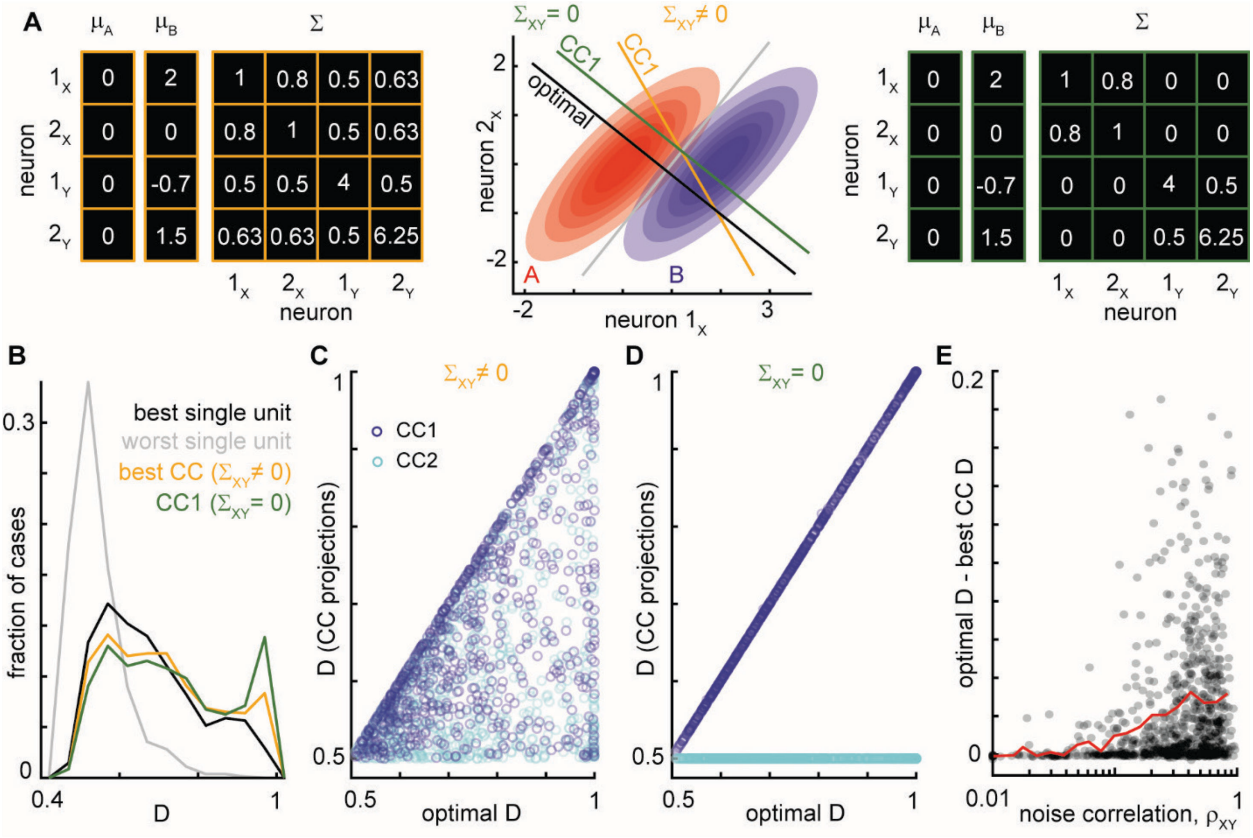

Figure 5. Cross-population noise correlations limit accuracy of coding subspaces. A) *Example parameters for population x (neurons 1_x and 2_x) and population y (neurons 1_y and 2_y) for two stimulus types (A and B). When stimulus-independent cross-population covariance between populations x and y are zero (green), the first component CC1 is the same as optimal decoding dimension. Non-zero cross-population covariance results in deviations from the optimal decoding dimension.* ***B)*** *Summary of 846 randomly chosen parameter sets for the 2x2 population. On average, decoding accuracy D is greater for CC1, even with non-zero cross-population covariance, but improved further if these covariances are set to zero (all other parameters held fixed.)* ***C)*** *Decoding along CC1 often reaches the optimal limit (63% of parameter sets).* ***D)*** *Decoding along CC1 always reaches optimal limit if cross-population covariance is set to zero.* ***E)*** *With non-zero cross-population covariance, the difference between CC1 decoding and the optimal limit was minimal for lower noise correlation across the populations. Red line – moving average of points.*

Coming back to the experimental data, these analytical results suggest that if there are noise correlations (i.e. correlations that have nothing to do with changing stimulus) between cortex and an upstream sensory region, this may limit the efficacy our coding subspaces. Moreover, the existence of cross-population noise correlations may result in the first CC being a worse coding subspace than higher CC's. Our analytical results offer a prediction; seeking coding subspaces among neural populations with weaker cross-population noise correlations, might result in even greater improvements in decoding.

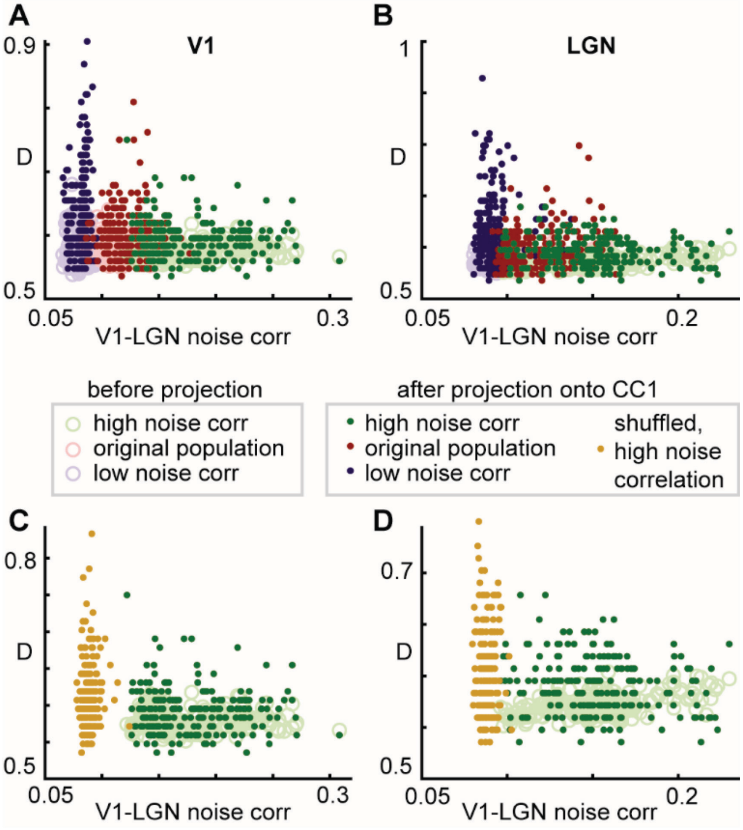

Figure 6. Predicted improvements in decoding among real neurons with low cross-region noise correlations. A,B) Projection onto CC1 improves decoding accuracy the most when CCA is performed on V1 and LGN neural populations with low noise correlations across regions. **C,D)** Similarly,shuffling responses across trials with the same stimulus type reduces noise correlations and affords greater improvements in decoding accuracy along CC1. This control was performed for the populations with high cross-population noise correlations.

Next, we set out to test this prediction. We asked, for a different population of neurons with different cross-population noise correlations, would our decoding subspaces exhibit different improvements in decoding? To make this a fair question, we also required each new population of neurons we considered to have the same average single-neuron decoding accuracy before any dimensionality reduction. In all the preceding analysis of experimental data we randomly chose 18 neurons from cortex and 18 neurons from the upstream region. To test our predictions, we took advantage of the fact that the recordings from the visual system had many more than 18 neurons to choose from for both LGN and V1. (This approach was not feasible for the olfactory system data because there were typically only 20 to 30 in the OB population, making it impossible to identify three different populations with substantially different cross-population noise correlations.) We considered two additional populations (each with 18 neurons in cortex and 18 LGN) that are chosen based on their cross-population noise correlation. One population was chosen to have high cross-population noise correlations, the other with low cross-population noise correlations (Methods). For these new populations, we repeated the same data analysis as before. In line with our predictions from the simplified 2x2 populations, we found that populations with lower cross-population noise correlations tended to have greater improvements in decoding when projected onto the first CC (Fig 6A, B). To quantify this result, we computed the correlation between D for CC1 and V1-LGN noise correlation. We found a significant negative correlation for both V1 (Pearson r=-0.26, $p<10^{-11}$) and LGN (r=-0.24, $p<10^{-9}$). As an additional test of our predictions, we also performed a control with responses shuffled across trials with a given stimulus type. This procedure also reduces noise correlations. We performed this control using the

neural population with high cross-population noise correlations (green points in Fig 6). In line with our predictions, for the visual system, we found that such shuffling increased decoding accuracy in the one-dimensional subspace defined by CC1. There was a significant negative correlation between D for CC1 and V1-LGN noise correlation ($r=-0.20$, $p<10^{-4}$) and LGN ($r=-0.20$, $p<10^{-4}$).

Discussion

We have shown that neurons in primary sensory cortices may appear to be noisy, poor decoders, but can become excellent decoders when projected onto low dimensional coding subspaces. The dimensions that make up these subspaces are those along which cortical activity is correlated with activity in upstream sensory regions. Furthermore, if we consider neurons with low noise correlations between cortex and upstream regions, these correlated subspaces become even better, reaching theoretical upper bounds for decoding accuracy for our simplified 2x2 theoretical population study.

An additional important point about the coding subspaces we study here is that, in principle, they can be found by only considering information internal to the brain, without information about the outside world. More specifically, the coding subspaces can be found using only neural activity without knowledge of which stimuli are presented on which trials. Thus, the brain could find these subspaces without a “cheating ideal observer”. However, it is still reasonable to ask, is it plausible that real neural circuits could implement the linear algebra operations like CCA and projections onto subspaces that are needed here? Theoretical studies suggest that this is possible[37–40]. Notably, a recent study showed that CCA can be implemented without biologically implausible non-local plasticity mechanisms[37].

Here we focused on sensory coding subspaces, but our findings suggest a more general principle for multiplexing many functions within the same neural circuit. Any two brain regions that cooperate to execute a particular function are likely to exhibit some correlated activity. But this shared signal is likely mixed in with other activity that is involved in other ongoing functions. Our results suggest that the common, correlated activity between the two regions can define a subspace (easily found by CCA) which effectively separates the function of interest from other ongoing functions, thus allowing the same circuits to execute many functions simultaneously.

References

1. Matyas, F. *et al.* Motor Control by Sensory Cortex. *Science (80-.).* **330**, 1240–1243 (2010).

2. Zagha, E., Casale, A. E., Sachdev, R. N. S., McGinley, M. J. & McCormick, D. A. Motor Cortex Feedback Influences Sensory Processing by Modulating Network State. *Neuron* 1–12 (2013) doi:10.1016/j.neuron.2013.06.008.

3. Ghazanfar, A. A. & Schroeder, C. E. Is neocortex essentially multisensory? *Trends Cogn. Sci.* **10**, 278–285 (2006).

4. Parker, P. R. L., Brown, M. A., Smear, M. C. & Niell, C. M. Movement-Related Signals in Sensory Areas: Roles in Natural Behavior. *Trends Neurosci.* **43**, 581–595 (2020).

5. Ayaz, A., Saleem, A. B., Schölvinck, M. L. & Carandini, M. Locomotion controls spatial integration in mouse visual cortex. *Curr. Biol.* **23**, 890–894 (2013).

6. Niell, C. M. & Stryker, M. P. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron* **65**, 472–479 (2010).

7. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science (80-.).* **364**, eaav7893 (2019).

8. Zatka-Haas, P., Steinmetz, N. A., Carandini, M. & Harris, K. D. Sensory coding and the causal impact of mouse cortex in a visual decision. *Elife* **10**, (2021).

9. Allen, W. E. *et al.* Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* **94**, 891-907.e6 (2017).

10. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).

11. Shuler, M. G. & Bear, M. F. Reward Timing in the Primary Visual Cortex. *Science (80-.).* **311**, 1606–1609 (2006).

12. Poo, C., Agarwal, G., Bonacchi, N. & Mainen, Z. F. Spatial maps in piriform cortex during olfactory navigation. *Nature* (2021) doi:10.1038/s41586-021-04242-3.

13. Allen, W. E. *et al.* Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science (80-.).* **364**, 0–10 (2019).

14. Zhang, X. *et al.* Active information maintenance in working memory by a sensory cortex. *Elife* **8**, 1–29 (2019).

15. Choi, G. B. *et al.* Driving opposing behaviors with ensembles of piriform neurons. *Cell* **146**, 1004–1015 (2011).

16. Siegle, J. H. *et al.* Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).

17. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

18. Dahmen, D. *et al.* Strong coupling and local control of dimensionality across brain areas. *bioRxiv* 2020.11.02.365072 (2020) doi:10.1101/2020.11.02.365072.

19. Bolding, K. A. & Franks, K. M. Recurrent cortical circuits implement concentration-invariant odor coding. *Science (80-.).* **361**, (2018).

20. Ebitz, R. B. & Hayden, B. Y. The population doctrine in cognitive neuroscience. *Neuron* **109**, 3055–3068 (2021).

21. Semedo, J. D., Gokcen, E., Machens, C. K., Kohn, A. & Yu, B. M. Statistical

methods for dissecting interactions between brain areas. *Curr. Opin. Neurobiol.* **65**, 59–69 (2020).

22. Libby, A. & Buschman, T. J. Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* **24**, 715–726 (2021).

23. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792 (2014).

24. Aoi, M. C., Mante, V. & Pillow, J. W. Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nat. Neurosci.* **23**, 1410–1420 (2020).

25. Tang, C., Herikstad, R., Parthasarathy, A., Libedinsky, C. & Yen, S. C. Minimally dependent activity subspaces for working memory and motor preparation in the lateral prefrontal cortex. *Elife* **9**, 1–23 (2020).

26. Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M. & Kohn, A. Cortical Areas Interact through a Communication Subspace. *Neuron* **102**, 249-259.e4 (2019).

27. Lanore, F., Cayco-Gajic, N. A., Gurnani, H., Coyle, D. & Silver, R. A. Cerebellar granule cell axons support high-dimensional representations. *Nat. Neurosci.* **24**, 1142–1150 (2021).

28. Flesch, T., Juechems, K., Dumbalska, T., Saxe, A. & Summerfield, C. Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron* **110**, 1258-1270.e11 (2022).

29. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification*. (John Wiley & Sons, 2001).

30. Quian Quiroga, R. & Panzeri, S. *Principles of Neural Coding*. (CRC Press, 2013).

31. Kohn, A. *et al.* Principles of Corticocortical Communication: Proposed Schemes and Design Considerations. *Trends Neurosci.* **43**, 725–737 (2020).

32. Bolding, K. A. & Franks, K. M. Simultaneous extracellular recordings from mice olfactory bulb (OB) and piriform cortex (PCx) and respiration data in response to odor stimuli and optogenetic stimulation of OB. *CRCNS.org* http://dx.doi.org/10.6080/K00C4SZB (2018).

33. Gao, P. *et al.* A theory of multineuronal dimensionality, dynamics and measurement. *bioRxiv* (2017) doi:10.1101/214262.

34. Umakantha, A. *et al.* Bridging neuronal correlations and dimensionality reduction. *Neuron* **109**, 2740-2754.e12 (2021).

35. Recanatesi, S., Ocker, G. K., Buice, M. A. & Shea-Brown, E. Dimensionality in recurrent spiking networks: Global trends in activity and local origins in connectivity. *PLOS Comput. Biol.* **15**, e1006446 (2019).

36. Wang, H. T. *et al.* Finding the needle in a high-dimensional haystack: Canonical

correlation analysis for neuroscientists. *Neuroimage* **216**, 116745 (2020).

37. Lipshutz, D., Bahroun, Y., Golkar, S., Sengupta, A. M. & Chklovskii, D. B. A biologically plausible neural network for multichannel canonical correlation analysis. *Neural Comput.* **33**, 2309–2352 (2021).

38. Lai, P. L. & Fyfe, C. A neural implementation of canonical correlation analysis. *Neural Netw.* **12**, 1391–1397 (1999).

39. Pehlevan, C., Zhao, X., Sengupta, A. M. & Chklovskii, D. Neurons as Canonical Correlation Analyzers. *Front. Comput. Neurosci.* **14**, (2020).

40. Vía, J., Santamaría, I. & Pérez, J. A learning algorithm for adaptive canonical correlation analysis of several data sets. *Neural Netw.* **20**, 139–152 (2007).

41. Bolding, K. A. & Franks, K. M. Complementary codes for odor identity and intensity in olfactory cortex. *Elife* **6**, (2017).

Online Methods

Experiments. The olfactory data set was recorded and first reported in (Bolding & Franks, 2018). Their methods were approved by Duke University Institutional Animal Care and Use Committee. Their methods for olfactory stimulation, head-fixation, respiration monitoring, electrophysiology, and spike-sorting were also described in detail previously[41]. Here studied the following recordings: 170608, 170609, 170613, 170615, 170618, 170619, 170621, 170622.
The visual data set was first reported in (Siegle et al, 2021) recorded by the Allen Institute for Brain Science. The visual stimulation, head-fixation, electrophysiology, and spike-sorting are described and the data is available for public download here: https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html Animal use protocols were approved by the Allen Institute's Institutional Animal Care and Use Committee. Here we analyzed experiments with the following session IDs: 754312389, 715093703, 750749662, 754829445, 757970808, 759883607, 761418226, 763673393, 799864342.

CCA with Matlab. To perform CCA on experimental data, we used the Matlab function 'cannoncorr', e.g. [A,B,~,U,V,~]=cannoncorr(Region1Data,Region2Data). Where each set of responses (e.g. Region1Data and Region2Data) is a T x N matrix (T trials, N neurons). Then the ith neuron's response from Region1Data projected onto the jth CC is U(i,j). And the vector that define the direction of the jth CC in Region1 is the jth column of A. Note that for the analytical study of 2x2 populations (Fig. 5), CCA was not done with Matlab; the CC directions were computed analytically (see below and Supplementary Material.)

Projections with Matlab. For an n-dimensional subspace spanned by n vectors arranged as the n columns in a matrix M, the projection matrix P is P=(M/(M'*M))*M', where the ' indicates transpose, M/X indicates multiplication of M by inverse of X, and * indicates matrix multiplication. Then if W is a vector of responses P*W is vector after projection onto the subspace spanned by the columns of M.

Decoding accuracy. For a given pair of stimulus types, the decoding accuracy D of a single neuron's responses before (Fig 1) or after (Figs 3 and 4) projection on a subspace was determined by trying every possible threshold between the minimum and maximum response. The fraction of correct classifications for the best threshold was reported as D. For the analytical study of 2x2 population a different approach (independent of number of trials) was taken (see below and Supplementary Material.)

Defining groups of neurons with different noise correlations. In Fig 6, we defined three groups of neurons with low, medium, and high cross-population noise correlations. The medium group was chosen randomly; this group was used in Figs 1,3, and 4. The low and high groups were defined such that, on average, the single-neuron D values before any projection were not significantly different across all three groups (Pearson correlation between D and V1-LGN noise correlation with p>0.05).

Analytical 2x2 study. In Figs 2 and 5, we consider the case of two populations (X and Y) of simulated neurons whose responses to two stimuli, A and B, are correlated both within and across populations. We assume the responses ($\mathrm{r_X}$ and $\mathrm{r_Y}$) to each stimulus can be described by a multivariate Gaussian,

$$\mathrm{P}(r_X,\, r_Y|S) = \mathrm{N}\left(\begin{bmatrix}\mu_{\mathrm{X},S}\\ \mu_{\mathrm{Y},S}\end{bmatrix}, \Sigma_{\mathrm{S}}\right),$$

where $S = \{A, B\}$, $\mu_{X,S} \in R^m$, $\mu_{Y,S} \in R^n$, and Σ_S is a symmetric, positive-definite matrix of size $(m+n)\times(m+n)$. Here populations X and Y contain m and n neurons respectively ($\mathrm{m} = 2$ and $\mathrm{n} = 2$ for the simulated results in Figs 2 and 5); for example, Y may be a cortical region and X may represent OB or LGN. Without loss of generality, we simplify notation by shifting the mean responses so that $\mu_{X,A} = 0,\ \mu_{Y,A} = 0$; thus, we can drop the stimulus subscript on the mean vectors and use $\mu_X = \mu_{X,B},\ \mu_Y = \mu_{Y,B}$. We further assume that the stimulus-conditioned noise covariance matrix is the same for each stimulus: i.e. that $\Sigma_A = \Sigma_B =: \Sigma$. For simplicity, we assume that noise correlations $\mathrm{c_{xy}}$ were equal for any pair of cells across the two populations

$$\Sigma = \Lambda \begin{bmatrix} 1 & c_{OB} & c_{XY} & c_{XY} \\ c_{OB} & 1 & c_{XY} & c_{XY} \\ c_{XY} & c_{XY} & 1 & c_{PC} \\ c_{XY} & c_{XY} & c_{PC} & 1 \end{bmatrix} \Lambda; \qquad \Lambda = \begin{bmatrix} \sigma_{\mathrm{X}_1} & 0 & 0 & 0 \\ 0 & \sigma_{\mathrm{X}_2} & 0 & 0 \\ 0 & 0 & \sigma_{\mathrm{Y}_1} & 0 \\ 0 & 0 & 0 & \sigma_{\mathrm{Y}_2} \end{bmatrix}$$

As noted earlier, mean response to stimulus A was 0 and mean response to stimulus B was:

$$\mu = [\mu_{X_1} \quad \mu_{X_2} \quad \mu_{Y_1} \quad \mu_{Y_2}]^T$$

Each of the eleven parameters μ_{X_1}, μ_{Y_1}, μ_{X_2}, μ_{Y_2}, σ_{X_1}, σ_{X_2}, σ_{Y_1}, σ_{Y_2}, c_{OB}, c_{PC}, c_{XY} was chosen randomly from the following distributions:

$$\sigma_{X_1},\ \sigma_{X_2},\ \sigma_{Y_1},\ \sigma_{Y_2} \sim N_F(0,2)$$

$$\mu_{X_1},\ \mu_{Y_1} \sim N(0,1);\ \mu_{X_2},\ \mu_{Y_2} \sim N_F(0,1)$$

$$c_{OB}, c_{PC}, \tilde{c} \sim U(0,1);\ c_{XY} = max(\tilde{c} - 0.01, 0)$$

Here N_F is the folded normal distribution (if $X \sim N(\mu, \sigma)$, then $|X| \sim N_F(\mu, \sigma)$), and U is the uniform distribution. Parameters are defined so that – without loss of generality – $\mu_{X_2}, \mu_{Y_2} \geq 0$, all noise correlation parameters are non-negative, and c_{XY} has about a 1% chance of being 0. For a single simulation, each parameter was chosen independently from the above distributions. This was then repeated 1000 times, allowing a robust and wide-ranging survey of possible signal and noise correlation structures.

We next explored decoding under different assumptions. First, we sought to determine how well the stimulus can be decoded from responses within each population. In this simplified setting (responses are Gaussian, and the noise covariance is stimulus-independent), the optimal decoder is linear and can be determined by a simple analytical formula (see Supplemental methods). That is, we decode the stimulus by projecting the population response onto a single vector, and then compare that value with a threshold. Next, we use the principal direction from canonical correlation analysis, or CC1, as a linear decoder. Finally, we artificially remove cross-population noise correlations by setting $c_{XY} = 0$, and recompute the canonical correlation analysis with the revised stimulus-unconditioned covariance. We show that in this setting, the most correlated direction CC1 is in fact equal to the optimal projection vector (see Supplemental methods). These three decoding vectors – optimal, CC1, and CC1 with $c_{XY} = 0$ – are demonstrated in Fig. 5A as black, yellow and green respectively.

We computed the single-population optimal decoding directions v_X, v_Y using Eqn. (S1) and found the decoding accuracy by integrating the resulting one-dimensional Gaussians (see Eqn. (S2)). Similarly, we calculated the CC1 for each population using Eqn. (S3), and found the decoding accuracy using $v_{X,CC1}$, $v_{Y,CC1}$ as projection vectors. To compute single-cell decoding accuracy, we integrated under the marginal distributions given by projecting onto the coordinate directions $e_1 = \begin{bmatrix}1\\0\end{bmatrix}$, $e_2 = \begin{bmatrix}0\\1\end{bmatrix}$.