

Cerebellar granule cell axons support high dimensional representations

Frederic Lanore^{1,2*}, N. Alex Cayco-Gajic^{1,3*}, Harsha Gurnani¹, Diccon Coyle¹ and R. Angus Silver^{1#}

¹Department of Neuroscience, Physiology, and Pharmacology, University College London, London WC1E 6BT, UK

²Univ. Bordeaux, CNRS, Interdisciplinary Institute for Neuroscience, IINS, UMR 5297, F-33000 Bordeaux, France

³Group for Neural Theory, Laboratoire de neurosciences cognitives et computationnelles, Département d'études cognitives, École normale supérieure, INSERM U960, Université Paris Sciences et Lettres, 75005 Paris, France.

*These authors contributed equally to this work

#Correspondence to: a.silver@ucl.ac.uk

Abstract

In classical theories of cerebellar cortex, high dimensional sensorimotor representations are used to separate neuronal activity patterns, improving associative learning and motor performance. Recent experimental studies suggest that cerebellar granule cell (GrC) population activity is low dimensional. To examine sensorimotor representations from the point-of-view of downstream Purkinje cell ‘decoders’, we used 3D acousto-optic lens two photon microscopy to record from hundreds of GrC axons. Here we show that GrC axon population activity is high dimensional and distributed with little fine-scale spatial structure during spontaneous behaviors. Moreover, distinct behavioral states are represented along orthogonal dimensions in neuronal activity space. These results suggest that the cerebellar cortex supports high dimensional representations and segregates behavioral state dependent computations into orthogonal subspaces, as reported in the neocortex. Our findings match the predictions of cerebellar pattern separation theories and suggest that the cerebellum and neocortex utilize population codes with common features, despite their vastly different circuit structures.

Main text

A core function of the cerebellum is to predict the sensory consequences of motor actions^{1,2} by learning sensorimotor associations³. This is achieved by combining sensory and motor information from multiple sources. These include the neocortex, which is extensively interconnected with the cerebellar cortex, forming multi-synaptic loops via the basal pontine nucleus and thalamus⁴. Sensorimotor information enters the cerebellar cortex via mossy fibers⁵⁻⁸, which are sampled by a much larger population of granule cells (GrCs), located in the input layer. This ‘expansion recoding’ involves mixing of mossy fiber inputs with diverse functional properties⁹ and nonlinear thresholding in GrCs combined with anatomical expansion, which is thought to increase the dimensionality of GrC representations¹⁰⁻¹². Such nonlinear mixing and expansion is proposed to separate neuronal activity patterns by projecting them into a high-dimensional space¹⁰⁻¹⁵. High dimensional codes have recently been observed in forebrain structures, including the neocortex when viewing natural scenes¹⁶, performing complex cognitive tasks¹⁷ and during spontaneous behaviors¹⁸. By contrast, the dimensionality of neural activity in the cerebellar cortex has been found to be much lower, encoding movement parameters in a small number of variables^{19,20}. But it is unclear whether this arises from an inability of feedforward cerebellar circuits to support high dimensional population codes, or

the nature of the behavioral tasks, which could limit the dimensionality of their neural representations²¹. Determining whether the cerebellar cortex can support high dimensional sensorimotor representations is therefore a key test of theoretical predictions that it performs expansion recoding^{10,11} and pattern separation¹²⁻¹⁴ and whether the neocortex and cerebellar cortex utilize distinct population-level sensorimotor representations.

Results

Axonal population activity

To investigate sensorimotor representations in the cerebellar cortex we selectively expressed GCaMP6f in cerebellar GrCs in mouse Crus I (Extended Data Fig. 1), an area that encodes information from the whiskers^{7,22,23}. Rather than imaging GrC somata^{19,20,24}, where synaptic and action potential linked Ca²⁺ influx could be mixed due to their close proximity²⁵, we monitored GrC axons in the molecular layer (parallel fibers), since their varicosities exhibit large action potential induced Ca²⁺ transients²⁶. We utilized the unique orthogonal arrangement of parallel fibers and Purkinje cell dendritic trees to read out GrC activity from the point of view of the ‘downstream decoder’ (i.e. Purkinje cells; Fig. 1a). To do this we used acousto-optic lens (AOL) 3D two-photon microscopy²⁷ (Methods) to simultaneously image multiple XY ‘patches’ (X: 48 - 110 μm , Y: 13 - 20 μm) positioned with a staircase arrangement through the molecular layer (Fig. 1a). Moreover, real-time and *post hoc* correction for brain movement enabled reliable recordings from parallel fiber varicosities during behavior (Methods; Video S1). Head-fixed mice were free to stand or run on a wheel and to whisk. Such spontaneous behaviors encompass many more individual movements than simple constrained behaviors and are therefore likely to have a higher intrinsic dimensionality²¹. Parallel fiber varicosities within each of the imaged patches were identified and grouped into putative axons on the basis of their spatial alignment along the averaged parallel fiber direction and the level of correlation in their activity (Fig. 1b, c and Extended Data Fig. 2; Methods). To validate our grouping procedure, we measured the distance between varicosities on the same putative GrC axon, as this was not one of the structural criteria used for grouping. The observed intervaricosity distance varied between 2 - 17 μm , with a mean of $5.50 \pm 0.08 \mu\text{m}$ (1080 putative axons with multiple varicosities), a range and mean that was similar to high resolution measurements from sparsely labeled parallel fibers in fixed tissue²⁸ (Fig. 1d). Following this analysis we identified 135 to 700 GrC axons per recording (Fig. 1e). Parallel fiber population activity had a rich and diverse

structure that was correlated to the whisker set point (low frequency changes in whisker angle) and to the locomotion speed of the animal (Fig. 1e and Extended Data Fig. 3).

Spontaneous behavior typically consisted of periods of quiet wakefulness (QW) when the mice rested on the wheel and exhibited little movement or whisking, and periods of pronounced whisking and locomotion, which we called the active state (AS), and which likely encompassed additional unobserved behaviors (Fig. 2a). Indeed, whisking and locomotion speed were highly correlated with one another ($p = 2.4 \times 10^{-4}$, Wilcoxon signed rank test, $n = 13$ experiments, $N = 5$ animals; Supp. Table 1; Extended Data Fig. 3). Parallel fiber activity showed a continuum of responses (Fig. 1e; Extended Data Fig. 4a, b), including both positively and negatively modulated responses during the AS (Fig. 1e and 2b). Comparison of the $\Delta F/F$ in axons during AS and QW revealed a majority of AS-preferring parallel fibers (positively modulated, 66%, $n = 13$, $N = 5$), with a smaller population of QW-preferring parallel fibers (negatively modulated, 19%; Fig. 2b). Correlation of parallel fiber activity with whisking and locomotor sensorimotor variables revealed a similar fraction of positively and negatively modulated axons associated with each of these behavioral parameters (Extended Data Fig. 3). Since it is possible that negatively modulated parallel fiber signals could arise from axial brain movement, we compared the intensity of beads embedded within the tissue with the activity of negatively modulated axons. No correlation between negatively modulated parallel fiber activity and bead fluorescence was observed, ruling out this possibility (Supp. Fig. 1). Moreover, both positively and negatively modulated responses were also observed when imaging larger GrC somata during whisking and locomotion (Supp. Fig. 2). This finding also argues against the possibility that negatively modulated axon responses arose from undetected off-target expression in molecular layer interneurons or Purkinje cells. A smaller proportion of parallel fibers were not significantly modulated by behavioral state (15%). For these parallel fibers, Ca^{2+} events were evident and the distribution of signal-to-noise ratios was similar to those of positively or negatively modulated parallel fibers, indicating that their lack of modulation was not simply due to noise (Extended Data Fig. 5). Overall, the proportions of negatively, positively and non-modulated parallel fibers were consistent across experimental sessions and animals (Extended Data Fig. 6). Due to the relatively low sensitivity of GCaMP6f for single spikes, these $\Delta F/F$ responses are likely to correspond to bursts or sustained spiking in parallel fibers. Nevertheless, these results show that spontaneous behaviors are represented in a bidirectional parallel fiber

population code in Crus I. This reveals a greater diversity in GrC responses than previously reported in awake behaving mice^{20,24,25}.

Previous findings in anesthetized mice showed that parallel fibers are activated in sparse clusters during discrete sensory stimulation of the perioral region²⁹. To investigate whether clusters of parallel fiber activity are present during spontaneous behavior, we computed the average pairwise cross correlation for each pair of axons and estimated the pairwise distance between axons in the recorded 3D volume (Extended Data Fig. 7a). No significant spatial dependence in the correlation coefficients was observed in the XY plane, except for a weak increase between parallel fibers within 2 μm ($p < 10^{-4}$, Wilcoxon rank sum test, $n = 13$, $N = 5$; Fig. 2c), likely due to our conservative grouping procedure. A similar result was obtained for ungrouped varicosities (Extended Data Fig. 7b), and when we included the Z dimension across imaging planes, albeit at lower spatial resolution (Extended Data Fig. 7c). Moreover, when positively and negatively modulated parallel fiber responses were examined separately, they showed no preferential clustering, as the distribution of within-group nearest neighbor (NN) distances remained similar after shuffling the group labels (positively modulated: 3403 putative axons, $p = 0.32$; negatively modulated: 896 putative axons, $p = 0.21$, Kolmogorov-Smirnoff test, $n = 13$, $N = 5$; Fig. 2d). Next, we investigated whether spatial clustering occurred during more defined behaviors. However, when our analysis was restricted to locomotion onsets, no significant spatial dependence in the correlation structure was observed (Extended Data Fig. 4c, d). These results show that parallel fiber activity in Crus I lacks spatial clustering during spontaneous behaviors.

Geometry of neural representations

We next explored how behavior is encoded across the parallel fiber population in Crus I by examining neural activity space, in which each dimension represents a different neuron, and each point in space corresponds to a unique pattern of activity across the population of axons. Because of the discrete behavioral state transitions in our data (Fig. 2a), we expected to observe two clusters of points corresponding to AS and QW. In principle, these clusters could overlap significantly, or alternatively, they could be encoded in distinct, well-separated representations (or ‘manifolds’; Fig. 3a). To visualize the structure of the representations of AS and QW, we reduced the dimensionality of the neural activity space by plotting the first three principal

components of the parallel fiber population activity (Fig. 3b). This revealed that parallel fiber activity represented AS and QW in well-separated manifolds, which were connected by distinct trajectories representing transitions in either direction (AS \rightarrow QW or QW \rightarrow AS; Video S2). Clearly separated manifolds for AS and QW were present in all five animals with > 100 parallel fibers recorded (Fig. 3b and Extended Data Fig. 8). The quantification of the average intra-manifold Euclidean distances to the inter-manifold distances revealed that the average distance between the AS and QW manifolds was 30 - 40% larger than either manifold ($p = 2.4 \times 10^{-4}$ (AS), $p = 2.4 \times 10^{-4}$ (QW), Wilcoxon signed rank test, $n = 13$, $N = 5$; Fig. 3c), indicating that these behavioral manifolds were well separated. In two animals, we observed isolated whisker movements in the absence of locomotion. Since these were excluded from the AS and QW state criteria, we wondered whether the neural representations of these isolated whisks would be embedded in the AS representation, or occupy a separate region of neural activity space. Analysis of these isolated whisking periods revealed that they indeed occupied a region of activity space that was distinct from the AS and QW manifolds (Supp. Fig. 3).

The geometry of neural representations can provide insight about the computations performed by neural populations³⁰. For example, in the motor and premotor cortices, orthogonal manifolds are thought to limit interference between different behaviors^{31,32}. Visualization and rotation of the AS and QW manifolds revealed an apparently orthogonal arrangement in activity space (Video S3). To quantify how the manifolds were orientated, we calculated the angle between the AS and QW subspaces within the neural activity space (Fig. 3d; Methods). Noisy estimates of the principal axes of the behavioral subspaces could make the subspaces appear artificially orthogonal, since random vectors are likely to be orthogonal in a neural activity space with high extrinsic dimensionality (i.e., large number of neurons). To control for measurement noise, we calculated the angle between random halves of the population activity after shuffling across time. Repeating this procedure gave a null distribution of angles for each experiment, which could then be compared with the angle observed in the data (Fig. 3e). The mean angle between the subspaces for the AS and QW was 1.4 ± 0.02 radians, suggesting they were nearly orthogonal, with significantly smaller values in the control (0.5 ± 0.09 radians, $p = 4.9 \times 10^{-4}$, Wilcoxon signed rank test, 12/13 experiments reached significance, $n = 13$, $N = 5$; Fig. 3f). These findings establish that population activity in the cerebellar cortex is organized into orthogonal subspaces representing different behavioral states.

We next asked whether subspace orthogonality simply arose from distinct populations of parallel fibers being active during the different behavioral states. To test this, we removed increasing fractions of the strongest positively and negatively modulated axons and recalculated the angle between AS and QW subspaces. As more positively and negatively modulated parallel fibers were excluded, the angle between these subspaces gradually decreased but remained significantly larger than in the shuffle control ($p < 4.5 \times 10^{-3}$ for 0th through 70th percentile of positively and negatively modulated parallel fibers excluded, Wilcoxon signed rank test with Bonferroni correction; Fig. 3g). The decrease in angle between AS and QW subspaces was not significantly different from a control in which we excluded the same number of neurons, randomly sampled from the entire distribution (0th through 100th percentile, Wilcoxon signed rank test with Bonferroni correction; Fig. 3g), suggesting that this decrement could be due to a fall in the number of neurons. The robustness to removing strongly negatively and positively modulated parallel fiber responses shows that they were not the sole determinant of the orthogonality of the AS and QW manifolds. This suggests that subspace orthogonality is not simply inherited from the bidirectionality of the parallel fiber responses.

Distributed sensorimotor representations

Since transitions between QW and AS were associated with protraction and retraction of the whiskers, we next asked whether widespread activity mediated by the positively and negatively modulated parallel fibers could be explained by changes in whisker set point. To investigate this, we examined how the first principle component (PC1), which captures widespread changes in parallel fiber activity, was related to whisker set point. While PC1 captured the transitions between AS and QW, it did not reflect different resting positions of the whisker set point during QW, even when it varied over the majority of its range (Fig. 4a, b). Across animals, PC1 was significantly correlated with whisker set point over all time (0.69 ± 0.05 , $n = 13$, $N = 5$; Fig. 4c), but there was little correlation during QW (0.04 ± 0.06 , $n = 13$, $N = 5$). Instead, PC1 was highly correlated with a binary variable reflecting the behavioral state (0.89 ± 0.02 , $n = 13$, $N = 5$). Moreover, PC1 was significantly correlated with whisker set point during the AS (0.49 ± 0.06 , $n = 13$, $N = 5$), indicating that it contains information about whisker position during active whisking. These results suggest that widespread modulation of parallel fiber activity in Crus I (i.e., PC1) is correlated with active behaviors rather than encoding detailed information on whisker set point.

We next asked whether more detailed information on the whisker set point was present in the GrC population activity as a whole. To investigate this, we used cross-validated linear regression to predict whisker set point from increasing numbers of principal components (PCs), and calculated the unexplained variance in held-out data that was not used for training (Fig. 5a, b; Methods). Across animals, decoding from the optimal number of PCs led to substantially better decoding performance than the first PC ($p = 2.4 \times 10^{-4}$, Wilcoxon signed rank test, $n = 13$, $N = 5$; Fig. 5c) or the first 10 PCs ($p = 2.4 \times 10^{-4}$, Wilcoxon signed rank test, $n = 13$, $N = 5$). This improvement was not due to an increased number of parameters since decoding performance was cross-validated. This suggests that more detailed information on whisker set point is available in the higher PCs of parallel fiber activity. Given the low correlation between whisker set point and PC1 during QW (Fig. 4c), we next investigated whether any information on whisker set point resting positions was present across GrCs during QW. To this end, we trained a decoder on activity exclusively during QW. The QW-only decoder was significantly better at predicting whisker set point during QW than a decoder trained on randomly sampled times during the experiment ($p = 2.4 \times 10^{-4}$, Wilcoxon signed rank test, $n = 13$, $N = 5$; Fig. 5d). These results suggest that detailed information about whisker set point is available in the population activity and more than one linear decoder (e.g. Purkinje cell) may be required to decode across different states.

The finding that many principal components are required to decode detailed whisker set point information raises the question of how such information is distributed across parallel fibers. Classic cerebellar theories have argued that sensorimotor information should be distributed across GrC populations rather than encoded in single GrCs¹⁰⁻¹⁴. To test this, we used lasso regression (L1 regularization; Methods) to quantify the minimal number of parallel fibers necessary for optimal decoding. This gave a minimum unexplained variance with 225 ± 22 parallel fibers, which was substantially lower than for the best performing parallel fiber (Fig. 5e, f; $p = 2.4 \times 10^{-4}$, Wilcoxon signed rank test, $n = 13$, $N = 5$). To investigate whether such distributed representations are present for other behavioral variables, we also investigated locomotion speed. Hundreds of parallel fibers (184 ± 22 parallel fibers, $n = 11$, $N = 5$) were required to minimize the cross-validated unexplained variance for locomotion speed (Extended Data Fig. 9a-c). Although there was a weak correlation between the decoders of whisker set point and locomotion (correlation between decoder coefficients: $r = 0.17 \pm 0.04$, $n = 11$, $N =$

5), there was an inverse relationship between decoding error and the similarity of the regression coefficients (Extended Data Fig. 9d), indicating that more complete representations of these variables tended to be partially aligned. These findings suggest that sensorimotor representations are distributed across the parallel fiber population.

Dimensionality of population activity

Theoretical work on cerebellar pattern separation predicts that sensorimotor representations in GrC populations are high-dimensional¹⁰⁻¹⁴. To test this, we quantified the dimensionality of parallel fiber population activity during spontaneous behaviors using a cross-validated variant of PCA (Methods). This revealed that the state dependent changes reflected in PC1 captured only $10.3 \pm 1.2\%$ of the variance (Fig. 6a inset, data subsampled to 300 axons, $n = 10$, $N = 3$). We then estimated the number of PCs required to attain the maximum variance explained, beyond which it decreased due to noise or other non-shared variability (Fig. 6a). This provided a lower bound on the dimensionality that could be inferred given the noise level within each experiment (21.6 ± 2.5 dimensions in 300 parallel fibers explaining $34.2 \pm 3.7\%$ of the variance, $n = 10$, $N = 3$; Methods). To obtain a more accurate estimate of the dimensionality, we noted that experiments with higher values of maximum variance explained tended to have a higher dimensionality (Fig. 6a). A simple model confirmed that a linear relationship is expected across a wide range of signal-to-noise levels (Extended Data Fig. 10). Linear extrapolation of the data suggested that 62 dimensions are required to explain the full variance of a population of 300 parallel fibers during spontaneous behaviors (Fig. 6b). This corresponds to a highly non-redundant population code with an average of only 5 parallel fibers for each encoded dimension. This ratio remained low for populations of up to 650 parallel fibers (4 - 5 neurons per dimension; Fig. 6c), indicating that population activity in GrC axons is high dimensional during spontaneous behaviors.

Discussion

Our recordings from hundreds of GrC axons in the molecular layer establish that the cerebellar cortex can support distributed, high dimensional representations during spontaneous behaviors. The presence of high dimensional population activity is consistent with the cerebellar input layer performing pattern separation¹², as proposed by Marr-Albus theory^{10,11}, and potentially explains why GrCs are so numerous³³. This contrasts with previous findings of low

dimensional GrC population activity during a mouse forelimb lever task²⁰ and tail movements in zebrafish larvae¹⁹. However, it was unclear whether these results were due to an inability of the cerebellar cortex to support high dimensional representations, or the low dimensionality of these defined behavioral tasks²¹. Our finding that only 5 GrC axons are required, on average, to encode each dimension is comparable to the low number of neurons per dimension found in the visual cortex, which has been shown to be as high dimensional as possible while also maintaining a smooth population code, which aids generalization to novel stimuli¹⁸. Thus, the dimensionality of parallel fiber activity that we observed could be near the optimum set by the trade-offs between pattern separability, robustness to noise, and generalizability.

Our results also show that GrC axonal populations employ a bidirectional coding strategy and that differentially modulated parallel fibers are spatially dispersed within the molecular layer. The fact that a subpopulation of GrCs are active in the absence of movement is consistent with previously reported cell-attached recordings from individual GrCs in Crus I which showed that although most GrCs fire during periods of active whisking, some exhibit substantial firing rates at rest²². Bidirectional coding is likely to be widespread across other lobules in the cerebellum where individual GrCs and mossy fibers exhibit tuning for a range of sustained variables including joint angle⁵ and angular head velocity⁶. It is possible that the positively and negatively modulated GrC responses we report here could contribute to the increased and decreased responses observed in downstream inhibitory interneurons in the molecular layer²² and Purkinje cells³⁴⁻³⁶.

Our choice to study spontaneous behaviors was motivated by recent work demonstrating that the dimensionality of neural representations is limited by the richness of the behavior²¹. However, this variability brings with it certain challenges that warrant consideration³⁷. Our definition of what constitutes an active state likely combines many behavioral parameters, and as a result the manifold structures that we have identified may aggregate multiple representations. Thus, while the representations we observe during spontaneous behaviors demonstrates the cerebellar cortex can support high dimensional activity, simpler behaviors and the individual behavioral parameters that contribute to spontaneous behaviors are both likely to be represented by lower dimensional manifold structures. These could be embedded within the coding subspace¹², consistent with the low dimensional representations reported for

well-defined behaviors^{19,20}. Future work will be required to explore the properties of the full manifold structure of these neural representations³⁸.

The high dimensionality and distributed nature of the parallel fiber population activity that we observe support the idea that the cerebellar input layer generates mixed sensorimotor representations¹³. This population coding strategy provides the capacity to encode vast numbers of different sensorimotor combinations that arise during complex behaviors^{15,17}. Moreover, the spatially uniform activity structure, when viewed in the plane of the Purkinje cell dendritic tree, suggests that parallel fiber synaptic inputs could be spatially distributed across the Purkinje cell dendritic tree during spontaneous behaviors. Such a configuration favours linear synaptic integration³⁹, potentially enabling Purkinje cells to act as linear decoders^{34,40} as originally proposed in classical theories of cerebellar function^{10,11}. However, parallel fiber activity only reflects potential synaptic inputs onto Purkinje cells (or molecular layer interneurons). Synaptic plasticity rules³ are likely to further select subsets of GrCs that form functional synapses, since the majority of synapses on an individual Purkinje cell are silent⁴¹. Thus, the pattern of synaptic input onto an individual Purkinje cell could still exhibit structure since it is likely to be a spatially⁴² and temporally⁴³ selected subset of the parallel fiber population activity. While further work is required to elucidate how individual Purkinje cells decode the parallel fiber activity that passes through their dendritic trees, our findings suggest that the functional and anatomical properties of parallel fibers are well suited for generating the wide array of sensorimotor associations required for predicting the sensory consequences of self-generated movements^{1,2}, and could be employed in coordinating other dynamical processes including those underlying cognitive processes⁴⁴.

Dimensionality reduction of the parallel fiber population activity in Crus I revealed that it forms distinct, well-separated manifolds representing AS and QW, that are orthogonally arranged. Orthogonal manifolds have been reported in the neocortex^{18,31,32,45}, but such properties have not previously been reported in the simpler, largely feedforward architecture of the cerebellar cortex. In premotor cortex, orthogonal ‘output-potent’ and ‘output-null’ subspaces have been proposed to separate neural activity that has a direct behavioral output from activity that reflects internal computations such as motor preparation^{31,45}. The finding that the cerebellum contributes to preparatory activity in the motor cortex^{46,47} raises the possibility that the orthogonal manifolds in the cerebellum perform a similar function.

Our finding that the cerebellar GrC population code shares several properties in common with the neocortex, including positively and negatively modulated responses⁴⁸, representation of behavioral state^{18,31}, orthogonal manifolds^{18,31,32,45} and, mixed¹⁷, high dimensional, distributed representations of spontaneous behaviors¹⁸, extend recent observations of a high level of coordination in the activity of individual cells in the cerebellum and neocortex²⁰. Indeed, both the cortico-pontine pathway, which conveys efferent copy information to the cerebellum⁴⁹, and the return loop of the cortico-thalamo-cerebellar pathway⁴ could be involved in generating and sharing common population level representations.

Our results establish that the cerebellar GrC population code can utilize a high dimensional neural activity space, as predicted for a general purpose pattern separation device¹². Moreover, we show that GrC population representations share several features in common with those in the neocortex, raising the possibility that sensorimotor information is shared through an effective communication subspace⁵⁰ in the cortico-cerebellar system.

References

1. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. *Trends Cogn. Sci.* **2**, 338–347 (1998).
2. Brooks, J. X., Carriot, J. & Cullen, K. E. Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nat. Neurosci.* **18**, 1310–1317 (2015).
3. Raymond, J. L. & Medina, J. F. Computational Principles of Supervised Learning in the Cerebellum. *Annu. Rev. Neurosci.* **41**, 233–253 (2018).
4. Kelly, R. M. & Strick, P. L. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J. Neurosci.* **23**, 8432–8444 (2003).
5. van Kan, P. L., Gibson, A. R. & Houk, J. C. Movement-related inputs to intermediate cerebellum of the monkey. *J. Neurophysiol.* **69**, 74–94 (1993).
6. Arenz, A., Silver, R. A., Schaefer, A. T. & Margrie, T. W. The contribution of single synapses to sensory representation in vivo. *Science* **321**, 977–980 (2008).
7. Proville, R. D. *et al.* Cerebellum involvement in cortical sensorimotor circuits for the control of voluntary movements. *Nat. Neurosci.* **17**, 1233–1239 (2014).
8. Rancz, E. A. *et al.* High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons. *Nature* **450**, 1245–1248 (2007).

9. Chabrol, F. P., Arenz, A., Wiechert, M. T., Margrie, T. W. & DiGregorio, D. A. Synaptic diversity enables temporal coding of coincident multisensory inputs in single neurons. *Nat. Neurosci.* **18**, 718–727 (2015).
10. Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1969).
11. Albus, J. S. A theory of cerebellar function. *Mathematical Biosciences* **10**, 25–61 (1971).
12. Cayco-Gajic, N. A. & Silver, R. A. Re-evaluating Circuit Mechanisms Underlying Pattern Separation. *Neuron* **101**, 584–602 (2019).
13. Cayco-Gajic, N. A., Clopath, C. & Silver, R. A. Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks. *Nat. Commun.* **8**, 1116 (2017).
14. Litwin-Kumar, A., Harris, K. D., Axel, R., Sompolinsky, H. & Abbott, L. F. Optimal Degrees of Synaptic Connectivity. *Neuron* **93**, 1153–1164.e7 (2017).
15. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: high dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016).
16. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361–365 (2019).
17. Rigotti, M. *et al.* The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
18. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, 255 (2019).
19. Knogler, L. D., Markov, D. A., Dragomir, E. I., Štih, V. & Portugues, R. Sensorimotor Representations in Cerebellar Granule Cells in Larval Zebrafish Are Dense, Spatially Organized, and Non-temporally Patterned. *Curr. Biol.* **27**, 1288–1302 (2017).
20. Wagner, M. J. *et al.* Shared Cortex-Cerebellum Dynamics in the Execution and Learning of a Motor Task. *Cell* **177**, 669–682.e24 (2019).
21. Gao, P. & Ganguli, S. On simplicity and complexity in the brave new world of large-scale neuroscience. *Curr. Opin. Neurobiol.* **32**, 148–155 (2015).
22. Chen, S., Augustine, G. J. & Chadderton, P. Serial processing of kinematic signals by cerebellar circuitry during voluntary whisking. *Nat. Commun.* **8**, 232 (2017).
23. Shambes, G. M., Gibson, J. M. & Welker, W. Fractured somatotopy in granule cell tactile areas of rat cerebellar hemispheres revealed by micromapping. *Brain Behav. Evol.* **15**, 94–140 (1978).
24. Giovannucci, A. *et al.* Cerebellar granule cells acquire a widespread predictive feedback signal during motor learning. *Nat. Neurosci.* **20**, 727–734 (2017).
25. Ozden, I., Dombeck, D. A., Hoogland, T. M., Tank, D. W. & Wang, S. S.-H. Widespread state-dependent shifts in cerebellar activity in locomoting mice. *PLoS One* **7**, e42650 (2012).
26. Rebola, N. *et al.* Distinct Nanoscale Calcium Channel and Synaptic Vesicle Topographies Contribute to the Diversity of Synaptic Function. *Neuron* **104**, 693–710.e9 (2019).
27. Nadella, K. M. N. S. *et al.* Random-access scanning microscopy for 3D imaging in awake

- behaving animals. *Nat. Methods* **13**, 1001–1004 (2016).
28. Pichitpornchai, C., Rawson, J. A. & Rees, S. Morphology of parallel fibres in the cerebellar cortex of the rat: an experimental light and electron microscopic study with biocytin. *J. Comp. Neurol.* **342**, 206–220 (1994).
 29. Wilms, C. D. & Häusser, M. Reading out a spatiotemporal population code by imaging neighbouring parallel fibre axons in vivo. *Nat. Commun.* **6**, 6464 (2015).
 30. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural Manifolds for the Control of Movement. *Neuron* **94**, 978–984 (2017).
 31. Li, N., Daie, K., Svoboda, K. & Druckmann, S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532**, 459–464 (2016).
 32. Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M. & Cunningham, J. P. Reorganization between preparatory and movement population responses in motor cortex. *Nat. Commun.* **7**, 13239 (2016).
 33. Lange, W. Cell number and cell density in the cerebellar cortex of man and some other mammals. *Cell Tissue Res.* **157**, 115–124 (1975).
 34. Chen, S., Augustine, G. J. & Chadderton, P. The cerebellum linearly encodes whisker position during voluntary movement. *Elife* **5**, e10509 (2016).
 35. Zhou, H. *et al.* Cerebellar modules operate at different frequencies. *Elife* **3**, e02536 (2014).
 36. De Zeeuw, C. I. Bidirectional learning in upbound and downbound microzones of the cerebellum. *Nat. Rev. Neurosci.* **22**, 92–110 (2021).
 37. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**, 480–490 (2017).
 38. Musall, S., Urai, A. E., Sussillo, D. & Churchland, A. K. Harnessing behavioral diversity to understand neural computations for cognition. *Curr. Opin. Neurobiol.* **58**, 229–238 (2019).
 39. Silver, R. A. Neuronal arithmetic. *Nat. Rev. Neurosci.* **11**, 474–489 (2010).
 40. Walter, J. T. & Khodakhah, K. The linear computational algorithm of cerebellar Purkinje cells. *J. Neurosci.* **26**, 12861–12872 (2006).
 41. Brunel, N., Hakim, V., Isope, P., Nadal, J.-P. & Barbour, B. Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell. *Neuron* **43**, 745–757 (2004).
 42. Valera, A. M. *et al.* Stereotyped spatial patterns of functional synaptic connectivity in the cerebellar cortex. *Elife* **5**, (2016).
 43. Suvrathan, A., Payne, H. L. & Raymond, J. L. Timing Rules for Synaptic Plasticity Matched to Behavioral Function. *Neuron* **92**, 959–967 (2016).
 44. Ito, M. Control of mental activities by internal models in the cerebellum. *Nat. Rev. Neurosci.* **9**, 304–313 (2008).
 45. Vyas, S. *et al.* Neural Population Dynamics Underlying Motor Learning Transfer. *Neuron* **97**, 1177–1186.e3 (2018).

46. Gao, Z. *et al.* A cortico-cerebellar loop for motor planning. *Nature* **563**, 113–116 (2018).
47. Chabrol, F. P., Blot, A. & Mrsic-Flogel, T. D. Cerebellar Contribution to Preparatory Activity in Motor Neocortex. *Neuron* **103**, 506–519.e4 (2019).
48. Peters, A. J., Lee, J., Hedrick, N. G., O’Neil, K. & Komiyama, T. Reorganization of corticospinal output during motor learning. *Nat. Neurosci.* **20**, 1133–1141 (2017).
49. Person, A. L. Corollary Discharge Signals in the Cerebellum. *Biol Psychiatry Cogn Neurosci Neuroimaging* **4**, 813–819 (2019).
50. 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).

Acknowledgments: This project was supported by the Wellcome Trust (095667, 203048) and the Agence Nationale de la Recherche (ANR-17-EURE-0017). R.A.S is in receipt of a Wellcome Trust Principal Research Fellowship. F.L. was supported by a postdoctoral Fondation Fyssen fellowship, a Marie Curie fellowship (Project No 331710, FP7 program) and the Wellcome Trust (203048). F.L. is now funded by the Centre National de la Recherche Scientifique. H.G. was funded by the Wellcome Trust PhD programme (203734). We acknowledge the GENIE Program and the Janelia Research Campus, Howard Hughes Medical Institute for making the GCaMP6 material available. We thank Adam Hantmann (Janelia Research Campus) for providing the *Slc17a7-Cre* transgenic mice and Ashok Litwin-Kumar, Tom Otis, Kenneth D Harris, David Attwell, Lena Justus, Thomas J. Younts, Antoine Valera, Jason S. Rothman and Tomas Fernandez Alfonso for their comments on the manuscript.

Author contributions: Conceptualization: F.L., N.A.C.G., and R.A.S. Methodology: F.L., N.A.C.G., H.G. and R.A.S. Software: N.A.C.G., H.G. and F.L. Formal analysis: N.A.C.G. Investigation: F.L. and D.C. performed experiments. Writing - original draft preparation: F.L., N.A.C.G., and R.A.S. Writing - review and editing: F.L., N.A.C.G., H.G., D.C., and R.A.S. Supervision: R.A.S. Funding acquisition: R.A.S.

Competing interests: R.A.S. is a named inventor on patents owned by UCL Business relating to linear and nonlinear Acousto-optic lens 3D laser scanning technology. The remaining authors declare no competing interests.

Figure legends

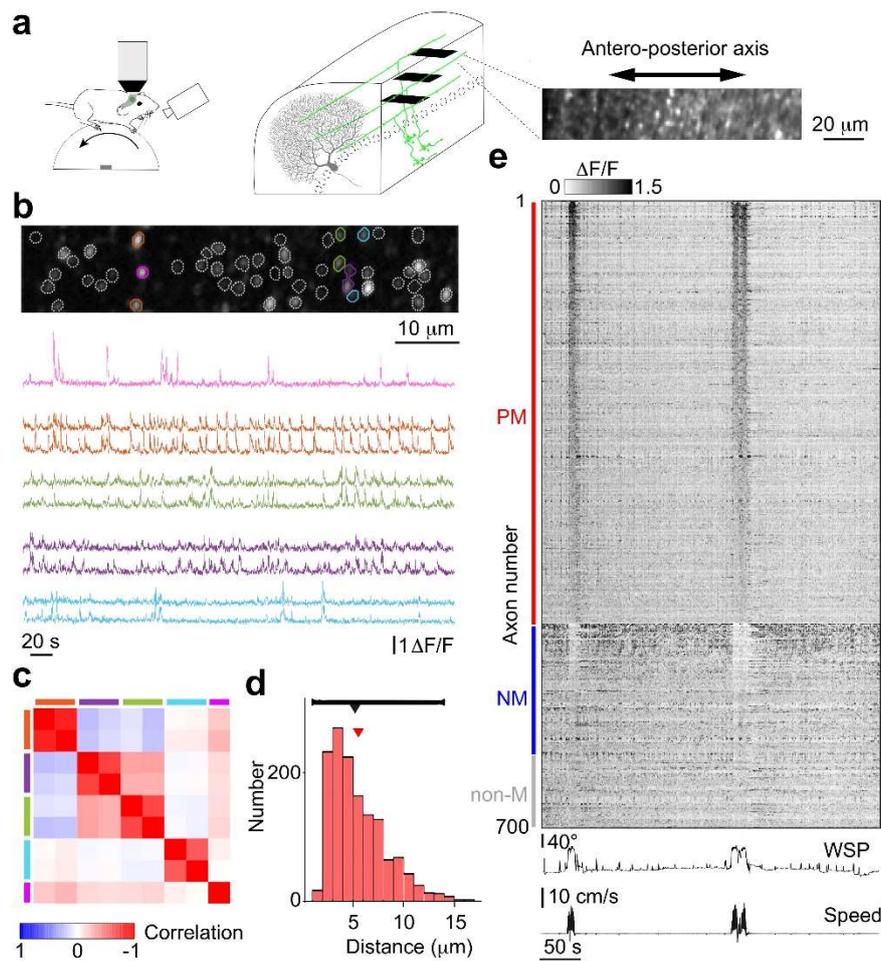


Fig. 1. Granule cell axon population activity during spontaneous behaviors. (a) Schematic of the experimental configuration for the acousto-optic lens (AOL) 3D imaging showing head-fixed mouse on a wheel, along with high-speed camera to track whisker movement (left). Spatial arrangement of multiple simultaneously acquired imaging planes ('patches') within the imaging volume in relation to granule cell (GrC) axons (in green) and Purkinje cell dendritic tree (in grey) in the molecular layer with example of imaged patch showing varicosities expressing GCaMP6f (average fluorescence image; right). (b) Example of varicosity grouping ($n=1$, $N=1$ of $n=13$, $N=5$). Top: Correlation image of a patch (13.7 μm x 68.4 μm) with identified varicosities outlined in white dots. Greyscale indicates correlation with the fluorescence of neighbouring pixels. The colored outlines show examples of grouped varicosities per axon, with each color corresponding to one axon. Bottom: $\Delta F/F$ traces for each varicosity highlighted in color. (c) Matrix showing correlation between $\Delta F/F$ traces of varicosities in b. Colored bars on the side show the grouping into putative axons. The strongest

correlations were between varicosities on the same putative axon. **(d)** Distribution of distances between varicosities grouped onto the same putative parallel fiber ($n = 13$, $N = 5$). The red arrow shows the mean intervaricosity distance. Black line and arrow indicate the range and mean intervaricosity distances as determined previously in fixed tissue with anatomical methods²⁸. The close match suggests our detection of varicosities and method of grouping into axons identifies the majority of boutons per active axon in the imaged patch. **(e)** Example of activity ($\Delta F/F$) of 700 putative GrC axons (parallel fibers) in a single experiment, grouped into positively modulated (PM, red), negatively modulated (NM, blue), and non-modulated parallel fibers (non-M, grey). Bottom: Whisker set point (WSP; slow-frequency component of whisker angle) and locomotion speed.

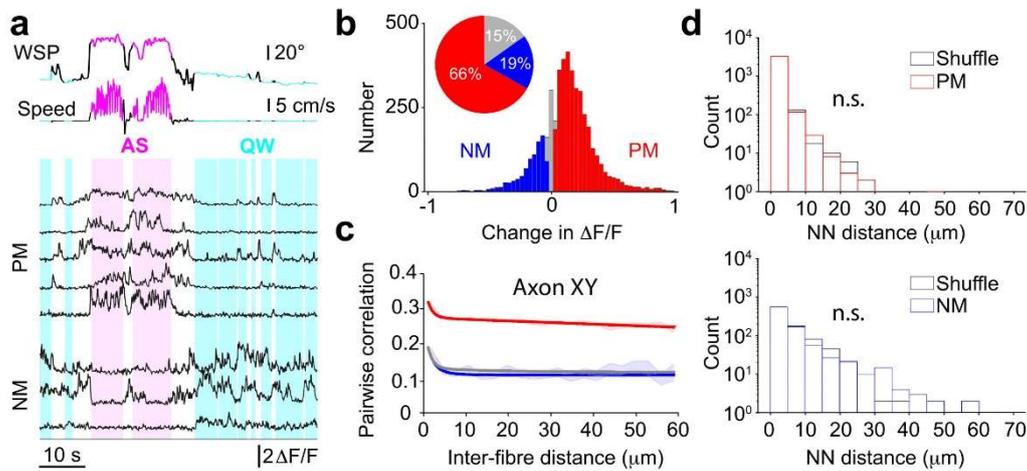


Fig. 2. Bidirectional spatially mixed parallel fiber responses during active behavioral state. **(a)** Example of behavioral state segmentation and parallel fiber responses. Top: time series of whisker set point (WSP) and locomotion speed labelled as periods of active state (AS, magenta), quiet wakefulness (QW state, cyan) or unclassified timepoints (black). Bottom: $\Delta F/F$ traces of parallel fibers that exhibited a significant increase or decrease during the AS, compared to QW ($p < 0.05$, two-sided shuffle test). **(b)** Histogram of changes in $\Delta F/F$ response during the AS relative to QW across all parallel fibers ($n = 13$, $N = 5$). Positively modulated (PM; red) and negatively modulated (NM; blue) parallel fibers, as well as axons which were not significantly modulated by behavioral state (grey). **(c)** Average pairwise correlation between parallel fiber activity as a function of the distance between axons ($n = 13$, $N = 5$), shown for positively modulated (red), negatively modulated (blue), and all parallel fibers (grey). Shading indicates s.e.m. and solid lines indicate double-exponential fits. **(d)** Within-

group nearest-neighbor (NN) distances for positively modulated (red) and negatively modulated (blue) parallel fibers, and shuffle controls (black) ($n = 13, N = 5$).

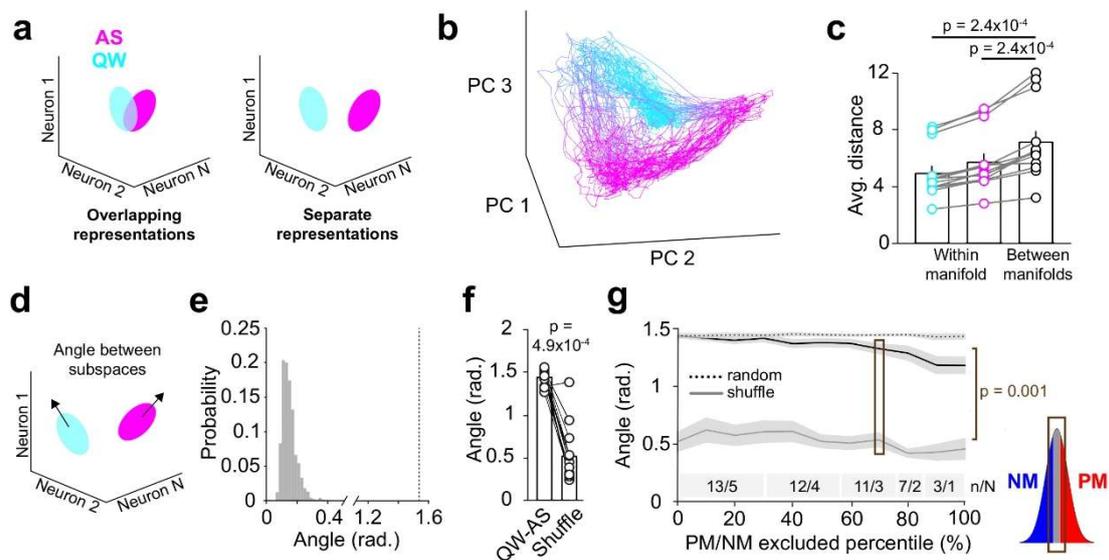


Fig. 3. Structure of population activity reveals separated orthogonal coding spaces during different behavioral states. (a) Schematic diagram illustrating possible overlapping (left) and separate (right) representations in neural activity space of the active state (AS, magenta) and quiet wakefulness (QW, cyan). (b) First three principle components (PCs) of parallel fiber population activity for a single experiment. Manifolds representing AS and QW and the transitions between them. Magenta to cyan color change indicates a continuous AS-QW scale for the state dimension (Methods). (c) Plot showing the average Euclidean distance between all pairs of neural activity patterns ($\Delta F/F$) within the QW manifold (cyan; 4.8 ± 0.5 ; mean \pm s.e.m.) within the AS manifold (magenta; 5.4 ± 0.5), or between the two manifolds (black; 6.9 ± 0.7). Each circle represents a different experiment ($n = 13, N = 5$; two tailed Wilcoxon signed rank test). (d) Schematic depicting quantification of the angle between the AS and QW subspaces (i.e., hyperplanes in which the AS and QW manifolds are embedded). (e) Example of null distribution obtained by calculating the angle between two halves of the data, after shuffling time for an individual experiment. Dashed line indicates the observed angle between AS and QW manifolds in the same experiment. (f) Plot showing angle between AS and QW manifolds, compared to the mean angle between random halves of the data after shuffling timepoints. Each circle indicates a different experiment ($n = 13, N = 5$; two-sided Wilcoxon signed rank test). (g) Angle between AS and QW manifolds (black) as increasing fractions of the most strongly positively and negatively modulated fibers are excluded. Schematic (right) depicts a distribution of the change of $\Delta F/F$ with positively modulated (PM, red), negatively

modulated (NM, blue), and non-modulated (grey) parallel fibers listed (cf. Fig. 1b). Brown box indicates the parallel fibers analysed when the 70th percentile is excluded (two-sided Wilcoxon signed rank test with Bonferroni correction). The gray curve indicates the shuffle control, and dotted black curve indicates the random control, in which the same number of neurons are analysed, but randomly sampled across the distribution. Grey boxes at bottom show the number of experiments (n) and animals (N) analysed. Shading indicates s.e.m.

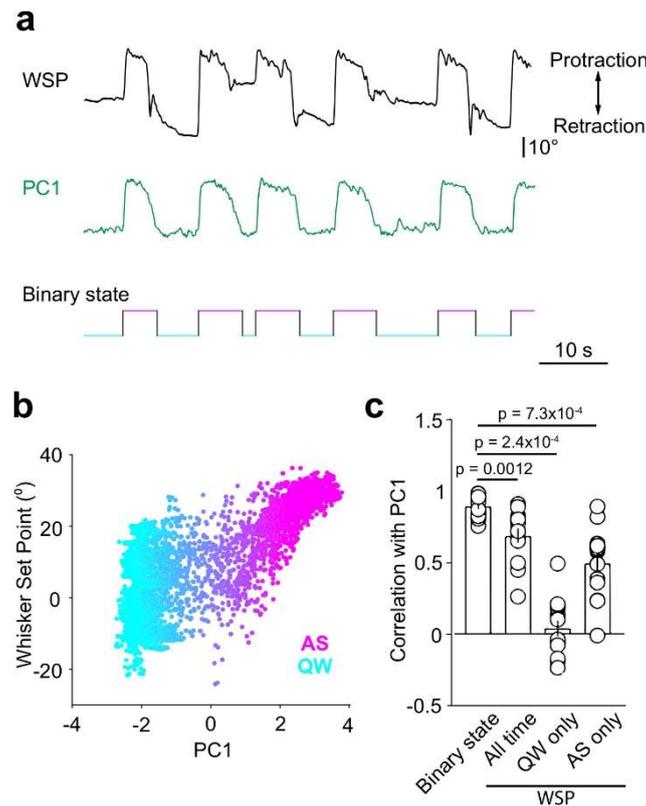


Fig. 4. Widespread parallel fiber population activity is correlated with changes in behavioral state. (a) Whisker set point (WSP; black) and first principal component (PC1; green) of parallel fiber population activity from a single experiment, together with binary representation of state. (b) WSP plotted against the first principal component (PC1) for the same experiment as (a). Color indicates a continuous active state (AS) to quiet wakefulness (QW) scale for the state dimension (Methods). (c) Correlation values between PC1 and different behavioral variables: binary state, or WSP over all time, during QW or AS. Each circle indicates a different experiment (n = 13, N = 5; two-sided Wilcoxon signed rank test). Error bars denote s.e.m.

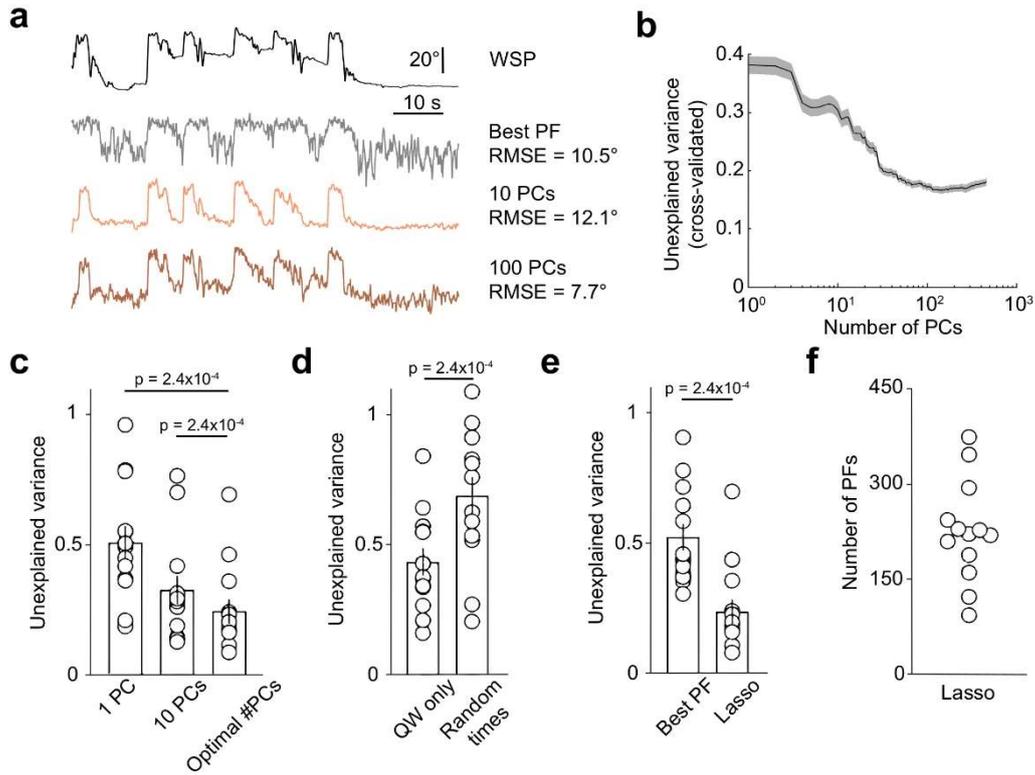


Fig. 5. Distributed representation of sensorimotor dynamics (a) Whisker set point (WSP) during the same experiment shown in figure 4a and 4b over a different period. Measured WSP (black), and its reconstruction using linear regression over the best performing parallel fiber (grey), first 10 principal components (PCs) (orange), and first 100 PCs (brown). Reconstruction error for each case is indicated as root mean square error (RMSE). (b) Example of unexplained variance (cross-validated) for WSP (an assay of the error in decoding performance) as a function of the number of PCs used for linear regression (same experiment as in Figure 4a, 4b and 5a). Shading indicates s.e.m. over random draws of held-out data. (c) Plot of the average cross-validated unexplained variance for WSP based on the first PC, the first 10 PCs, and the optimal number of PCs. Each circle indicates a different experiment ($n = 13$, $N = 5$; two-sided Wilcoxon signed rank test). (d) Plot of the average cross-validated unexplained variance for WSP during QW for a decoder trained only on QW times, compared to a decoder trained on random times across the experiment ($n = 13$, $N = 5$; two-sided Wilcoxon signed rank test). Both decoders were based on their optimal number of PCs, and were tested on the same held-out data during QW. (e) Plot of the average cross-validated unexplained variance for WSP based on the best parallel fiber (PF) for each recording and for lasso regression on parallel fiber population activity ($n = 13$, $N = 5$; two-sided Wilcoxon signed rank test). (f) Range of optimal number of parallel fibers to minimize the cross-validated unexplained variance in f. Each marker represents a different experiment. Error bars in c, d and e denote s.e.m.

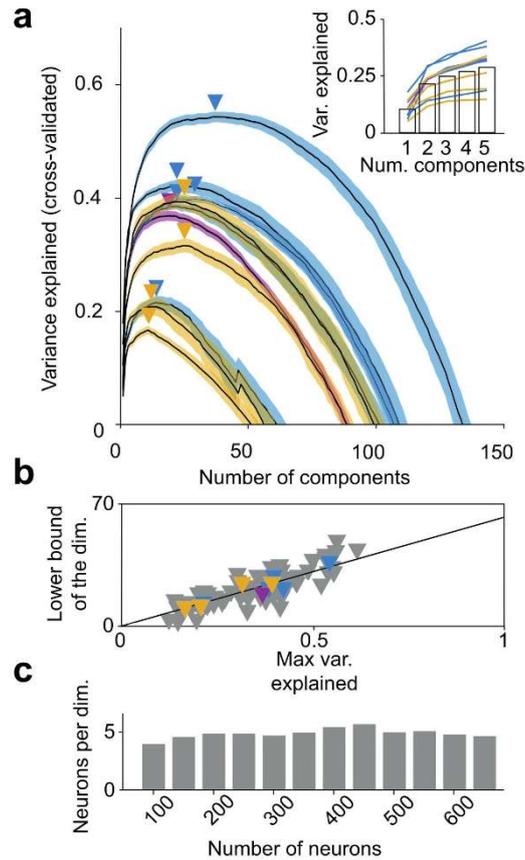
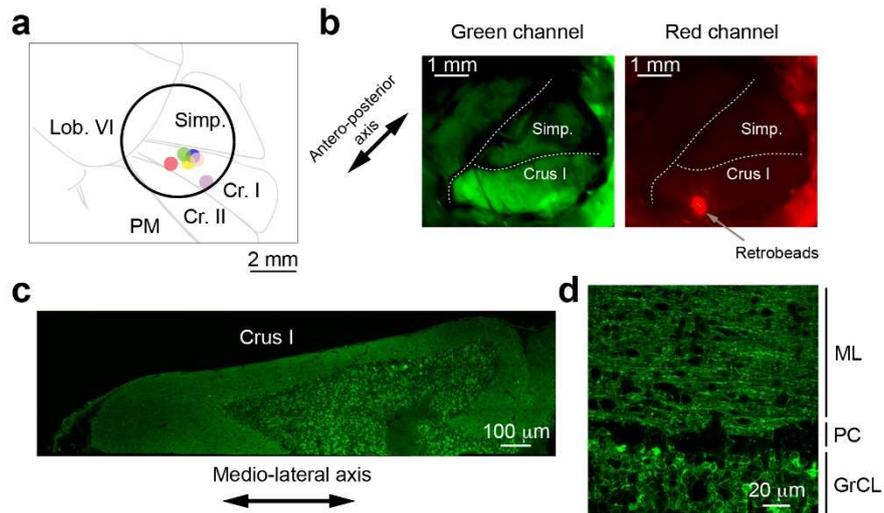
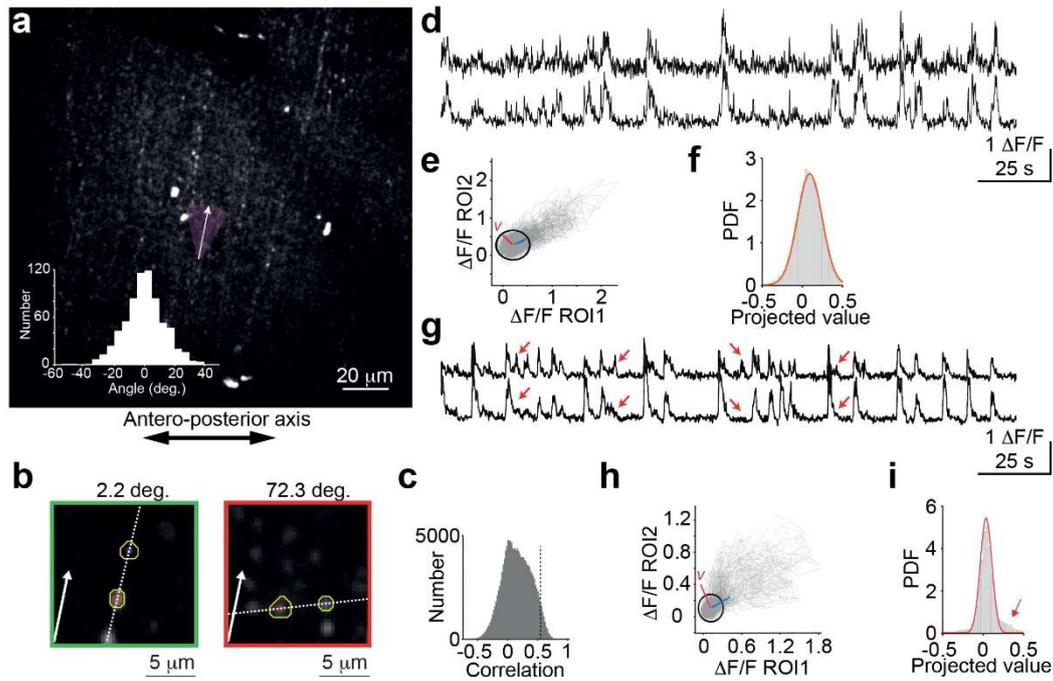


Fig. 6. Dimensionality of population activity during spontaneous behaviors. (a) Relationship between the variance of the population activity explained and number of principal components (PCs) based on cross-validated principal component analysis (PCA). Each black line represents the mean variance explained for a single experiment (all data randomly subsampled to 300 axons). Shading represents s.e.m. across different randomly subsampled populations and colors indicate different animals ($n = 10$, $N = 3$). The arrowheads represent the lower bound of the dimensionality for each experiment. Inset: Expanded region from main panel. Black bars indicate average over experiments. (b) Relationship between the lower bound of the dimensionality and the maximum variance explained. Grey and colored arrowheads indicate individual subsamples of held-out data and means for each experiment, respectively. Linear extrapolation predicts that 62 dimensions are necessary to explain all the variance for populations of 300 parallel fibers. (c) The ratio of number of neurons to the extrapolated dimensionality for all subsampled population sizes, ranging from 100 ($n = 13$, $N = 5$) to 650 parallel fibers ($n = 2$, $N = 1$).



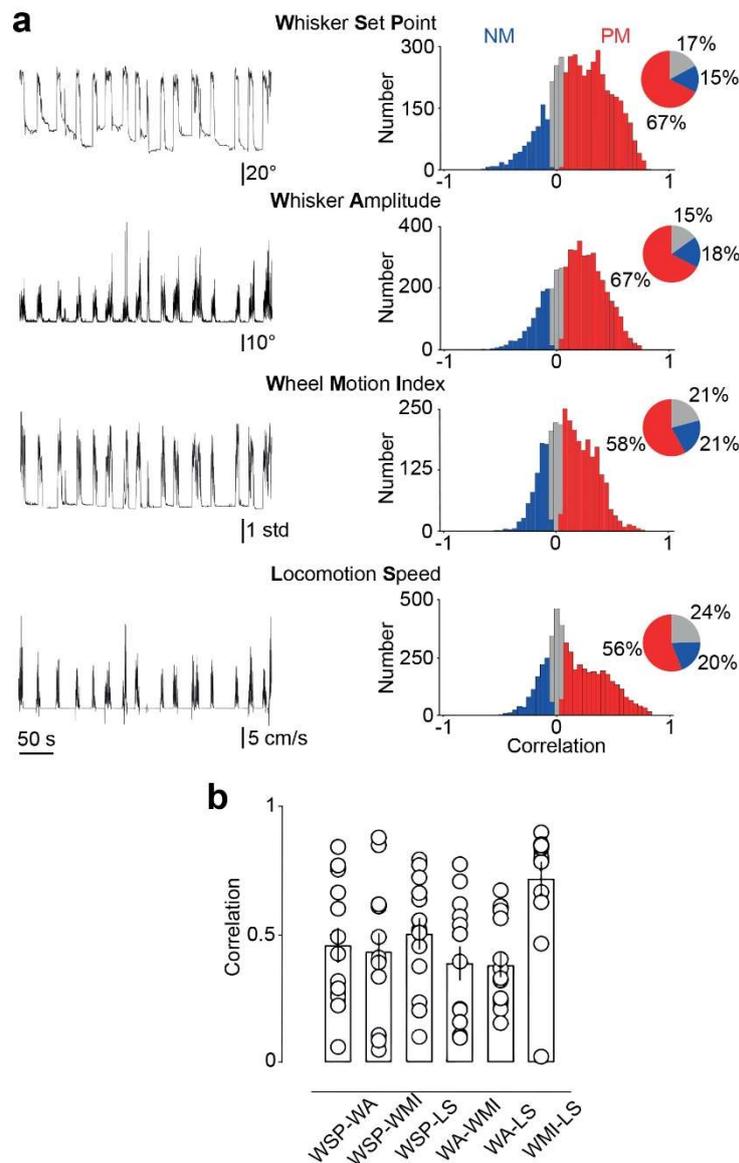
Extended Data Figure 1. Expression of GCaMP6f in granule cells in *Slc17-a7* Cre mice.

(a) Schematic representing a dorsal view of the cerebellum. The black circle represents the 5 mm cranial window above Crus I. Colored blobs show the approximate location of the virus injection and GCaMP6f expression for the animals in this study. (b) Top view of a cranial window above Crus I. The green channel (left) shows expression of GCaMP6f in lobule Crus I. Green fluorescence is widespread due to the spatial extent of parallel fiber projections. The red channel (right), shows a clump of retrobeads at the injection site (arrow). (c) Confocal tile scanning of a coronal section of Crus I where granule cells (GrCs) were transfected with GCaMP6f. Note the absence of labelled cell bodies in the molecular and Purkinje cell layers. (d) Confocal image with a smaller field of view to show GCaMP6f expression in GrC somata and axons. **Labels:** Cr. 1: Crus1 lobule; Cr. 2: Crus2; lob. VI: cerebellar lobule VI in the vermis, Simp.: simplex lobule, PM: paramedian lobule, ML: molecular layer, PC: Purkinje cell, GrCL: GrC layer.



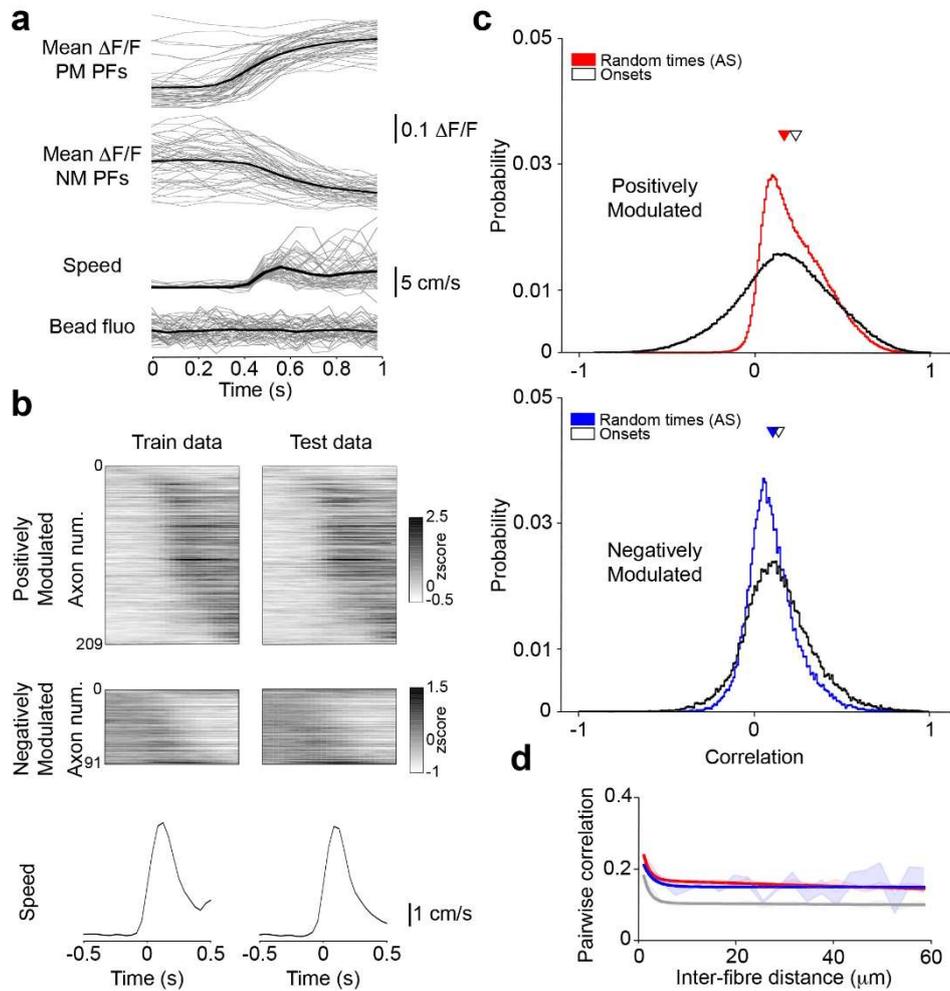
Extended Data Figure 2. Method of grouping varicosities into putative axons. (a) Strings of bright varicosities from active axons were traced by hand to obtain orientations of parallel fiber segments. Inset shows the histogram of the angle of individual parallel fiber segments from the average parallel fiber orientation ($n = 13$, $N = 5$). White arrow indicates average parallel fiber orientation for this experiment, and purple the acceptance angle for parallel fiber identification (two standard deviations of the distribution in the inset). (b) Examples of candidate varicosity groupings that pass (left, green box, each side $13.7 \mu\text{m}$) and fail (right, red box) the first grouping criterion. Varicosities indicated by yellow contours. Title indicates angle between candidate parallel fiber given by linear fit (dotted white line) and the average parallel fiber direction for that experiment (white arrow). (c) Example histogram of correlation coefficient for pairs of varicosities in different patches, used for the second grouping criterion. Dotted line indicates the threshold correlation (95th percentile) for this experiment. (d-f) Example of correlated varicosities that pass the third grouping criterion. (d) Example activity of the two varicosities ($r = 0.74$). (e) Activity of varicosity 1 plotted against activity of varicosity 2 (grey). Blue line indicates fit from linear regression. Black circle indicates baseline activity distribution (95% confidence interval). Red line indicates vector v onto which activity is projected to calculate the linear deviation ratio for the third criterion. (f) Histogram of activity from (e) projected onto v (grey histogram), and analytically calculated distribution of the baseline distribution projected onto v (orange curve). The ratio of the variances of these distributions is used for the third criterion (linear deviation ratio = 1.03). (g-i) Same as d-f for a pair of varicosities that fail the third grouping criterion ($r = 0.69$, linear deviation ratio =

3.12). Red arrows in (g) indicate transients that are missed in one varicosity. Red arrow in (i) shows the large tail of the distribution.



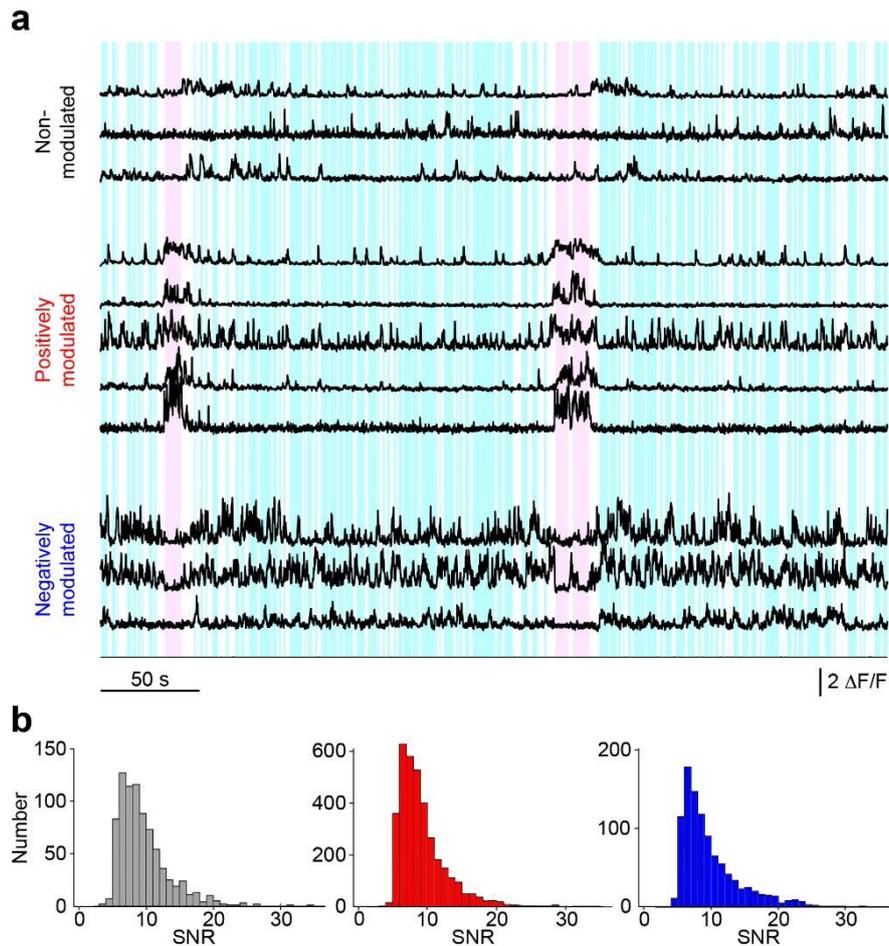
Extended Data Figure 3. Correlated locomotion speed and whisking during spontaneous behavior. (a) Left: Example traces of different behavioral variables: whisker set point (WSP), whisking amplitude (WA), wheel motion index (WMI), and locomotion speed (LS). Right: Histograms of correlations of parallel fiber Ca^{2+} activity ($\Delta F/F$) with WSP, WA, WMI and LS ($n = 13$, $N = 5$). Red and blue indicate parallel fibers that are positively or negatively correlated with each behavioral variable respectively ($p < 0.05$, two-sided shuffle test). Grey indicates parallel fibers that are not significantly correlated with that behavior. Pie charts reveal a similar fraction of positively modulated (PM, 58 – 67%), negatively modulated (NM, 16 – 22%) and non-modulated GrCs (13 – 20%) regardless of behavioral variable. (b) Correlation between all pairs of behavioral variables for each experiment (grey circles). Black bars indicate mean

across experiments ($p = 2.4 \times 10^{-4}$ for all pairs of behavioral variables, two-sided Wilcoxon signed rank test, $n = 13$, $N = 5$). Error bars indicate s.e.m.

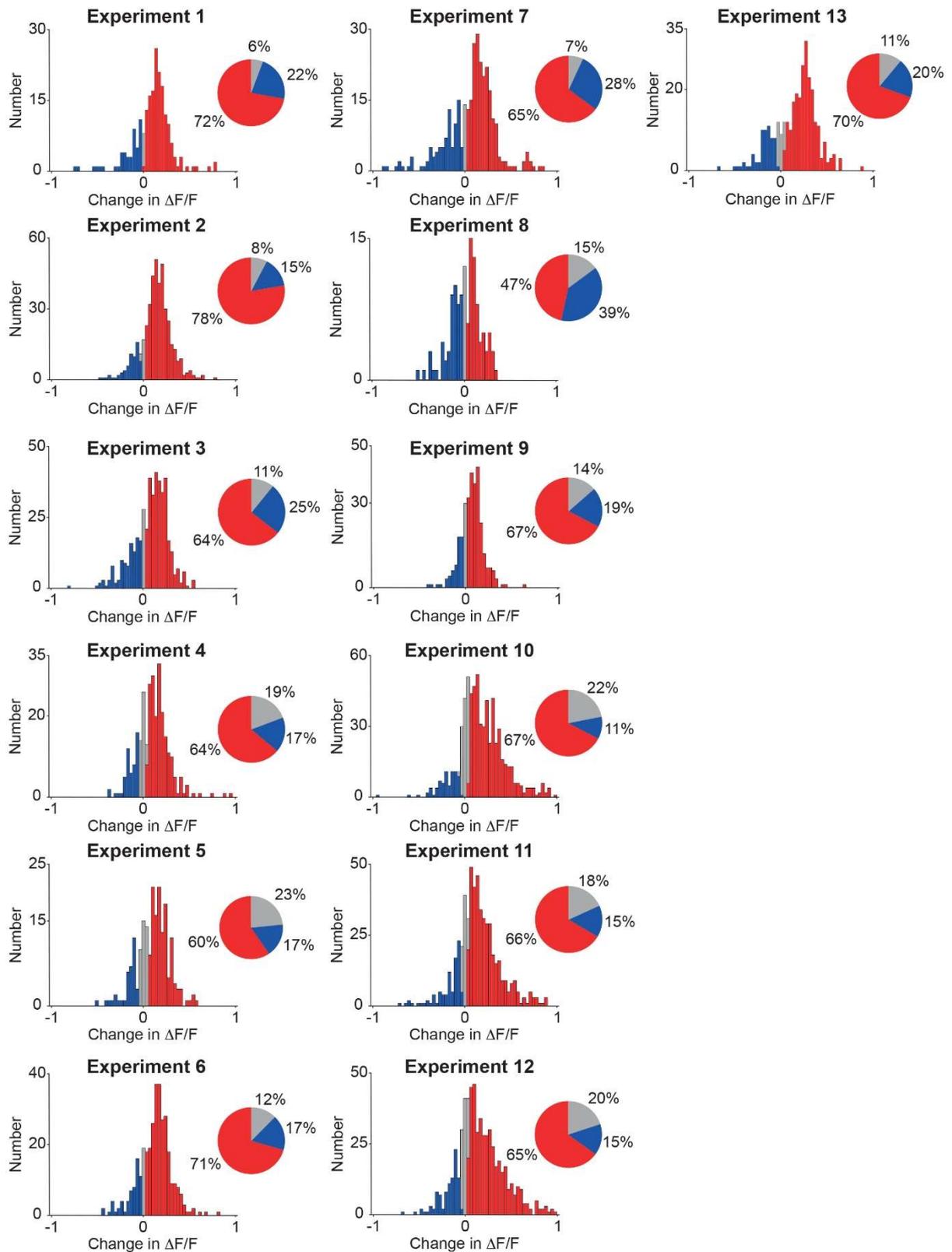


Extended Data Figure 4. Pairwise correlation and spatial dependence of parallel fiber correlations at the onset of locomotion. (a) $\Delta F/F$ traces of positively modulated (PM) and negatively modulated (NM) parallel fibers in grey (top) together with locomotion speed and bead fluorescence, from a single experiment aligned at locomotion onset. Bottom: Grey indicates individual traces and the black indicates the mean. (b) Example experiment showing temporal dispersion of parallel fiber activation during locomotion onsets. Top and middle panels show average $\Delta F/F$ (zscored) of PM and NM parallel fibers calculated over locomotion onsets. Locomotion onsets were randomly split into training (50%) and test (50%) data, and parallel fibers were sorted according to the time lag of their peak correlation (PM) or anticorrelation (NM) with locomotion speed during the training data. Bottom panels show average locomotion speed during training and test onsets. (c) Distribution of pairwise correlations for pairs of positively (black, top) and negatively (black, bottom) modulated parallel fibers during 1s interval surrounding locomotion onsets ($n = 12$, $N = 5$). Red and blue

curves indicate distributions of correlations during random periods in the active state (for positively modulated and negatively modulated parallel fiber pairs, respectively). Arrowheads represent the means. **(d)** Relationship between correlations between putative axons at locomotion onsets as a function of inter-fiber distance, for positively modulated pairs (red), negatively modulated pairs (blue), and all pairs (grey; $n = 12$, $N = 5$). Shaded regions indicate s.e.m. Thick lines indicate double exponential fit to the data.

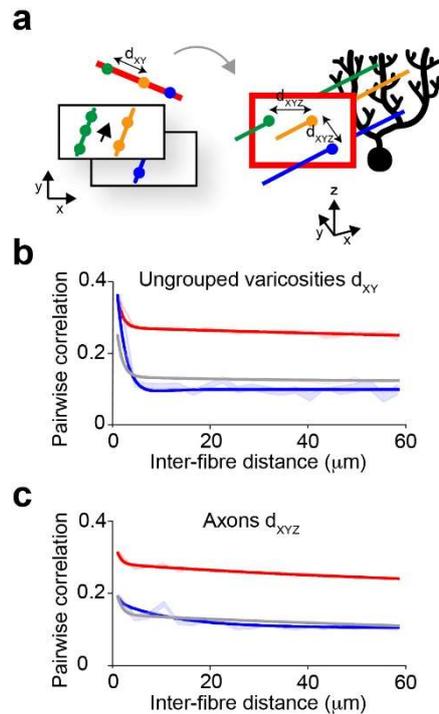


Extended Data Figure 5. Non-modulated parallel fibers are not noisier than modulated parallel fibers. **(a)** Example of three non-modulated parallel fibers (top) compared to positively modulated and negatively modulated parallel fibers (same example shown in Fig. 1e for full experiment). Magenta/cyan indicates AS/QW. **(b)** Distribution of signal-to-noise ratios (SNRs; Methods) for all non-modulated parallel fibers (top), as well as positively modulated (centre) and negatively modulated parallel fibers (bottom) ($n = 13$, $N = 5$).

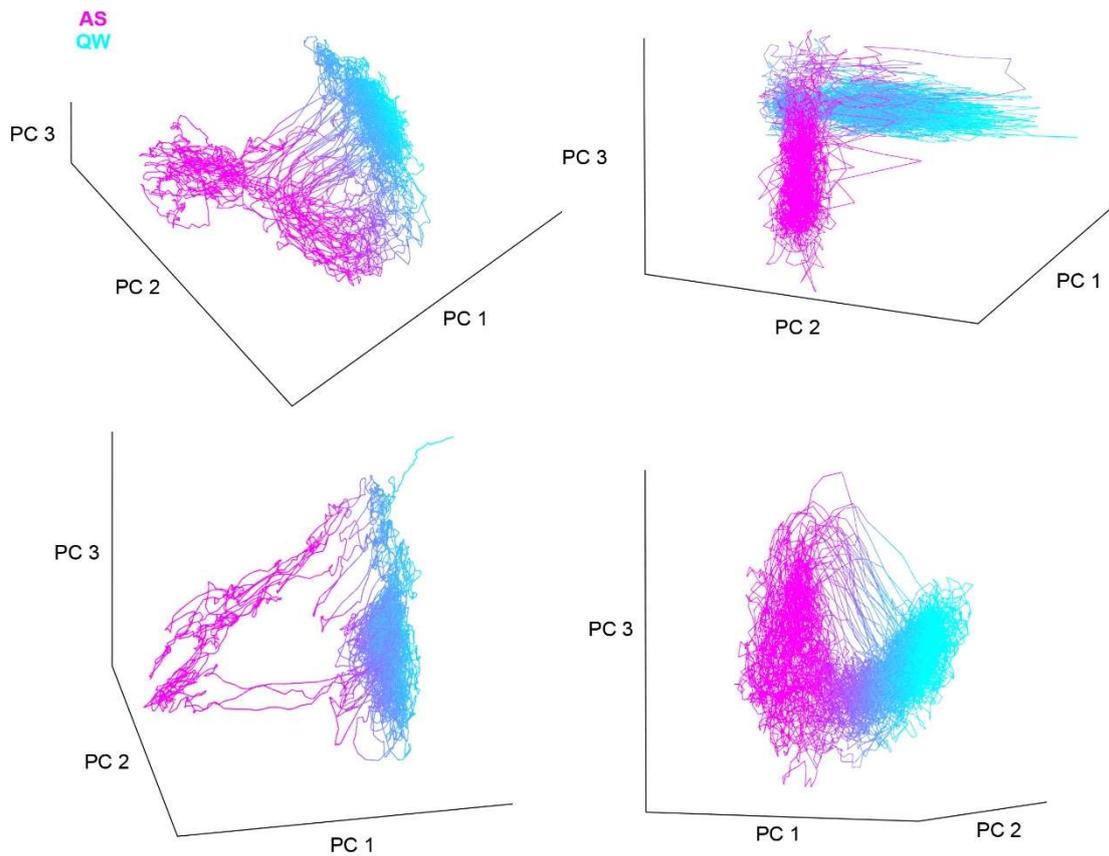


Extended Data Figure 6. Fraction of positively, negatively and non-modulated parallel fibers across experiments. Histograms of changes in $\Delta F/F$ response during the AS relative to QW across all parallel fibers for all 13 experiments across 5 mice. Positively modulated (red)

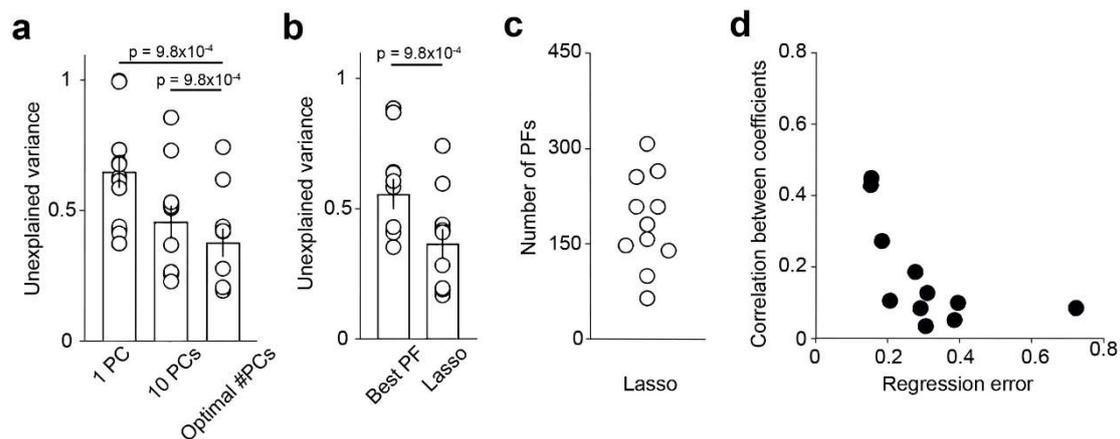
and negatively modulated (blue) parallel fibers, as well as parallel fibers which were not significantly modulated by behavioral state (grey). Pie charts indicate the proportion of each class across experiments.



Extended Data Figure 7. Spatial profile of parallel fiber correlations. (a) Schematic illustrating how distances between parallel fibers were calculated. Left: example of two patches with three parallel fibers, each with different numbers of varicosities. Black unidirectional arrow indicates average parallel fiber direction. To calculate the distance between parallel fibers, the position of the centre of its varicosities is projected onto the dimension orthogonal to the average fiber vector (red line). The XY distance (d_{XY}) is the distance in the projected dimension. Right: Same schematic, rotated to show Z-dimension. The XYZ distance (d_{XYZ}) is the distance in the projection plane (red). (b and c) Correlations between varicosities or putative axons as a function of inter-fiber distance, for positively modulated pairs (red), negatively modulated pairs (blue), and all pairs (grey; $n = 13$, $N = 5$). Shaded regions indicate s.e.m. Thick lines indicate double exponential fit to the data. (b) Correlations and XY distances (d_{XY}) for ungrouped varicosities (within the same patch). Note similar trend to grouped data, except for stronger peak at small distances ($< 2 \mu\text{m}$) (c.f. Fig. 1c). (c) Correlations and XYZ distances (d_{XYZ}) for putative axons across all patches.

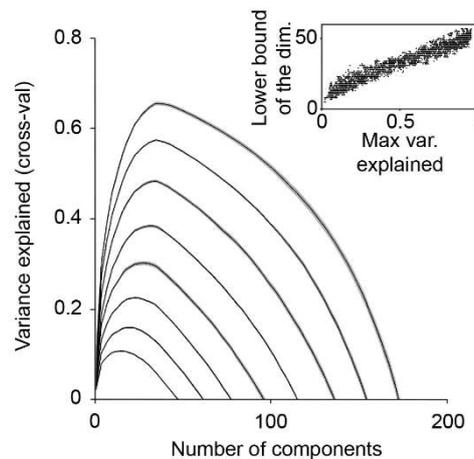


Extended Data Figure 8. Manifold structure across different mice. Parallel fiber population activity visualized by plotting first three principal components. Each panel indicates a different mouse ($N = 5$ in combination with Fig. 3b). Color indicates projection along the quiet wakefulness (QW; cyan) to active state (AS; magenta) state dimension.



Extended Data Figure 9. Distributed representation of locomotion speed. (a) Average cross-validated unexplained variance for locomotion speed based on the first principal component (PC), the first 10 PCs, and the optimal number of PCs. Each circle indicates a different experiment ($n = 11$, $N = 5$; two-sided Wilcoxon signed rank test). (b) Average cross-validated unexplained variance for locomotion speed based on the best parallel fiber (PF) for

each recording and for lasso regression on the population activity ($n = 11$, $N = 5$; two-sided Wilcoxon signed rank test). **(c)** Range of optimal number of parallel fibers to minimize the cross-validated unexplained variance. Each marker represents a different experiment. **(d)** Correlation between the lasso regression coefficients of the optimal decoders for locomotion speed and for whisker set point, plotted against average decoder error (unexplained variance averaged for speed and whisker set point; two-sided Spearman correlation: $r = -0.73$, $p = 0.02$; $n = 11$, $N = 5$). For each decoder, regression coefficients were averaged over 10 random samples of training/test data. Error bars in a and b denote s.e.m.



Extended Data Figure 10. Lower bound of dimensionality increases linearly with maximum variance explained in simulated data. We tested our procedure for estimating dimensionality in a simple model of random 60-dimensional representations in populations of 300 neurons corrupted with increasing levels of noise. Each black line represents the mean variance explained for a fixed standard deviation of the noise distribution. Shading represents s.e.m. across different random representations. Inset: linear relationship between lower bound of the dimensionality and the maximum variance explained.

Methods

Animal preparation for in vivo imaging

All experimental procedures were approved by the UCL Animal Welfare Ethical Review Body and the UK Home Office under the Animal (Scientific Procedures) Act 1986. To specifically express the Ca²⁺ indicator GCaMP6f⁵¹ in GrCs, we used the *Slc17a7-IRES-Cre* transgenic line^{52,53}, which is known to express Cre recombinase in VGlut1-expressing excitatory neurons. In the cerebellar input layer, GrCs are the only neurons expressing VGlut1⁵²⁻⁵⁴, which restricted the expression of GCaMP6f to this neuronal population. Stereotaxic injections were performed under sterile conditions on 6 - 12 weeks heterozygous *Slc17a7-IRES-Cre* mice (male and female). Following analgesic injection with buprenorphine (0.1 mg/kg), mice were deeply anesthetized with a ketamine/xylazine mix (100:10 mg/kg) and mounted in a stereotaxic frame (Kopf Instruments). 5 µl pipettes (Blaubrand 7087-07) were pulled on a Sutter P97 micropipette puller, cut to 10 - 20 µm internal diameter and suction filled with AAV9.CAG.Flex.GCaMP6f.WPRE.SV40 (AV-9-PV2816 - Upenn Vector Core). In 3 animals red retrobeads IX (0.02 - 0.2 µm, Lumafluor) were mixed with the AAV to be used as tracking objects for real-time movement correction (diluted 1:1000). A small craniotomy was performed above the injection site and the pipette slowly lowered to minimize tissue damage at coordinates of the cerebellar hemisphere in the Crus I region (6.5 mm anterior to Bregma, 2.5 mm lateral to the midline and 0.2 mm from the pia). A single injection of ~100 nl of virus was performed via a Toohey Spritzer pressure system (Toohey Company). Analgesia (bupivacaine 0.05 %) was then administered to the surgical wound site. Post-surgery, atipamezole (1 mg/kg) was administered for xylazine reversal.

Headplate and cranial window surgery

After 3 to 8 weeks of AAV expression, mice were implanted with a head plate for imaging. Mice received pre-surgery injections of dexamethasone (1 mg/kg), atropine (0.04 mg/kg) and carprofen (5 mg/kg) prior to induction of anaesthesia with a mixture of Fentanyl (0.075 mg/kg), Medetomidine (0.75 mg/kg) and Midazolam (7.5 mg/kg). Viscotears liquid eye gel application was used to prevent dehydration and body temperature was maintained throughout the surgery with a heat pad and temperature controller system (FHC, Inc.) After removal of overlying skin, a custom head plate was centred above Crus I and attached to the skull using dental acrylic cement (Paladur, Kulzer). A 5 mm craniotomy was performed over the Crus I region and the exposed cerebellar cortex cleared with sterile cortex buffer (125 mM NaCl, 5 mM KCl, 10 mM glucose, 10 mM HEPES, 2 mM MgSO₄, 2 mM CaCl₂ [pH 7.4]) to wash blood and remaining

debris from the craniotomy. The craniotomy was then sealed with a 5 mm glass coverslip (630-2112 VWR) and fixed with Cyanoacrylate glue. In 2 mice, red fluorescent beads (4 μm fluospheres, ThermoFisher) suspended in sterile cortex buffer (diluted 1:100) were placed between the coverslip and the brain surface to perform real-time movement correction. Post-surgery analgesia (buprenorphine 0.1 mg/kg) was administered prior to anaesthesia reversal via atipamezole (3.75 mg/kg), flumazenil (0.75 mg/kg) and naloxone (1.8 mg/kg). Mice were group housed and kept on a 12:12 h light dark cycle with food and water ad libitum.

In vivo two-photon imaging of head-fixed mice

Two-photon imaging was performed with an acousto-optic lens (AOL) 3D two-photon microscope which enables high-speed 3D random-access pointing and scanning^{27,55,56} and real-time movement corrected imaging⁵⁷. The excitation source was a Ti-sapphire laser (Chameleon Vision, Coherent) tuned to 920 nm and the optical configuration was set up to underfill a 20 X (1.0 NA, Olympus) objective. This gave an illumination NA of 0.6 - 0.7 and a two-photon point spread function of $0.69 \pm 0.04 \mu\text{m}$ in X-Y and $6.54 \pm 0.27 \mu\text{m}$ in Z (full width half maxima, mean \pm s.d.) as previously reported⁵⁷. The illumination power was controlled with a Pockels cell (Model 302CE, Conoptics) and was typically 60-70 mW at the back aperture of the objective. A two-channel data acquisition (DAQ) system was deployed using GaAsP photomultiplier tubes (PMTs) (H7422, Hamamatsu, Japan) in both the red and green channels. PMT outputs were digitized using high-speed 800 MSPS ADCs (NI-5772, National Instruments) via 200 MHz Pre-Amplifiers (Series DHPKA 100/200 MHz, FEMTO). A digital acquisition FPGA board (NI FlexRIO – 7966R, National Instruments) was used to down-sample the signals by integrating each pixel before sending frames to the host PC via the National Instruments PXIe interface. The 3D imaging was controlled with the custom SilverLab 3D imaging software (LabView, National Instruments). The microscope user interface acted as a master for the video acquisition system.

Two weeks after surgery, animals were habituated to the recording apparatus by head restraining them on a cylindrical Styrofoam wheel for thirty minutes per day during three consecutive days before imaging neural activity. A reference bead (0.2 or 4 μm) was identified within the imaging volume ($175 \times 175 \times 116 \pm 6 \mu\text{m}$, $n = 13$) and imaged with voxel dwell-time of 50 ns. Real-time tracking of brain movement and real-time movement corrected imaging was performed in 2D with a 500 Hz update rate⁵⁷. To control for axial brain movement during behavior, we recorded the bead fluorescence during AS and QW. There was no

significant correlation between $\Delta F/F$ of the bead fluorescence and locomotion (correlation coefficient 0.06 ± 0.06 , $n = 13$, $N = 5$). Next, a high-resolution movement corrected Z-stack image was performed by AOL raster scan imaging through the molecular layer. Elongated XY-patch regions-of-interest (ROIs) were then defined in a staircase arrangement at different depths from the pia, with their long axis orthogonal to the direction of the parallel fibers (Fig. 1a). Imaging patches were typically spaced 10 - 12 μm apart in Z. This minimized the chance of recording from the same parallel fiber in different patches. The line scans making up each patch had a voxel dwell-time of 200 – 400 ns. Imaging was performed for sets of 20 s trials lasting 5 min, where mice were free to run on the wheel.

Image processing

Imaging data for each patch were extracted and exported to tiff files by using in-house software written in LabView (National Instruments). The analysis was then performed using scripts and toolboxes in MATLAB. Before extracting calcium data from patches, *post hoc* movement correction was used to correct for any residual movement in the images⁵⁸. In one experiment, where there was more movement, 10 pixels were trimmed from each edge of each patch to improve *post hoc* movement correction. In Extended Data Figure 6d, we quantified residual movement in image patches by quantifying the mean square displacement of an imaged bead following *post hoc* movement correction in a 500 ms time window centered around the onset of locomotion speed (as determined in Extended Data Fig. 4a).

Measurement of whisker position and locomotion

Two video cameras with far IR LED illumination were used to monitor the face and whiskers. Facial areas were recorded at 1280 x 960 resolution at 30 Hz (The Imaging Source), while whisking was recorded at 644 x 484 resolution at 300 Hz (Mako). All behavioral data was acquired with the SilverLab custom software running under LabView (National Instruments). Whisker position was extracted from videos using DeepLabCut⁵⁹ by tracking 3 markers on a single whisker. Whisker angle was measured as the angle between the linear fit between the 3 markers and a line parallel to the whisker pad of the mouse. The angle was denoised using a 30 Hz 4th order forward-backward Butterworth filter. Whisker set point was determined by Gaussian smoothing the whisker angle using a 500 ms window. Whisker amplitude was calculated as the magnitude of the Hilbert transform of the whisker angle after being bandpass filtered using a 4th order Butterworth filter from 8-30 Hz^{60,61}. Besides the locomotion speed of the mice recorded every 2 ms with a rotary encoder, a wheel motion index was calculated using

a small ROI selected on the wheel, as the average difference in pixel values between successive frames⁶², smoothed over 200 ms. This provides an estimate of wheel motion without distinguishing between forward movement (e.g., running) or backward movement (e.g., startle responses). Datasets with no locomotion or whisking were not analysed as no comparison could be made with the representation of the active state in the same population of parallel fibers.

Calcium imaging processing

Parallel fiber varicosities were identified in imaged patches by adapting signal detection tools from a publicly available toolbox⁶³. In brief, following Zhou et al. (2018)⁶⁴, we identified varicosities by first identifying seed pixels, defined as the pixels having a peak correlation with their neighboring pixels. To remove spurious seed pixels due to background noise, we required that this correlation be above a threshold which was determined from the bimodal distribution of pixel correlations over all data. Corresponding spatial filters were then detected by using linear regression to fit the fluorescence of all pixels within a local region (1.7 x 1.7 μm) to the fluorescence trace of the seed pixel for the varicosity. Masks were then defined by thresholding the resulting spatial filter weights at 80% of their total value and trimming overlapping pixels. However, because there was very little overlap between the spatial filters of different varicosities, we did not proceed with demixing the fluorescence data⁶³. For quality control, we removed varicosities with a signal to noise ratio (SNR) below the 95th percentile of the distribution of SNRs of varicosity-sized regions within the neuropil. Following Pnevmatikakis et al., (2016)⁶³, the SNR was defined as the peak $\Delta F/F$ for that varicosity normalized by the noise standard deviation (estimated from the power spectrum). In addition to this, a small number of varicosities (2%) were manually removed following visual inspection for artefacts. Neuropil fluorescence was calculated using masks of size 20 x 20 μm excluding any pixel within twice the average varicosity radius. We also excluded pixels whose correlation with their neighboring pixels was above the 95th percentile, to avoid bleaching in localized saturated regions. The resulting neuropil signal was small and only accounted for $4.6 \pm 4.5 \times 10^{-4} \%$ of the variance of the activity of the corresponding varicosity (Supp. Fig. 4). As a result, this was not subtracted to avoid inflating noise due to low baseline fluorescence. $\Delta F/F$ was then calculated as follows: $\frac{(F-F_0)}{F_0}$ where F is raw fluorescence (averaged over all pixels within the varicosity, or within all varicosities corresponding to the same putative axon after grouping procedure described below) and F_0 is the baseline fluorescence (10th percentile of F).

Varicosity grouping into putative axons

Comparison of the normalized fluorescence transients ($\Delta F/F$) revealed that some varicosities exhibited highly correlated activity. To isolate responses from putative parallel fiber axons, we grouped varicosities using a semi-automated procedure based on correlations in functional activity as well as spatial alignment along the overall direction taken by the parallel fiber population (Extended Data Fig. 2a). For each experiment, we first obtained the average fiber direction by hand tracing small segments of parallel fibers observed in the Z stacks in ImageJ. Next, within each patch, we iteratively grouped a pair of candidate varicosities (or putative parallel fibers) into a new putative axon if the following three criteria were satisfied:

(1) Spatial arrangement: we estimated the putative parallel fiber connecting the candidate varicosities as the best linear fit to the locations of the identified varicosities. If any of the candidate varicosities were further than 1 μm from the fiber, the putative parallel fiber was rejected. We also rejected putative parallel fibers that were orientated at an angle greater than 27° of the average fiber direction for that experiment (this number was estimated as two standard deviations of fiber angles across all datasets; Extended Data Fig. 2a, b).

(2) Functional correlation: we required that the total activity ($\Delta F/F$) correlation between candidate varicosities be greater than a threshold value. For this threshold value, we required a null distribution for the correlation between varicosities on different parallel fibers. For this null distribution, we used the distribution of correlations between varicosities on different patches of the same experiment (Extended Data Fig. 2c), as they were unlikely to be on the same parallel fiber due to our staggered patch arrangement. For the threshold correlation, we took the 95th percentile of this distribution.

(3) Deviation from linear scaling: if the candidate varicosities belong to the same parallel fiber, and assuming they are in the linear regime of GCaMP6f, their activity would be a scaled version of each other. We estimated the deviation from linear scaling as the minimum projected variance of the data. We denote this projection vector as v (Extended Data Fig. 2d-i; we also tried using the vector orthogonal to the vector obtained from linear regression, which yielded similar results). To take into account varying noise levels, we normalized this quantity by the variance of the distribution of baseline fluorescence projected onto v . The baseline distribution, which presumably represented noise, was obtained by fitting a mixture of two 2D Gaussians to the activity of the varicosities, and taking the lower-mean Gaussian. Finally, we rejected all pairs for which this “linear deviation ratio” exceeded 1.5.

After this automated procedure, we visually inspected all data for clear misclassifications, which we corrected manually, including misses (due to sparsely active varicosities whose low

event rates precluded the second condition, 6.0% of total groupings), and false positives (due to varicosities with visually distinct events missed by the third condition, 16.4% of total groupings). This resulted in an average of 1.27 varicosities per putative parallel fiber. To cross check our grouping algorithm, we measured the distance between neighbouring varicosities within each putative axon comprising more than one identified varicosity, as this was not used as a criterion in our procedure, and found that it was consistent with the mean and range of intervaricosity distances previously reported²⁸ (Fig. 1d). The analysis was repeated for every patch placed in the volume for the recording session. All putative axons, from the different patches, were used for further analysis.

Granule cell somatic calcium analysis

We performed GrC somatic two photon imaging in two mice used for the parallel fiber imaging. As GrCs are densely packed, we imaged a single plane in the GrC layer with a field of view of 250 μm , rather than using patches at different depths. To identify GrC somata and extract their $\Delta\text{F}/\text{F}$ traces, we used the software package Suite2p⁶⁵ available on GitHub (github.com/cortex-lab/Suite2P).

Identification of the active and quiet wakefulness states

We labelled timepoints as belonging to periods of AS and QW based on behavioral recordings. We first smoothed whisker amplitude and wheel motion index over 500 ms, centred around their modes, and normalized by their standard deviation. Timepoints in which these assays of locomotion and whisking variables were both below 0.1 were defined as QW; periods in which they were above 0.1 for at least 3 seconds were defined as AS. Note that the criteria are purposefully strict to avoid mislabelling.

Definition of positively and negatively modulated parallel fibers

For each parallel fiber, we calculated the difference between the average $\Delta\text{F}/\text{F}$ during AS and the average $\Delta\text{F}/\text{F}$ during QW. We then calculated the two-sided p-value compared to a null distribution obtained with 1000 trials in which we shuffled time in 1 s blocks. Positively and negatively modulated parallel fibers were defined by having a significant increase or decrease in mean $\Delta\text{F}/\text{F}$ during AS compared to QW, when compared to the shuffle control ($p < 0.05$).

Analysis of the spatial structure of parallel fiber activity

To analyse spatial structure in parallel fiber activity, we calculated the correlation coefficient of parallel fiber $\Delta F/F$ over the full recording as well as the distance between fibers. The XY distance (Fig. 2c and Extended Data Fig. 7b) was calculated for each patch within an experiment. For each putative axon, we first found the centroid of the spatial filter (comprising the ROIs for each varicosity associated with that axon), then projected each centroid onto the dimension orthogonal to the average fiber direction for that experiment. The distance between the projected centroids was then the distance between fibers in the same patch. For the XYZ distance (Extended Data Fig. 7c), we instead considered all pairs of parallel fibers in an experiment (across all patches) by projecting the centroids of each parallel fiber onto the plane orthogonal to the average fiber direction, and calculating the 2D distance between centroids. In the nearest-neighbor (NN) analysis, NN distances were calculated as the average distance of each positively modulated parallel fiber to the nearest positively modulated parallel fiber (similar for negatively modulated). In the shuffle control, positively modulated (or negatively modulated) labels were randomly shuffled.

Identification of locomotion onset and analysis of activity

We calculated the correlations between parallel fibers during locomotion onsets (Extended Data Fig. 4). Locomotion was defined as any time point in which wheel speed exceeded 1.5 cm/s. Locomotion onsets were identified by finding locomotion time points with a gap of at least 500 ms from the previous instance of locomotion. The correlation coefficient between parallel fibers was calculated after concatenating 1 s periods centered around every locomotion onset in the experiment. To compare these defined behaviors against the multiple behaviors encompassed in the AS, we also calculated the correlation coefficient during the same number of time points randomly selected within the AS.

Analysis of manifolds

To analyse the structure of the activity space associated with different behaviors, we defined the AS manifold as the set of neural activity patterns during timepoints labelled as AS (similarly for the QW representation). Unlabelled timepoints were excluded from these definitions. For visualization purposes only, in some figures we additionally labelled all timepoints (including unlabelled timepoints) according to a continuous AS-QW scale (Fig. 3b, 4b, Extended Data Fig. 8). To do this, we projected the data onto the ‘state dimension’ (the vector separating the

means of the AS vs QW representations), z-scored, and capped the resulting value between -1 to 1.

The orthogonality of the manifolds was assessed by measuring the first principal angle between the QW and AS subspaces. To calculate the principal angle, we first found a planar embedding for both AS and QW manifolds. We used singular value decomposition to find a rank-2 approximation to the population activity during the AS as $X_{AS} \approx U_{AS}S_{AS}V_{AS}$ (similarly for QW). The first principal angle between U_{AS} and U_{QW} is given by⁶⁶

$$U_{QW}^r = U_{QW} - U_{AS}'U_{AS}U_{QW}$$

$$\theta = \arcsin(\sigma_{max}(U_{QW}^r))$$

where $\sigma_{max}(U_{QW}^r)$ is the maximum singular value of U_{QW}^r (the residual of U_{QW} that is orthogonal to U_{AS}). We calculated θ using the function `subspace.m` in MATLAB. If U_{QW} and U_{AS} are orthogonal, the residual U_{QW}^r and thus the angle θ will be large. In a vector space of high extrinsic dimensionality (i.e., large number of neurons), random vectors are likely to be orthogonal. Therefore, to ensure that the orthogonality of the AS and QW subspaces is a feature of the data, and not due to added measurement noise, we compared to a control in which we first shuffled the time indices of the data in 1 s blocks, then calculated the principal angle between the first and second halves of this random data. This shuffling breaks the structure of the two defined manifolds so that they no longer represent different behavioral states. For manifold analyses, datasets with fewer than 100 axons were excluded.

Linear regression

We used cross-validated linear regression to predict a behavioral variable (whisker set point or locomotion speed) either based on the first K PCs (principal component regression) or on parallel fiber population activity (lasso regression). For training, we used 80% of the data (in random 1 s blocks), and calculated the error as the fraction of unexplained variance of the behavioral variable in the held-out data (Fig 5a, b). This was repeated for 10 random samples of training/test data to obtain the average cross-validated unexplained variance. For principal component regression, the ‘optimal’ K was defined as the number of PCs that minimized this average cross-validated unexplained variance (Fig. 5b). We used a similar protocol for lasso regression, instead varying the penalty from $\lambda = 10^{-3}$ to 1. To determine whether GrC representations are distributed across the population, we quantified the number of parallel

fibers with a nonzero coefficient at the optimal λ . For comparison, we also calculated the unexplained variance when regressing against a single neuron ('best' parallel fiber, i.e., which minimized the unexplained variance). Finally, to quantify QW-only decoding, we repeated principal component regression constraining both the test data and the training data to QW periods (taking the optimal K). For a control, we compared the performance of the QW-only decoder to one in which the timestamps of the training data were randomly sampled as 1s blocks from all timepoints in the experiment (combining both QW and AS periods). The shuffled decoder was tested on the same held-out data as the QW-only decoder.

Dimensionality analysis

We used a bi-cross-validated version of principal component analysis (PCA) to infer the dimensionality of neural representations during spontaneous behavior^{16,18,64,67}. To control for differing population sizes across experiments, we randomly subsampled a fixed number of parallel fibers from the population. We randomly selected 80% of the data (training data X^{tr} , chosen in 1 s blocks) to calculate the first K PCs, resulting in the following low-rank approximation:

$$X^{tr} = US^{tr}V^{tr}$$

To cross-validate these PCs, we split the remaining 20% of the test data into a second partition of training neurons (X_1^{te} , 80% of the population) and test neurons (X_2^{te}). The low-rank matrix decomposition for the test data can be written in block format:

$$[X_1^{te}; X_2^{te}] = [U_1; U_2]S^{te}V^{te}$$

We use the upper block to estimate the latent dynamics ($S^{te}V^{te}$) via linear regression, and use the lower block to predict X_2^{te} . Note that this linear regression step is only well-defined if the latent dynamics is shared across neurons. The lower bound of the dimensionality is the number of PCs required to maximize the explained variance of X_2^{te} . For each experiment, this procedure was repeated for 10 random samples of the population, as well as random selections of training/test data. The extrapolated dimensionality was then inferred as the number of PCs that would be required to attain 100% variance explained, using linear extrapolation across experiments. To validate this procedure, we tested a simple model (Extended Data Fig. 10) in

which we used exponentially distributed singular values (S) and random orthonormal vectors (U , V) to create a 60-dimensional representation embedded in a space of 300 (extrinsic) dimensions. We then tested our procedure with different amounts of zero-mean normally distributed noise, verifying that the lower bound of the dimensionality increases linearly with the maximum variance explained (Extended Data Fig. 10).

Statistical analyses

All statistical tests were two-tailed. All error bars indicate s.e.m. Throughout the manuscript, n refers to the number of experiments, N to the number of animals.

Code availability

The SilverLab LabVIEW Imaging Software is available on GitHub at <https://github.com/SilverLabUCL/SilverLab-Microscope-Software>. Analysis scripts are available at <https://github.com/SilverLabUCL/ParallelFibres>.

Data availability statement

Data presented in main figures and extended data figures are available in the data source files or on FigShare (<https://doi.org/10.5522/04/14482977>). Raw data is available on request due to its size.

References

51. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
52. Huang, C.-C. *et al.* Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. *Elife* **2**, e00400 (2013).
53. Nunzi, M. G., Russo, M. & Mugnaini, E. Vesicular glutamate transporters VGLUT1 and VGLUT2 define two subsets of unipolar brush cells in organotypic cultures of mouse vestibulocerebellum. *Neuroscience* **122**, 359–371 (2003).
54. Hioki, H. *et al.* Differential distribution of vesicular glutamate transporters in the rat cerebellar cortex. *Neuroscience* **117**, 1–6 (2003).
55. Kirkby, P. A., Srinivas Nadella, K. M. N. & Silver, R. A. A compact Acousto-Optic Lens for 2D and 3D femtosecond based 2-photon microscopy. *Opt. Express* **18**, 13721–13745 (2010).
56. Fernández-Alfonso, T. *et al.* Monitoring synaptic and neuronal activity in 3D with synthetic and genetic indicators using a compact acousto-optic lens two-photon microscope. *J. Neurosci. Methods* **222**, 69–81 (2014).
57. Griffiths, V. A. *et al.* Real-time 3D movement correction for two-photon imaging in behaving

- animals. *Nat. Methods* **17**, 741–748 (2020).
58. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156–158 (2008).
 59. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
 60. Sofroniew, N. J., Cohen, J. D., Lee, A. K. & Svoboda, K. Natural whisker-guided behavior by head-fixed mice in tactile virtual reality. *J. Neurosci.* **34**, 9537–9550 (2014).
 61. Hill, D. N., Curtis, J. C., Moore, J. D. & Kleinfeld, D. Primary motor cortex reports efferent control of vibrissa motion on multiple timescales. *Neuron* **72**, 344–356 (2011).
 62. Jelitai, M., Puggioni, P., Ishikawa, T., Rinaldi, A. & Duguid, I. Dendritic excitation-inhibition balance shapes cerebellar output during motor behaviour. *Nat. Commun.* **7**, 13722 (2016).
 63. Pnevmatikakis, E. A. *et al.* Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron* **89**, 285–299 (2016).
 64. Zhou, P. *et al.* Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *Elife* **7**, (2018).
 65. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. doi:10.1101/061507.
 66. Björck, Åke & Golub, G. H. Numerical methods for computing angles between linear subspaces. *Math. Comput.* **27**, 579–594 (1973).
 67. Owen, A. B. & Perry, P. O. Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Ann. Appl. Stat.* **3**, 564–594 (2009).