Mechanisms of pattern decorrelation by recurrent neuronal circuits

Martin T Wiechert[1], Benjamin Judkewitz[2], Hermann Riecke[3] & Rainer W Friedrich[1]

Decorrelation is a fundamental computation that optimizes the format of neuronal activity patterns. Channel decorrelation by adaptive mechanisms results in efficient coding, whereas pattern decorrelation facilitates the readout and storage of information. Mechanisms achieving pattern decorrelation, however, remain unclear. We developed a theoretical framework that relates high-dimensional pattern decorrelation to neuronal and circuit properties in a mathematically stringent fashion. For a generic class of random neuronal networks, we proved that pattern decorrelation emerges from neuronal nonlinearities and is amplified by recurrent connectivity. This mechanism does not require adaptation of the network, is enhanced by sparse connectivity, depends on the baseline membrane potential and is robust. Connectivity measurements and computational modeling suggest that this mechanism is involved in pattern decorrelation in the zebrafish olfactory bulb. These results reveal a generic relationship between the structure and function of neuronal circuits that is probably relevant for pattern processing in various brain areas.

Sensory stimuli and neuronal activity often represent relevant information in a highly inefficient manner. Natural images, for example, may convey different messages even though their pixel distributions are nearly identical. A fundamental strategy to improve neuronal codes is decorrelation, which can reduce the redundancy between neuronal responses, distribute neuronal activity patterns more evenly over coding space and enable the brain to extract information from small differences between overlapping sensory inputs. Decorrelation is therefore an important computation for neuronal processing[1–3], as well as engineering and computer science.

Two forms of decorrelation have to be distinguished that perform different tasks and are referred to as channel and pattern decorrelation. Channel decorrelation decreases the overlap, and thus the redundancy, between response profiles of individual channels (neurons) to a set of stimuli, resulting in a code that is efficient because the information conveyed by different channels is largely complementary. Efficient coding has been implicated in various neuronal processing tasks[4,5]. For example, the receptive fields of neurons in primary visual cortex and the tuning of auditory nerve fibers can be described by filters that achieve channel decorrelation and result in efficient coding of natural images or sounds, respectively[6–9]. In non-neuronal systems, channel decorrelation methods such as independent component analysis have been used for blind source separation and other operations[10–13]. Most channel decorrelation methods must be adapted to a particular task on the basis of prior knowledge about inputs. As a consequence, adaptive methods can efficiently decorrelate channels with known properties, but may perform poorly on unexpected inputs.

In contrast, pattern decorrelation reduces the overlap between combinatorial patterns of activity across populations of neurons. Pattern decorrelation can therefore make neuronal representations more

distinct, which facilitates subsequent readout by a simple classifier. In addition, decorrelated representations are most likely important for robust memory storage by associative networks because correlated representations may be confused or erased by catastrophic interference[1,2,14,15]. Pattern decorrelation has been observed experimentally in various brain areas, including the olfactory bulb[16–18] and hippocampus[19], and may be involved in cerebellar motor learning[20]. However, the mechanisms underlying pattern decorrelation are not well understood.

It was previously found that the correlation between patterns or channels can be decreased by thresholding, a fundamental nonlinearity in neuronal transfer functions imposed by the mechanism of action potential generation[20–22]. To examine pattern decorrelation, we mathematically analyzed networks of randomly connected threshold-linear units. We first proved that, given normally distributed input patterns, thresholding invariably causes decorrelation. In feedforward circuits, however, this decorrelation is accompanied by sparsening of output activity and therefore requires large numbers of output units. We extended our analysis to recurrent networks and proved that random feedback connections amplify pattern decorrelation. Recurrent networks can therefore achieve pattern decorrelation with far fewer output units. Analysis of a simplified computational model indicated that this mechanism is likely to contribute strongly to pattern decorrelation in the olfactory bulb. These results reveal a generic, nonadaptive and biologically plausible pattern decorrelation mechanism, and provide stringent mathematical insight into the underlying principles.

RESULTS

Pattern decorrelation by SNOREs

To examine how networks of neurons can achieve pattern decorrelation, we mathematically analyzed a generic class of recurrent

[1]Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland. [2]Wolfson Institute for Biomedical Research, University College London, London, UK. [3]Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois, USA. Correspondence should be addressed to R.W.F. (rainer.friedrich@fmi.ch).
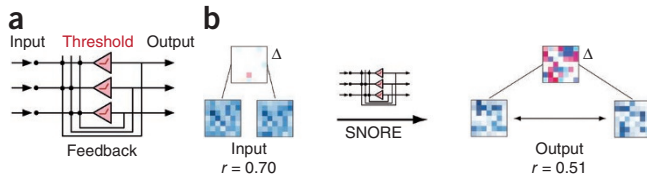
Figure 1 Pattern decorrelation by SNOREs. (**a**) Schematic illustration of a SNORE. Red line indicates thresholding (rectification). (**b**) Example of pattern decorrelation. Two random patterns across 10,000 units with a correlation of $r = 0.7$ were processed by a SNORE with sparse connectivity. Images show input and output of 49 units and their differences, Δ (blue, positive; red, negative). Pearson correlation coefficient, r, represents the similarity of the full input and output patterns.

networks that we refer to as stochastic networks of rectifying elements (SNOREs; see Online Methods). SNOREs consist of threshold-linear units that are randomly connected by synapses of uniform weight (**Fig. 1a**). The output (firing rate) of each unit is zero if its state variable (membrane potential) is below a given threshold and linearly increases when the threshold is exceeded. We consider input patterns with joint normal intensity distribution so that pairs of input patterns are binormally distributed. The overlap between input patterns is thus parameterized by the Pearson correlation coefficient, r_a. If the number of channels (neurons) is large, the overlap, r_x, between the corresponding steady-state response patterns depends only on r_a and on the structural parameters of the SNORE.

For simplicity, we only refer to purely inhibitory circuits; however, all of our results generalize to mixed excitatory-inhibitory SNOREs (see Online Methods). A SNORE is then characterized by six parameters: the mean $\mu(a)$ and s.d. $\sigma(a)$ of input patterns, the absolute firing threshold η, the time constant τ, the synaptic strength λ, and the number p of connections per neuron. The baseline membrane potential is included in $\mu(a)$. For a wide range of these parameters, the system converges to a steady state (**Supplementary Theory**). The first three parameters can be condensed into a single normalized threshold

$$\eta_a = \frac{\eta - \mu(a)}{\sigma(a)}$$ because the readout of interest, the correlation r_x

between steady-state response patterns, depends on η_a, but not on $\mu(a)$, $\sigma(a)$ and η individually. For example, changing the input mean $\mu(a)$ and the threshold η by a common offset will simply add the same offset to the response membrane potentials and will not affect the response firing rates. Thus, r_x only depends on the difference $\eta - \mu(a)$. Furthermore, r_x is independent of the time constant because τ has no influence on the steady state. The steady-state response correlation r_x is therefore a function of the three parameters: λ, p and η_a.

To explore how pattern decorrelation $r_a - r_x$ depends on these parameters, we first simulated SNOREs containing 10,000 units and found that some SNOREs produced substantial pattern decorrelation (**Fig. 1b**). We noticed that pattern decorrelation depended systematically on network parameters. Briefly, decorrelation increased with the inhibition strength λ, which is expected because stronger coupling should generally

enhance network effects. Notably, pattern decorrelation increased with connection sparseness $\frac{1}{p}$ when the total coupling strength $\Lambda = -\lambda p$

(number of synapses × synaptic weight) was kept constant. Thus, networks with sparse, but strong, connections decorrelated input patterns more effectively than networks with dense, but weak, connections, even though overall activity levels remained similar. Furthermore, for networks with sufficiently strong coupling (**Supplementary Theory**), pattern decorrelation increased with the baseline membrane potential of the units, which is included in $\mu(a)$. This is noteworthy because, in feedforward networks, correlations increase, rather than decrease, with increasing baseline membrane potential[21].

Mathematical analysis of pattern decorrelation by SNOREs

To understand the mechanism underlying pattern decorrelation by SNOREs and to explain its unexpected dependence on connection sparseness and baseline membrane potential, we mathematically analyzed the equations describing SNOREs (Online Methods). As nonlinearities are known to have well-defined effects on the Pearson correlation of a given distribution[23], we first determined how pattern decorrelation depends on the threshold in the input-output function of SNORE units. For correlated jointly normally distributed input, we mathematically proved that this nonlinearity always results in decorrelation and that decorrelation monotonically increases with the threshold level (Online Methods and **Fig. 2**; for a rigorous mathematical analysis of this threshold-induced decorrelation (TIDe), see **Supplementary Theory** and **Supplementary Fig. 1**). Thresholding of inputs approximates the passage through an array of neurons that do not interact with each other. Thus, a simple feedforward array of neurons acts as a pattern decorrelator, consistent with previous observations[20,21].

Increasing the threshold level can, in theory, result in arbitrarily strong decorrelation by TIDe (**Fig. 2b,c**). However, high thresholds will silence the majority of neurons so that a large number of neurons are required to generate meaningful output (**Fig. 3**). Thus, TIDe alone may not be sufficient to achieve substantial pattern decorrelation in circuits with limited numbers of neurons.
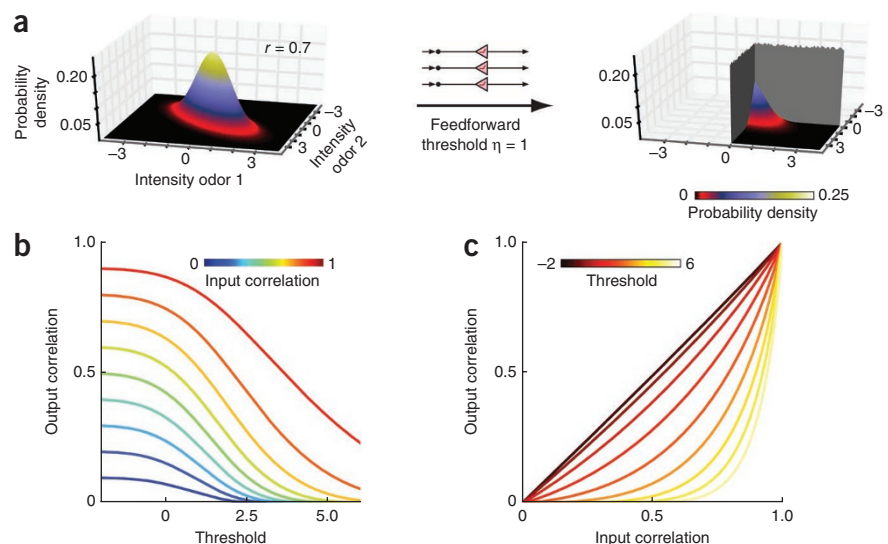
Figure 2 TIDe. (**a**) Standard binormal probability density with correlation $r = 0.7$ before and after applying threshold $\eta = 1$. Walls indicate peaks of Dirac distributions. (**b**) Output Pearson correlation as a function of threshold for different input Pearson correlations. (**c**) Output Pearson correlation as a function of input Pearson correlation for different thresholds.
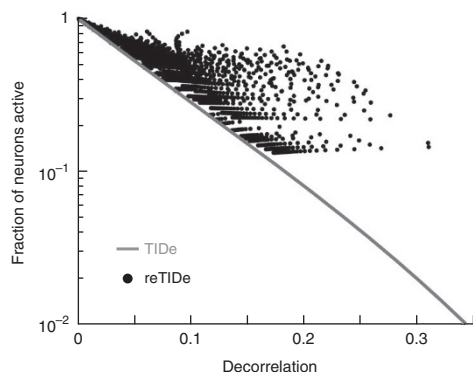
Figure 3 Sparsening of output activity by thresholding. The fraction of active neurons is plotted against the decorrelation achieved by TIDe (gray line) and reTIDe (dots). Different dots correspond to SNOREs with different parameter combinations (**Fig. 4e**). As decorrelation increases, the fraction of active neurons decreased systematically for TIDe, but not for reTIDe.

We next analyzed how TIDe is affected by recurrent connectivity. Recurrent connections feed the thresholded, and therefore decorrelated, output patterns back into the network where they become part of the input (**Fig. 1a**). Consequently, the Pearson correlation of the total input, that is, the sum of the external and the feedback inputs, is reduced. This in turn further decorrelates the output patterns until the steady state is reached. We proved that this effect increases the decorrelation produced by TIDe (Online Methods; for a rigorous mathematical analysis, see **Supplementary Theory**). In recurrent networks, TIDe therefore seeds a regenerative loop that progressively amplifies pattern decorrelation (**Supplementary Fig. 2**). This recurrence-enhanced TIDe (reTIDe) is more powerful than TIDe alone and can achieve substantial decorrelation with markedly fewer units (**Fig. 3**).

The formal proof of reTIDe is based on a mathematical theory (Online Methods) that predicts pattern decorrelation by a SNORE from its parameters. We validated this SNORE theory against simulations

and found that theoretical predictions were in excellent quantitative agreement with simulation results throughout a wide parameter range (**Fig. 4** and **Supplementary Fig. 3**). Small deviations were detected only for very sparse connectivity. This is expected because the assumption of normally distributed feedback, which depends on the central limit theorem, starts to break down in this regime (**Fig. 4e**).

SNORE theory can therefore be used to analytically derive relationships between network parameters and pattern decorrelation. Analysis of SNORE equations showed that reTIDe necessarily increases with connection sparseness $\frac{1}{p}$ when the total coupling strength Λ is kept constant (**Fig. 4a,e**) and that reTIDe increases with the baseline membrane potential when coupling is sufficiently strong (**Fig. 4c,e** and **Supplementary Fig. 3**). Thus, SNORE theory not only accurately predicts pattern decorrelation, but also analytically explains the relationships between pattern decorrelation and network parameters (Online Methods; **Supplementary Theory**).

Our analytical results lead to an intuitive understanding of the dependence of pattern decorrelation on connection sparseness (Online Methods; **Supplementary Fig. 2**). The effectiveness of the regenerative loop underlying reTIDe depends on the relative contributions of the external input pattern and the feedback pattern to the Pearson correlation of the total input pattern. Larger variance of the feedback pattern leads to a lower Pearson correlation of the total input and therefore enhances pattern decorrelation. In networks with dense and weak connections, the variance in the feedback pattern is low because the recurrent input to each unit is an average over many presynaptic neurons. In sparsely connected networks, however, the variance can become large because each neuron averages only over a small number of recurrent inputs, resulting in more pronounced pattern decorrelation (**Supplementary Fig. 2**; for a rigorous mathematical analysis, see **Supplementary Theory**).

The dependence of pattern decorrelation on baseline membrane potential can be understood by considering the interaction between baseline membrane potential and thresholding (Online Methods). Changes in baseline membrane potential have two opposing effects

Figure 4 reTIDe. (**a**) Difference in pattern decorrelation between recurrent networks (SNOREs) and feedforward networks (Δdecorrelation) as a function of connection density (fan-in, variable p in equations). The average over the ten most similar stimulus pairs is shown. The line represents the predictions of SNORE theory. Dots represent simulation results. Baseline membrane potential, $\mu(a)$, was 34.5 Hz (potentials and firing rates have the same units in the model). (**b**) Mean (black) and s.d. (gray; across units) of network activation patterns as a function of connection density. $\mu(a) = $ 34.5 Hz. (**c**) Δdecorrelation as a function of $\mu(a)$ (in Hz) (fan-in, $p = 12$). (**d**) Mean (black) and s.d. (gray) of network activity patterns as a function of $\mu(a)$ (in Hz) (fan-in, $p = 12$). (**e**) Δdecorrelation for a wide range of SNORE parameter combinations. The first three columns show prediction from SNORE theory, simulation results (10,000 randomly connected threshold-linear units) and their difference. The right column shows the deviation of the steady state across-population distribution of activation in simulation results from a normal distribution, quantified by the Cramer-von Mises statistic (high values indicate large deviation from normality).

Note that differences between predicted and observed Δdecorrelation are well explained by departure from normality. Gray areas depict parameter combinations for which the theory is not applicable because coupling is too strong ($\lambda \geq 1$). Hatched areas depict parameter regimes where convergence to a steady state does not occur in theoretical predictions or simulations. Note that the match between predicted and observed convergence is perfect.
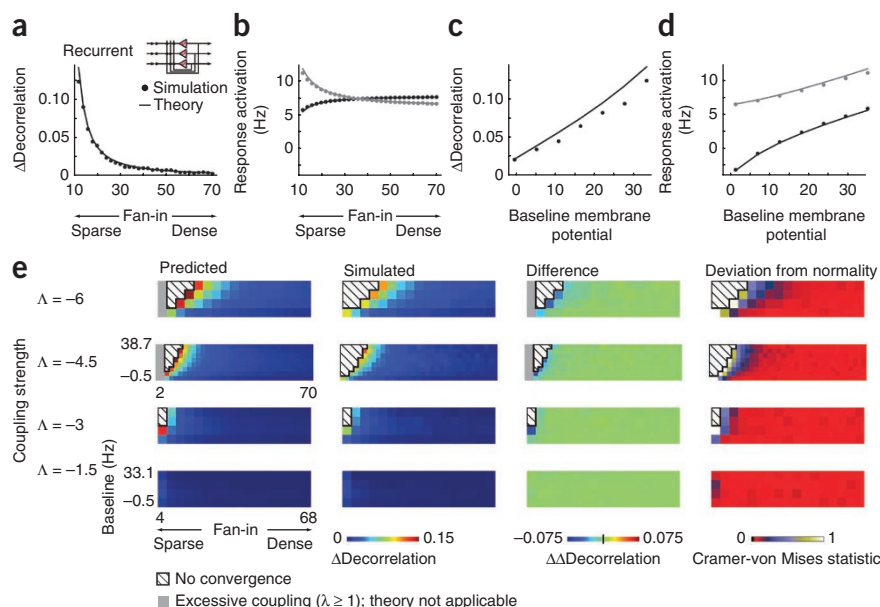
Figure 5 Functional connectivity in the zebrafish olfactory bulb measured by forward optical probing. (**a**) Left, mitral cell marker expression (HuC-YC) in an optical section through the lateral olfactory bulb. Shadow indicates patch pipette targeted onto a mitral cell (trigger neuron). Center, raw fluorescence of the calcium indicator (rhod-2) in the same view. Right, calcium signals evoked by current injection into the trigger neuron (average over 20 repetitions). Note the strong signals in the soma and dendrites of the trigger neuron (black arrowhead) and discrete responses of three followers in the interneuron layer (medium gray arrowheads). The light gray arrowhead indicates a nonresponsive region. (**b**) Top, example of the trigger neuron response to current injection. Bottom, time course of calcium signals in the soma of the trigger neuron (black traces) in the three followers (medium gray) and in the nonresponsive region (light gray). Shown is the average over 20 repetitions. (**c**) Scheme illustrating projection of follower neurons onto a sphere and measurement of the distance on the sphere (d_{sphere}). (**d**) Probability of finding a follower as a function of d_{sphere}. Dashed line shows exponential fit. Data represented by darker bars are based on more voxels and weighted higher in the fit. Inset, cumulative probability distributions for measured data (solid gray line) and exponential fit (dashed black line), both corrected for limited field of view.
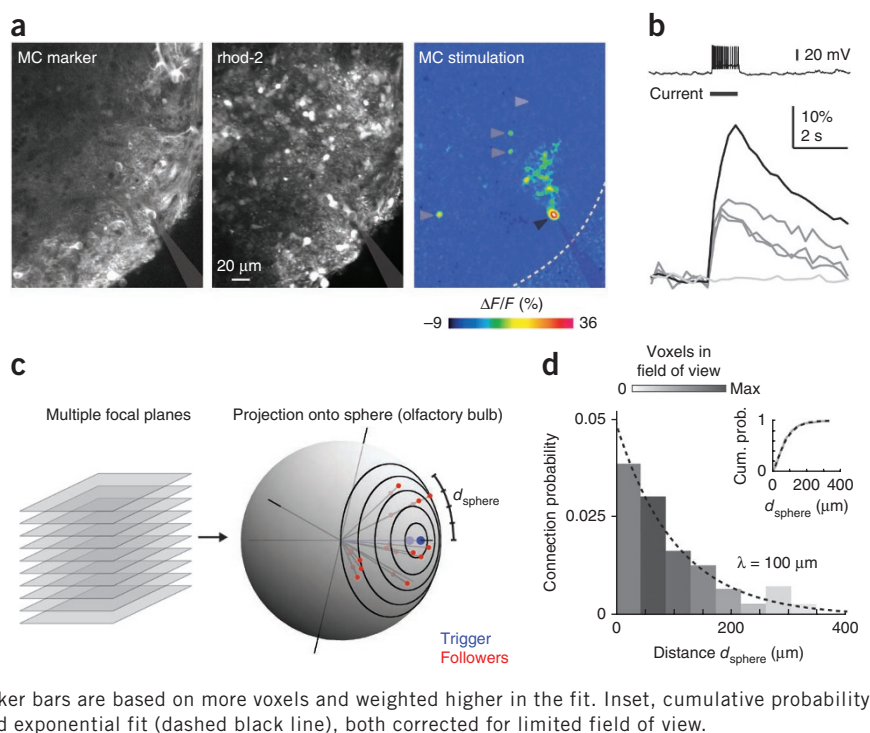
on reTIDe. First, when baseline membrane potential is increased, thresholding eliminates less of the variance in the input signal. As a consequence, reTIDe is enhanced because the variance in the feedback pattern is increased. Second, increasing baseline membrane potential reduces TIDe (Online Methods) and therefore results in less efficient seeding of reTIDe. The net effect of baseline membrane potential on reTIDe depends on the relative strengths of these opposing effects, which in turn depends on network parameters. Because the first effect scales approximately linearly with connection strength, whereas the second effect does not, reTIDe will grow with baseline membrane potential if the network is sufficiently coupled (**Supplementary Fig. 3**). This regime includes most networks with sparse and strong connections (for a rigorous mathematical treatment, see **Supplementary Theory**).

Decorrelation could also be achieved by a chaotic system, but such a mechanism would be of little biological use because even minimal input corruption would result in an unpredictable change in the output. In contrast, TIDe and reTIDe are well conditioned; that is, the effect small input changes have on output is limited (Online Methods; **Supplementary Theory**). The amount of imprecision in the input that can be tolerated depends on network parameters and there is an obvious trade-off between robustness and decorrelation. This kind of robustness facilitates pattern classification because sets of moderately similar patterns are decorrelated, whereas correlations between near-identical patterns that may convey the same message are largely preserved (**Supplementary Fig. 4**).

Outputs could also be compromised by fluctuations in the units and connections of a SNORE. However, the steady states of SNOREs are linearly attractive. Departure from the fixed point will be corrected for, rather than amplified by, network dynamics. In principle, multiple fixed points might exist and fluctuations during the transient phase of the response might affect which attractor the system converges to. However, Monte-Carlo simulations indicate that there is only one globally attractive fixed point (Online Methods), consistent with the theoretical prediction that identical inputs will result in an output correlation of 1. TIDe and reTIDe are therefore numerically stable, implying that they can be implemented with noisy units and connections.

Because SNORE theory applies not only to inhibitory networks, but also to excitatory or mixed networks, we validated theoretical predictions of SNORE theory in this regime. Again, we found that theoretical predictions were in excellent quantitative agreement with simulation results (**Supplementary Fig. 5**).

SNORE theory indicates that the nonlinearity in neuronal input-output functions is essential for TIDe and reTIDe. Indeed, we found that any nonzero linear map will, on average, leave the angle between two vectors unchanged (**Supplementary Theory**). In particular, a linear adaptive method can be trained to effectively decorrelate a limited set of input patterns, but will, on average, perform poorly on unexpected inputs. Hence, TIDe and reTIDe are superior to any linear mechanism when input patterns are unpredictable.

Pattern decorrelation by a model of the olfactory bulb

To explore whether reTIDe can also explain pattern decorrelation by a biologically plausible network that lacks some of the idealizations of SNOREs, we examined a computational model of the olfactory bulb, the first olfactory processing center in the brain. Structurally similar odors evoke overlapping patterns of distributed activity across the input channels of the olfactory bulb, the glomeruli[24–26]. These input patterns are transformed into decorrelated activity patterns across the output neurons, the mitral cells, as has been seen in zebrafish[16–18]. Local GABAergic interneurons mediate inhibition between mitral cells via multiple synaptic pathways, presumably in an action potential–dependent manner[27]. We first characterized the functional connectivity between mitral cells and interneurons in the zebrafish olfactory bulb by 'forward optical probing'. In this approach, vigorous action potential firing is evoked in an individual mitral cell (trigger) by whole-cell current injection, whereas neuronal activity in the surrounding tissue is monitored by multiphoton calcium imaging to identify functionally connected 'follower' neurons.

Individual mitral cells were identified by a transgenic marker[28,29] and stimulated to fire action potentials at a rate of 30–50 Hz for 1 s (20 repetitions; **Fig. 5a,b**). This stimulus evoked calcium signals
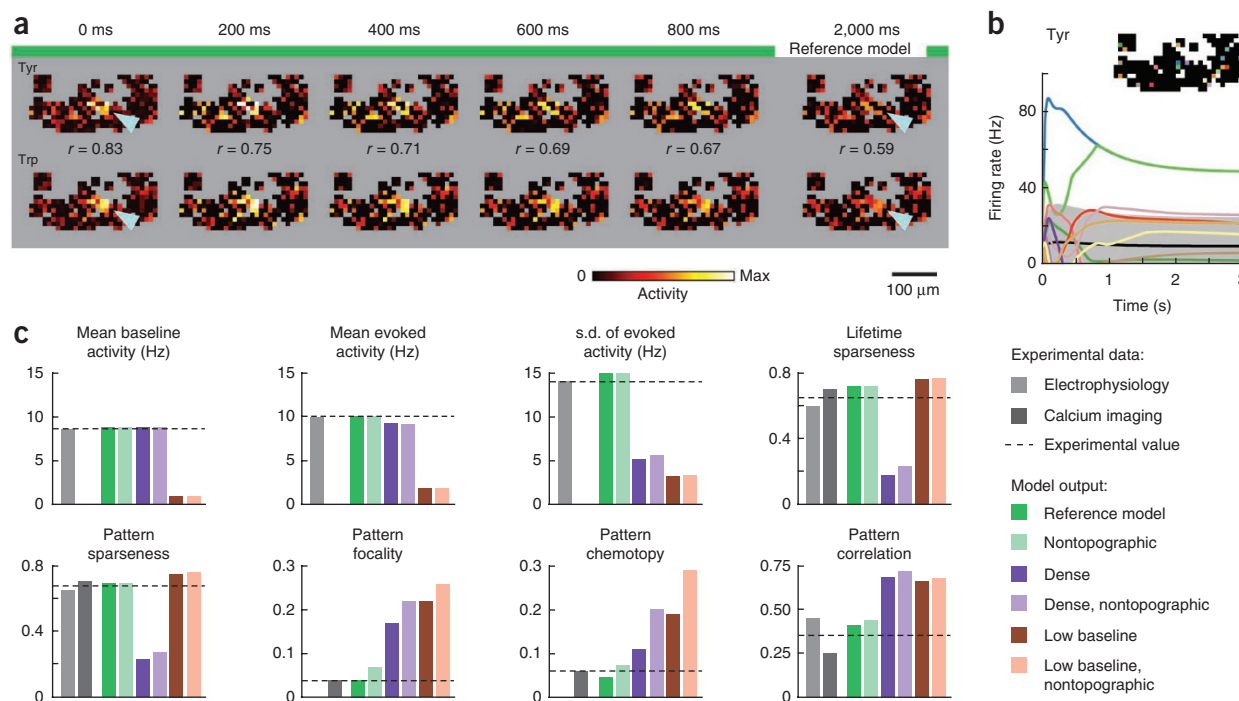
Figure 6 Output of a computational model of the olfactory bulb. (**a**) Mean output activity patterns evoked by tyrosine (tyr) and tryptophan (trp) (**Supplementary Fig. 8**) in successive 200-ms time windows. Note that foci of active mitral cells (arrowheads) became less pronounced over time and Pearson correlation coefficients decreased. (**b**) Response time courses of 20 mitral cells to stimulation with Tyr pattern. Mitral cell positions are indicated by the color code in the inset. Three mitral cells were chosen from the central cluster (arrowhead in **a**); the remaining 17 mitral cells were selected randomly. (**c**) Comparison of model output to experimental data determined by electrophysiology[16,17] and/or calcium imaging[32] for eight readouts that characterize single-neuron and population activity. Dashed line shows the experimentally observed value (mean value if the same readout was measured by both methods). Model output was quantified for the reference model (green) and for models with dense connectivity (blue), low baseline membrane potential (red) and nontopographic connectivity (light colors). All values except baseline firing rate were determined in the steady state. For additional information, see **Supplementary Table 1**.

in the soma and dendrite of the trigger neuron, as well as in sparse sets of follower neurons (**Fig. 5a**). Because mitral cells make synaptic connections predominantly onto inhibitory interneurons, neurons more than one synapse downstream of the trigger mitral cell are unlikely to be excited by the stimulus. Indeed, the majority of follower neurons (260 of 262, $n = 32$ trigger neurons in 20 olfactory bulbs) were interneurons. The remaining two followers expressed the mitral cell marker and projected dendrites to the same glomerulus as the trigger mitral cell. These follower mitral cells were therefore probably connected to the trigger neuron by gap junctions or intraglomerular glutamatergic connections[30,31] and were excluded from further analysis. To facilitate distance measurements in a metric that relates to inter-glomerular distances, we radially projected the three-dimensional coordinates of followers onto a sphere representing the surface of the olfactory bulb. The probability of finding a follower (connection probability) was then determined as a function of surface distance from the trigger (equivalent to their angular separation; **Fig. 5c**).

Connection probability decreased with distance in a fashion that was well fit by an exponential model (**Fig. 5d**) with a length constant (~100 μm) and maximal distance (~400 μm) that were substantially larger than the diameter of a glomerulus in the recorded region (~20 μm)[24]. In contrast, the amplitude of the follower response decreased only slightly with distance (**Supplementary Fig. 6**), implying that the distance-dependent decay of connection probability does not reflect a decrease in coupling strength. To estimate the absolute connection probability, we extrapolated the mean number of follower

neurons per mitral cell (68 ± 20, mean ± s.d.; Online Methods) and estimated the total number of interneurons in the olfactory bulb (~20,000; **Supplementary Fig. 7**). Each mitral cell is thus estimated to contact ~0.34% of interneurons. Because the olfactory bulb contains ~1,500 mitral cells[32], each interneuron is estimated to receive approximately five mitral cell inputs that are detectable by forward optical probing. These results indicate that strong synaptic connections between mitral cells and interneurons are sparse.

The computer model consisted of a single layer of threshold-linear analog neurons (mitral cells) equivalent to the units in SNOREs. Mitral cells made bidirectional inhibitory synapses of uniform weight with other mitral cells, reflecting the reciprocal connectivity in the olfactory bulb (**Supplementary Fig. 8**). Connection probability decayed exponentially with distance (**Supplementary Fig. 8**) with a length constant of 100 μm. The model time constant was set to $\tau = 20$ ms. The model was stimulated using previously recorded glomerular afferent responses to 16 different amino acids[24] (Online Methods; **Supplementary Fig. 8**) with a stereotyped time course derived from odor responses of sensory neurons[16]. The spatial organization (chemotopy) of measured glomerular activation patterns in the olfactory bulb was therefore directly imposed onto the model. Mitral cells were arranged in a square grid and assigned to the nearest glomerulus or pruned if there was no detected glomerulus in a given radius. The resulting model contained a geometric arrangement of 239 mitral cells that reflected the experimentally observed pattern of glomeruli. This model differed from generic SNOREs because connections were symmetrical, input patterns were not normally distributed, and
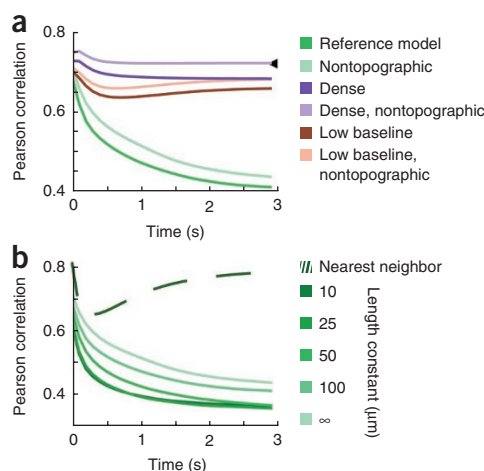
Figure 7 Dependence of pattern decorrelation by the olfactory bulb model on topography and network parameters. (**a**) Mean correlation coefficients of output activity patterns evoked by the ten most similar stimulus pairs as a function of time for different networks. Note that pattern decorrelation by the reference model (dark green) was not substantially affected by randomizing the spatial distribution of connections (light green). However, pattern decorrelation was severely reduced when connectivity was dense (purple) or baseline membrane potential was low (red). (**b**) Varying the length constant of connection probability in the sparse model showed that pattern decorrelation increased as the spatial extent of connectivity was decreased from random (infinity, light green) to narrow-range connectivity (10 µm, dark green). However, the effect was small compared with changes in connection density or baseline membrane potential (**a**). Nearest-neighbor connectivity (dashed) did not result in effective pattern decorrelation.

inputs and connections were topographically organized. Moreover, the number of mitral cells after pruning (239) was low, but biologically realistic[16,32].

The model responded to the 16 input patterns with odor- and mitral cell–specific spatio-temporal activity patterns (**Fig. 6a,b** and **Supplementary Figs. 9** and **10**). We adjusted network parameters (coupling strength, connection sparseness, firing threshold, baseline membrane potential and input strength) to match the model output to experimental data at the single-neuron and population level determined previously by electrophysiology[16,17] and/or two-photon calcium imaging[32]. The following readouts were compared quantitatively between model and experiment: mean baseline firing rate, mean odor-evoked firing rate, s.d. of odor-evoked firing rates across mitral cells, lifetime sparseness of response profiles, pattern sparseness across the mitral cell population, focality of response patterns[32], chemotopy of response patterns[32] and the mean pattern correlation evoked by the ten most similar stimulus pairs. In addition, we qualitatively compared the response profiles of model mitral cells and their dynamics (**Supplementary Fig. 9**) to electrophysiological data[16].

Good agreement between model output and experimental observations was obtained with relatively strong synaptic coupling, sparse connectivity and relatively high mitral cell baseline membrane potential, consistent with experimental data from zebrafish and other vertebrates[17,33–36]. Deviations of model data from experimentally determined values were less than 25% for all variables and were less than 10% in most cases (**Fig. 6c** and **Supplementary Table 1**). This is not trivial considering the simplifications contained in the model and the small number of parameters. Moreover, the model reproduced many of the dynamical and topological features of odor response patterns in the zebrafish olfactory bulb[16,17,32] (**Supplementary Figs. 9–12**) and produced a gradual pattern decorrelation (**Fig. 7a** and **Supplementary Fig. 11**) similar to that observed experimentally[16–18].

In vertebrates, glomeruli responding to some molecular features are preferentially located in loosely defined regions, although not necessarily clustered[24–26]. This weakly chemotopic organization raises the possibility that decorrelation results from distance-dependent lateral inhibition between glomeruli[37], similar to edge enhancement in the retina[38]. To differentiate between such a topographic mechanism and reTIDe, we tested how pattern decorrelation was affected when model parameters were modified. Pattern decorrelation, as well as other measurements, was only marginally affected when connections

or glomerular positions were redistributed randomly (**Figs. 6c** and **7a** and **Supplementary Figs. 9–14**). Moreover, decorrelation depended only weakly on the length constant of connection probability and was markedly reduced by nearest-neighbor connectivity (**Fig. 7b** and **Supplementary Fig. 15**). Thus, a topographic mechanism is not required for pattern decorrelation.

In contrast, pattern decorrelation strongly depended on connection sparseness and baseline membrane potential of mitral cells (**Fig. 7a** and **Supplementary Figs. 10** and **11**). Dense connectivity essentially abolished pattern decorrelation (**Fig. 7a** and **Supplementary Figs. 10**, **11** and **13–15**). Similarly, pattern decorrelation was substantially reduced when the baseline membrane potential of mitral cells was decreased (**Fig. 7a** and **Supplementary Figs. 10**, **11** and **13–15**). These results are fully consistent with the predictions from SNORE theory and provide strong evidence that reTIDe is the primary mechanism underlying pattern decorrelation in our olfactory bulb model.

DISCUSSION

We describe a generic pattern decorrelation mechanism, reTIDe, that is a natural consequence of sparse and stochastic recurrent connectivity among nonlinear neuronal units. Simulations indicate that reTIDe is a plausible mechanism involved in pattern decorrelation in the olfactory bulb. These results reveal basic relationships between the structure and function of neuronal circuits that are likely to be relevant for the processing of activity patterns in various brain areas.

Pattern decorrelation by reTIDe

Pattern decorrelation by reTIDe comprises two crucial steps. First, an initial decorrelation of input patterns is caused by thresholding (TIDe), which is a necessary consequence of nonlinear neuronal input-output functions. Unless the number of neurons is very high, this thresholding step alone cannot produce strong decorrelation, but acts as a seed for subsequent amplification. Second, TIDe is amplified by feeding output patterns back into the network via recurrent connections (reTIDe). This amplification is particularly strong when connectivity is sparse because the variance in the feedback pattern, and therefore the contribution of the thresholded feedback pattern to the total input, is high (**Supplementary Fig. 2**). Moreover, the effectiveness of reTIDe depends on the baseline membrane potential, which is also a consequence of thresholding and subsequent amplification. reTIDe is therefore a generic decorrelation mechanism that emerges from basic properties of recurrent circuits.

reTIDe differs from other decorrelation strategies in at least three respects. First, although most other decorrelation mechanisms perform channel decorrelation, reTIDe performs pattern decorrelation.

reTIDe therefore reduces the overlap between neuronal population representations and facilitates the readout and storage of activity patterns.

Second, most decorrelation mechanisms must be adapted to their inputs on the basis of prior knowledge. For example, source separation by independent component analysis requires training of a network on representative input data[10–13] and the filter properties of auditory or visual neurons have likely been optimized for processing typical sensory input by evolution and experience[6,7,9]. In contrast, pattern decorrelation by reTIDe is universal and does not depend on prior adaptation of the network. As a consequence, reTIDe will decorrelate a wide range of patterns that do not need to have a specific structure. This cannot be achieved by any linear method, including adaptive strategies such as principal component–based approaches (**Supplementary Theory**). reTIDe is therefore particularly useful when inputs are unpredictable or when input patterns lack specific statistical properties.

Third, for many decorrelation mechanisms, it is unclear how they may be implemented in neuronal circuits. Adaptive decorrelation strategies, for example, often rely on a global cost function that does not reflect biologically plausible learning rules. In contrast, reTIDe occurs in generic neuronal networks (SNOREs) and relies only on thresholding and recurrent connectivity, which are among the most basic properties of neurons and circuits in the brain. The possible biological implementation of reTIDe is therefore obvious.

Pattern decorrelation in the olfactory bulb

Forward optical probing revealed two basic features of neuronal connectivity in the zebrafish olfactory bulb. First, the weak decay of connection probability with distance implies that neuronal interactions are widespread, but coarsely topographically organized. A field with a radius equivalent to the estimated length constant encompasses about 75 glomeruli in the lateral olfactory bulb of zebrafish, allowing for potential interactions among neurons associated with diverse sets of glomeruli. Second, connections between mitral cells and follower interneurons are sparse. This result may be biased toward strong connections because forward optical probing may fail to detect weak or silent synapses. Nevertheless, sparse connectivity in the olfactory bulb has also been inferred in rodents by trans-synaptic viral tracing[34] and from the low probability of finding connected mitral-granule cell pairs[39]. Moreover, sparse functional interactions between mitral cells and glomeruli have been shown *in vivo*[33], and sparse connectivity resulted in a good match between experimental data and our olfactory bulb model.

Our modeling results indicate that pattern decorrelation in the olfactory bulb does not depend on topographic mechanisms. Instead, efficient pattern decorrelation by our olfactory bulb model depended on sparse connectivity and high baseline membrane potential, two conserved and salient features of the olfactory bulb whose functions have been elusive. The behavior of the olfactory bulb model is therefore consistent with strong predictions of SNORE theory and the reTIDe mechanism. In principle, pattern decorrelation could also be achieved by other nontopographic mechanisms[40,41] that result in a form of histogram modification[42]. One prediction of these mechanisms is that weak mitral cell activity should be systematically suppressed, whereas strong activity should be systematically preserved. Neuronal responses in the olfactory bulb and in the insect antennal lobe, however, do not generally follow this prediction[17,43]. Thus, experimental results are difficult to reconcile with these mechanisms but are consistent with reTIDe. reTIDe is therefore a plausible mechanism for pattern decorrelation in the olfactory bulb, although other mechanisms may also be involved.

Implementations and implications of TIDe and reTIDe

Classical work has proposed that thresholding of convergent inputs produces a sparse and decorrelated code in the cerebellum[20] and that similar mechanisms create non-overlapping and sparse odor representations in the mushroom body of insects[22,44]. These processes may represent examples of decorrelation by TIDe in circuits with large numbers of neurons. In hippocampus, pattern decorrelation occurs in the dentate gyrus[19] and its target area CA3 (refs. 19,45). Both of these areas contain intra- and inter-areal recurrent connections, and connections between some neuron types are sparse[14,46]. It may therefore be interesting to explore whether pattern decorrelation in hippocampal areas involves reTIDe. In general, sparse recurrent connectivity is also abundant in neocortex[47]. Thus, pattern decorrelation by reTIDe may occur not only in the olfactory bulb but also in other brain areas.

Given that reTIDe decorrelates a wide range of patterns without the need for adaptation, decorrelation by reTIDe could generally facilitate pattern discrimination and enhance the coding capacity of downstream circuits[1,3]. Moreover, reTIDe could reduce correlations among inputs to associative memory circuits to prevent misclassification and catastrophic interference[1,2]. These functions might, for example, be important in the olfactory bulb and dentate gyrus because their target areas, the piriform cortex and area CA3, respectively, are thought to function as associative networks for pattern storage[1,14,15].

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

AUTHOR CONTRIBUTIONS
M.T.W. performed all of the mathematical analyses and computational modeling and wrote the manuscript. B.J. performed physiological experiments and was involved in early modeling work. H.R. was involved in early stages of theoretical work and wrote the manuscript. R.W.F. participated in modeling and data analysis and wrote the manuscript.

1. Srivastava, V., Parker, D.J. & Edwards, S.F. The nervous system might 'orthogonalize' to discriminate. *J. Theor. Biol.* **253**, 514–517 (2008).
2. French, R.M. Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* **3**, 128–135 (1999).
3. Itskov, V. & Abbott, L.F. Pattern capacity of a perceptron for sparse discrimination. *Phys. Rev. Lett.* **101**, 018101 (2008).
4. Barlow, H.B. Possible principles underlying the transformations of sensory messages. in *Sensory Communication* (ed. W.A. Rosenblith) 217–234 (MIT Press, Cambridge, Massachusetts, 1961).
5. Barlow, H. Redundancy reduction revisited. *Network* **12**, 241–253 (2001).
6. Lewicki, M.S. Efficient coding of natural sounds. *Nat. Neurosci.* **5**, 356–363 (2002).
7. Olshausen, B.A. & Field, D.J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).

8. Atick, J.J. & Redlich, A.N. Convergent algorithm for sensory receptive-field development. *Neural Comput.* **5**, 45–60 (1993).
9. Vinje, W.E. & Gallant, J.L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
10. Weinstein, E., Feder, M. & Oppenheim, A.V. Multi-channel signal separation by decorrelation. *IEEE Trans. Speech Audio Process.* **1**, 405–413 (1993).
11. Amari, S., Cichocki, A. & Yang, H.H. Recurrent neural networks for blind separation of sources. in *International Symposium on Nonlinear Theory and its Applications* 37–42 (NOLTA'95, Las Vegas, 1995).
12. Brown, G.D., Yamada, S. & Sejnowski, T.J. Independent component analysis at the neural cocktail party. *Trends Neurosci.* **24**, 54–63 (2001).
13. Parra, L.C. & Spence, C.D. Separation of non-stationary natural signals. in *Independent Component Analysis: Principles and Practice* (eds. S. Roberts & R. Everson) 135–157 (Cambridge University Press, Cambridge, Massachusetts, 2001).
14. Rolls, E.T. & Kesner, R.P. A computational theory of hippocampal function, and empirical tests of the theory. *Prog. Neurobiol.* **79**, 1–48 (2006).
15. Hasselmo, M.E., Wilson, M.A., Anderson, B.P. & Bower, J.M. Associative memory function in piriform (olfactory) cortex: computational modeling and neuropharmacology. *Cold Spring Harb. Symp. Quant. Biol.* **55**, 599–610 (1990).
16. Friedrich, R.W. & Laurent, G. Dynamic optimization of odor representations in the olfactory bulb by slow temporal patterning of mitral cell activity. *Science* **291**, 889–894 (2001).
17. Friedrich, R.W. & Laurent, G. Dynamics of olfactory bulb input and output activity during odor stimulation in zebrafish. *J. Neurophysiol.* **91**, 2658–2669 (2004).
18. Friedrich, R.W., Habermann, C.J. & Laurent, G. Multiplexing using synchrony in the zebrafish olfactory bulb. *Nat. Neurosci.* **7**, 862–871 (2004).
19. Leutgeb, J.K., Leutgeb, S., Moser, M.B. & Moser, E.I. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007).
20. Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202**, 437–470 (1969).
21. de la Rocha, J., Doiron, B., Shea-Brown, E., Josic, K. & Reyes, A. Correlation between neural spike trains increases with firing rate. *Nature* **448**, 802–806 (2007).
22. Jortner, R.A., Farivar, S.S. & Laurent, G. A simple connectivity scheme for sparse coding in an olfactory system. *J. Neurosci.* **27**, 1659–1669 (2007).
23. Baum, R. The correlation function of Gaussian noise passed through nonlinear devices. *IEEE Trans. Inf. Theory* **15**, 448–456 (1969).
24. Friedrich, R.W. & Korsching, S.I. Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging. *Neuron* **18**, 737–752 (1997).
25. Soucy, E.R., Albeanu, D.F., Fantana, A.L., Murthy, V.N. & Meister, M. Precision and diversity in an odor map on the olfactory bulb. *Nat. Neurosci.* **12**, 210–220 (2009).
26. Meister, M. & Bonhoeffer, T. Tuning and topography in an odor map on the rat olfactory bulb. *J. Neurosci.* **21**, 1351–1360 (2001).
27. Egger, V., Svoboda, K. & Mainen, Z.F. Dendrodendritic synaptic signals in olfactory bulb granule cells: local spine boost and global low-threshold spike. *J. Neurosci.* **25**, 3521–3530 (2005).
28. Yaksi, E. & Friedrich, R.W. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^{2+} imaging. *Nat. Methods* **3**, 377–383 (2006).
29. Li, J. *et al.* Early development of functional spatial maps in the zebrafish olfactory bulb. *J. Neurosci.* **25**, 5784–5795 (2005).
30. Urban, N.N. & Sakmann, B. Reciprocal intraglomerular excitation and intra- and interglomerular lateral inhibition between mouse olfactory bulb mitral cells. *J. Physiol. (Lond.)* **542**, 355–367 (2002).
31. Christie, J.M. *et al.* Connexin36 mediates spike synchrony in olfactory bulb glomeruli. *Neuron* **46**, 761–772 (2005).
32. Yaksi, E., Judkewitz, B. & Friedrich, R.W. Topological reorganization of odor representations in the olfactory bulb. *PLoS Biol.* **5**, e178 (2007).
33. Fantana, A.L., Soucy, E.R. & Meister, M. Rat olfactory bulb mitral cells receive sparse glomerular inputs. *Neuron* **59**, 802–814 (2008).
34. Willhite, D.C. *et al.* Viral tracing identifies distributed columnar organization in the olfactory bulb. *Proc. Natl. Acad. Sci. USA* **103**, 12592–12597 (2006).
35. Kapoor, V. & Urban, N.N. Glomerulus-specific, long-latency activity in the olfactory bulb granule cell network. *J. Neurosci.* **26**, 11709–11719 (2006).
36. Rinberg, D., Koulakov, A. & Gelperin, A. Sparse odor coding in awake behaving mice. *J. Neurosci.* **26**, 8857–8865 (2006).
37. Yokoi, M., Mori, K. & Nakanishi, S. Refinement of odor molecule tuning by dendrodendritic synaptic inhibition in the olfactory bulb. *Proc. Natl. Acad. Sci. USA* **92**, 3371–3375 (1995).
38. Hartline, H.K. & Ratliff, F. Inhibitory interaction of receptor units in the eye of *Limulus. J. Gen. Physiol.* **40**, 357–376 (1957).
39. Isaacson, J.S. Mechanisms governing dendritic gamma-aminobutyric acid (GABA) release in the rat olfactory bulb. *Proc. Natl. Acad. Sci. USA* **98**, 337–342 (2001).
40. Cleland, T.A. & Sethupathy, P. Non-topographical contrast enhancement in the olfactory bulb. *BMC Neurosci.* **7**, 7 (2006).
41. Arevian, A.C., Kapoor, V. & Urban, N.N. Activity-dependent gating of lateral inhibition in the mouse olfactory bulb. *Nat. Neurosci.* **11**, 80–87 (2008).
42. Hummel, R.A. Histogram modification techniques. *Comp. Graph. Image Proc.* **4**, 209–224 (1975).
43. Wilson, R.I., Turner, G.C. & Laurent, G. Transformation of olfactory representations in the *Drosophila* antennal lobe. *Science* **303**, 366–370 (2004).
44. Perez-Orive, J. *et al.* Oscillations and sparsening of odor representations in the mushroom body. *Science* **297**, 359–365 (2002).
45. Wills, T.J., Lever, C., Cacucci, F., Burgess, N. & O'Keefe, J. Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005).
46. Lisman, J.E. Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions. *Neuron* **22**, 233–242 (1999).
47. Braitenberg, V. & Schüz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer, Berlin, 1998).

ONLINE METHODS

Theoretical framework. Our theory applies to SNOREs in the steady state. Neurons are modeled as threshold-linear units, that is, the state of cell j under stimulus α at time t is fully described by its activation, $x_{j,\alpha}(t)$. We often use the suggestive term membrane potential even though this term is not fully accurate in the supra-threshold regime. The instantaneous firing rate of cell j equals $[x_{j,\alpha}(t) - \eta_0]_+$, where η_0 and subscript '+' denote the threshold and half-wave rectification, respectively. Given the vector a_α of afferent firing rates, the vector x_α of neuronal activation and the connectivity matrix L the equations of motion take the form

$$\tau \dot{x}_\alpha(t) = -x_\alpha(t) + a_\alpha(t) + L\left[x_\alpha(t) - \eta_0\right]_+ \quad (1)$$

where the dot indicates the time derivative. Given a stationary stimulus α, it is essential to know whether equation (1) will reach a stable steady state

$$x_\alpha = a_\alpha + L\left[x_\alpha - \eta_0\right]_+ \quad (2)$$

For symmetrical L, a convergence criterion based on Lyapunov functions has been given[48]. We derived a heuristic criterion (gain and variance limitedness) that predicts convergence to a steady state with high accuracy for random L and a_α (**Supplementary Theory** and **Supplementary Fig. 3**).

The idea of SNORE theory is to adopt a statistical view of equation (2) and replace the population vectors with across-population distributions. We make the simplifying assumption that each cell receives p recurrent contacts of strength $0 < \lambda < 1$ ($\lambda \geq 1$ leads to group-winner-take-all behavior if L is symmetric and inhibitory[48]) of which p_+ are excitatory and p_- are inhibitory. We also allow for global feedback of strength γ such that the total synaptic weight on each cell is $\Lambda = (p_+ - p_-)\lambda + \gamma$. In purely inhibitory networks such as the olfactory bulb model, p_+ and γ vanish. If connectivity is sparse ($p^2 \ll$ network size) and random, equation (2) leads to

$$\mu(x_\alpha) = \mu(a_\alpha) + \Lambda\mu\left(\left[x_\alpha - \eta_0\right]_+\right) \quad (3)$$

$$\text{var}(x_\alpha) = \text{var}(a_\alpha) + P\,\text{var}\left(\left[x_\alpha - \eta_0\right]_+\right) \quad (4)$$

$$\text{cov}(x_\alpha, x_\beta) = \text{cov}(a_\alpha, a_\beta) + P\,\text{cov}\left(\left[x_\alpha - \eta_0\right]_+, \left[x_\beta - \eta_0\right]_+\right) \quad (5)$$

between means, variances and covariances, respectively, each taken across the population of cells or input channels. We abbreviated $P = p\lambda^2$. Equation (5) compares two stimuli α and β.

If the input $a = (a_\alpha, a_\beta)$ is bivariate normally distributed and $p \gg 1$, then, by the central limit theorem, $x = (x_\alpha, x_\beta)$ is approximately binormally distributed and completely determined by equations (3–5). Solving equations (3–5) thus requires an understanding of the effect of thresholding on the binormal distribution and its mean, variance and covariance. Using the properties of these moments, solutions to equations (3–5) can be tightly characterized. In particular, the dependence of pattern separation on model parameters is determined by analytical results (**Supplementary Theory**).

Theorems. All proofs are given in the **Supplementary Theory**. Let $_{\eta,\eta}\Phi_\rho$ be the distribution obtained from the standard binormal distribution with correlation ρ by applying the threshold η to both variates.

Theorem 1 (TIDe theorem): assume $\rho \neq \pm 1$. The Pearson correlation $_\eta r(\rho)$ of $_{\eta,\eta}\Phi_\rho$ is infinitely differentiable in ρ and η,
1. a) strictly increasing in ρ,
   b) strictly convex in ρ,
2. strictly increasing in η if $\rho < 0$ and strictly decreasing in η if $\rho > 0$ and
3. tends to zero for $\eta \to \infty$.

The most important part of this theorem is assertion 2, which states that thresholding reduces pattern overlap.

Returning to equations (3–5), let ρ_a and ρ_x be the correlation coefficients of $a = (a_\alpha, a_\beta)$ and $x = (x_\alpha, x_\beta)$, respectively, and let the stimuli α and β be of the same mean $\mu(a) = \mu(a_\alpha) = \mu(a_\beta)$ and variance $\text{var}(a) = \text{var}(a_\alpha) = \text{var}(a_\beta)$. Then, we have the following corollary of theorem 1:

Theorem 2 (reTIDe theorem): assuming binormality of input and feedback, $0 < \rho_a < 1$ implies $\rho_x < \rho_a$.

The theorem states that without accounting for the final thresholding step, $\rho_x \mapsto {}_\eta r(\rho_x)$ Pearson correlation is reduced. It thus isolates the gain in pattern separation of reTIDe over TIDe. This also applies to theorem 3.

Theorem 3: assume $0 < \rho_a < 1$, binormality of input and feedback, and gain and variance limitedness (**Supplementary Theory**).
1. ρ_x is strictly decreasing in P and
2. strictly convex in ρ_a
If, in addition, the system is sufficiently coupled (**Supplementary Theory**)
3. ρ_x is strictly decreasing in $\mu(a)$.

Because parameters that are either not gain or variance limited will typically not permit a steady state (**Supplementary Theory** and **Supplementary Fig. 16**), the premises of theorem 3 are mild.

Making connectivity sparser while keeping Λ, γ and the excitation-inhibition ratio constant leads to larger P. Assertion 1 of theorem 3 therefore states that connection sparseness promotes pattern decorrelation. In the olfactory bulb, sufficient coupling is likely to be fulfilled because connectivity is sparse and strong. Assertion 3 therefore states that higher mitral cell baseline activation promotes pattern decorrelation in olfactory bulb–like networks.

Forward optical probing of neuronal connectivity. Zebrafish (crosses of wild-type strains Ab/Ab, Ab/TÜ or TÜ/TÜ and transgenic strain HuC:YC) were bred and kept at 26–27 °C on a 14-h/10-h light/dark cycle. HuC:YC transgenic fish express the fluorescent protein yellow cameleon 2.1 in mitral cells of the adult olfactory bulb[29]. Experiments were performed in an explant preparation of the nose and brain from adult zebrafish as described[16]. All animal procedures were performed in accordance with official animal care guidelines and approved by the Federal Republic of Germany and the Veterinary Department of the Canton of Basel-Stadt (Switzerland).

Multiphoton images were acquired at 256 ms per frame (256×256 pixels) or 128 ms per frame (128×256 pixels) using a custom-built two-photon microscope as described previously[28]. Whole-cell patch-clamp recordings from identified mitral cells were performed using an Axoclamp 2B or Multiclamp 700B amplifier (Axon Instruments, Molecular Devices) and were usually stable for 1–2 h as assessed by changes in input resistance. All recorded trigger neurons were located in the lateral olfactory bulb, 200–300 μm dorsal from the ventral pole of the olfactory bulb. The intracellular solution contained 130 mM potassium gluconate, 10 mM sodium gluconate, 10 mM sodium phosphocreatine, 4 mM NaCl, 4 mM Mg-ATP, 0.3 mM Na-GTP and 10 mM HEPES (pH 7.25).

For forward optical probing, 1-s current steps were injected into mitral cells at a rate of 0.1 Hz. Current amplitudes were adjusted for each mitral cell to evoke 30–50 action potentials (38 ± 13, mean ± s.d.). At each focal plane, the stimulus was repeated 20 times. Peri-simtulus time series of two-photon images were averaged over repetitions and images during a 2-s period following stimulus onset were averaged over time. The resulting image was then divided pixel-wise by an image time-averaged over 2 s preceding stimulus onset to generate a two-dimensional map of relative fluorescence changes ($\Delta F/F$).

Somata of followers were outlined manually in each focal plane. For all followers, the time course of fluorescence change was time-locked to the stimulus. A nonparametric Mann-Whitney U test was used to test that the $\Delta F/F$ values during a 3-s period after stimulus onset were significantly different ($P < 0.05$) from the average $\Delta F/F$ values during the 2 s before stimulus onset. Potential followers that did not fulfill this criterion were excluded from the analysis.

Analysis of follower distribution. XY positions of followers were determined by the center of mass of the outlined soma. The Z position was given by the focal plane. The XYZ positions of all follower neurons were then transformed into spherical coordinates assuming a spherical olfactory bulb of 250-μm radius with its origin in the focal plane of the trigger neuron. The polar axis was defined to run through the trigger neuron. The probability of finding a follower neuron at a given polar angle was then fitted to an exponential function by minimizing the Cramer-von Mises statistic. To reduce the effect of spatial jitter, we discarded follower neurons near the center of the sphere (radius < 50 μm, <1% of all followers). To correct for the limited field of view, we used the following model distribution (**Fig. 5d**).

$$F(d) = \frac{\int_0^d e^{-\frac{x}{l}} f(x)\,dx}{\int_0^\pi e^{-\frac{x}{l}} f(x)\,dx}$$

Here l is the fit parameter, d is the polar angle and f is the relative frequency with which a voxel imaged and the trigger neuron subtended an angle of x at the center of the olfactory bulb. As a consequence of measurement inaccuracies, we expect this procedure to overestimate the length constant. Thus, networks with smaller length constants were also simulated (**Fig. 7b** and **Supplementary Fig. 15**).

In each olfactory bulb, positive calcium signals were detected in multiple focal planes (5.4 ± 3.3 planes, $n = 20$ olfactory bulbs). Of the 262 follower neurons, 2 neurons expressed the mitral cell marker and projected dendrites to the same glomerulus as the trigger mitral cell. These follower mitral cells were therefore likely to be connected to the trigger neuron by gap junctions or intraglomerular glutameric connections[30,31] and were excluded from further analysis.

The expected number $N_{MC} \approx 68.4$ of followers per mitral cell was calculated by the formula

$$N_{MC} = \frac{\frac{2\pi}{3} N \int_0^\pi e^{-\frac{\vartheta}{l}} \sin(\vartheta) d\vartheta}{\int_0^\pi \int_{-\pi}^\pi \int_0^1 e^{-\frac{\theta}{l}} F(r,\vartheta,\varphi) r^2 \sin(\vartheta) dr d\varphi d\vartheta}$$

where l is the length constant, N is the total number of observed followers and F is the number of experiments for which a given pixel (in normalized spherical coordinates r, ϑ and φ) was in the field of view. The s.d. $SD_{MC} \approx 20.4$ was estimated by the formula

$$SD_{MC} = \sqrt{\frac{\frac{2\pi}{3} \int_0^\pi e^{-\frac{\vartheta}{l}} \sin(\vartheta) d\vartheta}{\int_0^\pi \int_{-\pi}^\pi \int_0^1 e^{-\frac{\theta}{l}} F(r,\vartheta,\varphi) r^2 \sin(\vartheta) dr d\varphi d\vartheta} \times \sum_i \left[N_i - \frac{N \int_0^\pi \int_{-\pi}^\pi \int_0^1 e^{-\frac{\theta}{l}} F_i(r,\vartheta,\varphi) r^2 \sin(\vartheta) dr d\varphi d\vartheta}{\int_0^\pi \int_{-\pi}^\pi \int_0^1 e^{-\frac{\theta}{l}} F(r,\vartheta,\varphi) r^2 \sin(\vartheta) dr d\varphi d\vartheta} \right]^2}$$

where summation is over trigger neurons, N_i is the number of follower neurons observed for trigger neuron i and F_i is the indicator function of the field of view of the corresponding experiment.

Numerical procedures and algorithms. All simulations and most analyses were run on a three-way dual-core IBM x3755 with 16 GB RAM using custom software written in Python with time-critical subsystems implemented in C. Differential equations were solved using an adaptive step size embedded Runge-Kutta-Fehlberg (4, 5) scheme[49]. Moments of the half-wave rectified binormal distribution were obtained by adaptive numerical quadrature using a 61-point Gauss-Kronrod rule[49] on the expression

$$_{\eta,\kappa} M_{n,n}(\rho) = \sum_0^{n-1} {}_\eta M_{n-i} {}_\kappa M_{n-i} \frac{\rho^i}{i!}$$
$$+ \int_0^{\arcsin(\rho)} e^{-\frac{1}{4}\left[\frac{(\eta+\kappa)^2}{1+\sin(\psi)} + \frac{(\eta-\kappa)^2}{1-\sin(\psi)}\right]} \frac{(\rho - \sin(\psi))^n}{2\pi n!} d\psi$$

(cf. M18; **Supplementary Theory**). Amongst other advantages this method avoids loss of significance when calculating covariances. Equations (M31, M32; **Supplementary Theory**) were solved using the secant method.

Olfactory bulb model: connectivity matrices. Synaptic weights were uniform and contacts were drawn suppressing across-population variation of fan-in and fan-out; that is, adjacency matrices were required to have constant row and column sums p. By the Birkhoff-von Neumann theorem[50], any such matrix is a positive linear combination of permutation matrices M_{Π_l}. Inspecting the proof of the theorem, we find that all coefficients in this linear combination can be assumed to equal 1. In the case of topographic connectivity, the M_{Π_l} were generated using a heuristic algorithm that strives to respect preset connection probabilities $P(j, k)$ between pairs of mitral cells x_j, x_k.

The algorithm first determined a set of boundaries β_l such that the expectation $\sum_{\beta_{l-1} \le P(j,k) < \beta_l} P(j,k) - \beta_{l-1} + \sum_{P(j,k) \ge \beta_l} \beta_l - \beta_{l-1}$ was just above a half-integer multiple $(l - \frac{1}{2})N$ of the number N of mitral cells. Starting with $l = 1$ for each 'terrace' $T_l = \{(j,k) | P(j,k) \ge \beta_l\}$, an N permutation was drawn and then shuffled into $T_l - \bigcup_{q<l} \Gamma(\Pi_q)$; that is, random transpositions were applied until the graph $\Gamma(\Pi_l)$ of the resulting permutation Π_l was contained in $T_l - \bigcup_{q<l} \Gamma(\Pi_q)$. Here '−' indicates the relative complement that was taken to avoid collisions. To enforce convergence in acceptable time transpositions that would have increased the number of points of $\Gamma(\Pi_l)$ outside $T_l - \bigcup_{q<l} \Gamma(\Pi_q)$ were rejected. For the same reason, boundaries were adjusted where necessary to ensure that terraces satisfied $\min_j \left| \{k | (j,k) \in T_l\} \right| \ge l + 2$ and $\min_k \left| \{j | (j,k) \in T_l\} \right| \ge l + 2$ with vertical bars denoting the size of a set.

To generate bidirectional synapses only even boundaries β_{2l} were considered. Permutations Π_{2l} were generated subject to the additional constraint that they have no one or two cycles. Π_{2l-1} was chosen to be the inverse of Π_{2l}.

Olfactory bulb model: stimulus time course. All simulations started with resting mitral cells. In most trials, the stimulus was constant in time and only the mitral cell fixed point was recorded. In time-resolved experiments, the stimulus was modeled to evolve according to the formula

$$a_j(t) = \xi + \frac{a_{j,\infty}}{1-\alpha}\left[1 - e^{-\tau_j t+} - \alpha + \alpha e^{-\tau_d t+} \right]$$

where ξ is a constant offset accounting for ORN/mitral cell spontaneous activity, $a_{j,\infty}$ is the afferent strength after adaptation, $\alpha = 0.8$ parameterizes the degree of adaptation, τ_j and $\tau_d = 1/600$ ms denote a rise and a decay rate, respectively. The rise rate τ_j was assumed to be proportional to $a_{j,\infty}$, consistent with electrophysiological recordings of odor responses from olfactory sensory neurons in zebrafish (R.W.F., unpublished observations). The proportionality factor was adjusted such that the time-to-peak ranged between 50 and 400 ms, as has been observed experimentally[16].

Artificial stimulus patterns. Multinormally distributed stimuli with the mean, variance and pair-wise correlation coefficients of the 16 measured afferent patterns were obtained as follows. In a first step, half-wave rectified normal distributions approximating histograms of the 16 measured afferent patterns were determined by minimizing the Cramer-von Mises statistic. The means and variances of these distributions before half-wave rectification were averaged to yield $\mu(a)$ and $\sigma^2(a)$. A matrix $C(a)$ of correlation coefficients was then determined such that application of the threshold $-\frac{\mu(a)}{\sigma(a)}$ to the corresponding standard multinormal would recover the correlation coefficients between the 16 measured afferent patterns. As $C(a)$ was not positive definite, negative entries were replaced with zeroes before the Cholesky decomposition $^TUU = C(a)$ was computed. Finally, the desired pattern size N was chosen, an $N \times 16$ matrix Ξ drawn from the standard normal distribution and the columns of $\mu(a) + \sigma(a)\Xi U$ were used as artificial stimuli.

Monte Carlo simulations. Repeated simulations with different initial values were run with the 10,000 neuron network for six different parameter combinations ($p = 12, 36, 60$; $\mu(a) = 10.7, 33.1$; $\Lambda = -4.5$). For each network, the steady states reached from 128 independently drawn normally distributed (8.5 ± 100, mean ± s.d.) start vectors were computed and found to be numerically identical.

48. Hahnloser, R.H., Seung, H.S. & Slotine, J.J. Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.* **15**, 621–638 (2003).
49. Galassi, M. *et al. Gnu Scientific Library: Reference Manual* <http://www.gnu.org/software/gsl/manual/html_node/> (2009).
50. Bapat, R.B. & Raghavan, T.E.S. *Nonnegative Matrices and Applications* (Cambridge University Press, Cambridge, 1997).