



# Hippocampal Neurogenesis Reduces the Dimensionality of Sparsely Coded Representations to Enhance Memory Encoding

Anthony J. DeCostanzo<sup>1,2\*</sup>, Chi Chung Alan Fung<sup>1</sup> and Tomoki Fukai<sup>1\*</sup>

<sup>1</sup> Laboratory for Neural Coding and Brain Computing, RIKEN Center for Brain Science, Saitama, Japan, <sup>2</sup> Ascent Robotics Inc., Tokyo, Japan

## OPEN ACCESS

### Edited by:

Rong Chen,  
University of Maryland, Baltimore,  
United States

### Reviewed by:

Ryota Kobayashi,  
National Institute of Informatics, Japan  
Shaoyu Ge,  
Stony Brook University, United States

### \*Correspondence:

Anthony J. DeCostanzo  
anthony@ascent.ai  
Tomoki Fukai  
tfukai@riken.jp

**Received:** 09 July 2018

**Accepted:** 29 November 2018

**Published:** 07 January 2019

### Citation:

DeCostanzo AJ, Fung CCA and Fukai T (2019) Hippocampal Neurogenesis Reduces the Dimensionality of Sparsely Coded Representations to Enhance Memory Encoding. *Front. Comput. Neurosci.* 12:99. doi: 10.3389/fncom.2018.00099

Adult neurogenesis in the hippocampal dentate gyrus (DG) of mammals is known to contribute to memory encoding in many tasks. The DG also exhibits exceptionally sparse activity compared to other systems, however, whether sparseness and neurogenesis interact during memory encoding remains elusive. We implement a novel learning rule consistent with experimental findings of competition among adult-born neurons in a supervised multilayer feedforward network trained to discriminate between contexts. From this rule, the DG population partitions into neuronal ensembles each of which is biased to represent one of the contexts. This corresponds to a low dimensional representation of the contexts, whereby the fastest dimensionality reduction is achieved in sparse models. We then modify the rule, showing that equivalent representations and performance are achieved when neurons compete for synaptic stability rather than neuronal survival. Our results suggest that competition for stability in sparse models is well-suited to developing ensembles of what may be called memory engram cells.

**Keywords:** dimensionality reduction, hippocampus, pattern separation, neuromorphic computing, feed-forward neural network, synaptic pruning, synaptic turnover, synaptic plasticity

## 1. INTRODUCTION

### 1.1. What Is Known

The hippocampal dentate gyrus (DG) is known to participate in the generation and maintenance of spatio-contextual memories via groups of cells whose activity is causally responsible for the recollection of particular associations (Josselyn et al., 2015; Tonegawa et al., 2015). The DG is noted for a combination of distinctive properties, including adult neurogenesis of the principle granule cells (Wu et al., 2015; Gonçalves et al., 2016) and extremely sparse activity (Jung and McNaughton, 1993; Leutgeb et al., 2007; Danielson et al., 2016; Diamantaki et al., 2016).

Since most adult-born neurons rapidly die, it has long been hypothesized that they must compete amongst themselves, and with mature neurons, for survival dependent upon their contribution to behavior (Bergami and Berninger, 2012). Consistent with this notion, newly adult-born cells integrate into the DG in an experience-dependent manner (Kempermann et al., 1997b; Gould et al., 1999; Bergami et al., 2015; Alvarez et al., 2016; Zhuo et al., 2016), and numerous studies have demonstrated that either ablation (Clelland et al., 2009; Sahay et al., 2011), or *in vivo* silencing of activity (Danielson et al., 2016; Zhuo et al., 2016) or synaptic output (Nakashiba et al., 2012) of these

cells impairs discrimination of hippocampus-dependent associative memories, while enhancing survival of these cells can enhance such performance (Sahay et al., 2011). Similar interventions that silence adult-born cells after learning have shown that retrieval of recent memories is impaired (Gu et al., 2012).

Experience induces synaptic competition among adult-generated granule cells for contacts to CA3 neurons resulting in axonal retraction by mature cells induced by young cells (Yasuda et al., 2011). Elsewhere in both the central (Fitzsimonds et al., 1997; Tao et al., 2000; Du and Poo, 2004) and peripheral nervous systems (Sharma et al., 2010; Zhou et al., 2012), the strength of a neurons output synapses can retrogradely adjust the strength of its input synapses. It has been suggested that this biological phenomenon could encode a neurons performance errors to achieve a similar effect to the artificial backpropagation of error so commonly employed in training neural networks (Harris, 2008). Adult-born DG granule cells reach their targets in CA3 after about 4–6 weeks (Toni et al., 2008), overlapping with when they begin to participate in memory encoding (Clelland et al., 2009; Sahay et al., 2011; Nakashiba et al., 2012; Danielson et al., 2016; Zhuo et al., 2016), and thus may begin to receive signals from CA3 that indicate the success of their contribution to useful representations. The combination of these results suggests that neurogenesis may endow the DG with a kind of learning rule—DG neurons compete with each other for target-derived factors through their synaptic contact to CA3, in turn, influencing their probability of survival. Such a learning rule is the focus of our study.

In an apparently distinct thread of research, sparse activity in recurrent Hopfield-like networks is shown to reduce the interference between stored memories (Tsodyks and Feigel'man, 1988; Amit and Fusi, 1994) and, in models of vision, to enable the efficient representation of naturalistic images as combinations of statistically independent components (Olshausen and Field, 1996; Bell and Sejnowski, 1997), ideas that have roots in the efficient coding hypothesis (Barlow, 1961). In cortical models consisting of a single hidden layer multilayer perceptron with random input weights, it has been shown that pattern decorrelation (often called pattern separation in the neurogenesis literature) is not sufficient to yield proper memory retrieval in the presence of noise (Barak et al., 2013; Babadi and Sompolinsky, 2014). Instead, memory retrieval depends upon a balance between decorrelation of input patterns and generalization of those patterns to the correct class. In such models, sparseness improves memory retrieval by reducing the tradeoff between decorrelation and generalization (Barak et al., 2013). This apparent tradeoff has been analytically expressed in terms that reflect the counterintuitive amplification of noise by sparse coding (Babadi and Sompolinsky, 2014). As a result, there is a theoretical limit on the benefits provided by sparseness in a hidden layer with random input weights (Barak et al., 2013; Babadi and Sompolinsky, 2014). This limitation led some authors to suggest that random weighting is at least partly responsible for limiting the benefits of sparse coding (Babadi and Sompolinsky, 2014).

## 1.2. Our Contribution

One interpretation of these studies is that pattern classification performance, rather than pattern separation, as it has been defined in the neurogenesis literature, may be the appropriate measure of memory performance. We hone our questions into a framework similar to that employed in previous studies of sparse cortical representations (Barak et al., 2013; Babadi and Sompolinsky, 2014), a single-hidden layer, randomly connected feedforward neural network. Within this framework we represent the activities of the neurogenic cells of the dentate gyrus in the hidden layer. With only minimal assumptions, such a network can learn generalizable, nonlinear classifications (Barak et al., 2013), while allowing us to implement sparse coding, synaptic plasticity, and competition among DG neurons for contact with CA3. By supervising the output, the network is trained and then tested for discrimination between sets of input patterns.

We first demonstrate that our neuronal turnover rule, employing randomly drawn input weights, markedly increases the discrimination performance over the initial condition of random projection that was previously studied (Barak et al., 2013; Babadi and Sompolinsky, 2014). The rule exploits sparse coding such that the longer neuronal turnover is allowed to proceed, the sparser the optimal coding level. Since our input weights are always drawn randomly, our results suggest that the sparsening of the optimal code is due to the achievement of a particular hidden layer representation rather than a structuring of the input weights, as was the case explored by Babadi and Sompolinsky. Thus our work complements theirs by suggesting a learning rule via which very sparse codes are optimal for random input weights without require fine tuning.

We show that our rule induces a contextual preference among DG neurons, partitioning the population into ensembles whose average activities are biased for their respective contexts. This is equivalent to dimensionality reduction of the contextual representations in the DG. The final classification performed by the CA3 readout thereby suffers less errors during generalization. We demonstrate that the final achievable discrimination between contextual memories is constrained by the distribution of singular values of the DG representation, such that the sparse code can evolve to a greater difference in the representation space. We then construct a more general model based on evidence that the strength of a neurons output synapses can influence that of its input synapses via internal signals (Fitzsimonds et al., 1997; Tao et al., 2000; Du and Poo, 2004; Sharma et al., 2010; Zhou et al., 2012). This rule similarly reduces the dimensionality of the representation while shifting the activity-dependence toward sparser levels, improving memory performance. Our results suggest that axonal competition for target-mediated stability in sparse models is a novel form of encoding that does not require synaptic fine-tuning, and could be employed across many sparsely coded systems of the brain.

## 2. MATERIALS AND METHODS

### 2.1. Representations of Contexts

We represent the activity state of a population of EC neurons in response to a stimulus as a vector  $\xi$ , the elements of which

neurons that are either spiking,  $\xi_j = +1$ , or not spiking  $\xi_j = -1$ . Patterns are split evenly into two contexts representing the two contexts that the network must learn. The synaptic current of a given DG unit  $i$  for pattern  $\mu$  is defined as:

$$g_i^\mu = \sum_j^M J_j \xi_j^\mu \quad (1)$$

and its activity is given by a threshold function of the synaptic current controlled by  $\theta$ :

$$S_i^\mu = \text{sgn}(g_i^\mu - \theta) \quad (2)$$

The CA3 synaptic current is defined similarly as the weighted sum of the input from DG:

$$h^\mu = \sum_i^N W_i S_i^\mu \quad (3)$$

For every  $\mu$ 'th pattern we want the output of the trained network,  $\hat{\eta}^\mu = \text{sgn}(h^\mu)$  to be equal to a randomly pre-chosen target output state for the CA3 unit, either spiking  $\eta^\mu = +1$ , or not spiking,  $\eta^\mu = -1$ , for all patterns.

## 2.2. Training the Network With Neurogenesis

The task of the network is to use the training patterns to find a  $W$  such that when presented with patterns of a given class to which the network has not been explicitly trained it can correctly generalize, i.e., it will still output the correct class. We train the CA3 output weights in a similar manner to Barak et al. (2013). We assume that the activity of the EC consists of random, uncorrelated prototype patterns,  $\xi$ , that determine their corresponding current in the DG,  $g_i = \sum_j^M J_j \xi_j^\mu$ . We then assume there is noise, or variability in the system such that each prototype pattern is actually represented by a group of noisy instances of the prototype that are generated by flipping the sign of elements of the vector  $\xi$  with a fixed probability  $\nu = 0.2$ . This allows us to calculate the mean synaptic current of a given DG unit  $i$  for pattern  $\mu$  as:

$$\bar{g}_i^\mu = g_i^\mu (1 - 2\nu). \quad (4)$$

Consider the difference between two noisy instances of a prototype pattern, say  $g_i^\mu(t)$  at the  $t$ -th iteration and  $g_i^\mu(t')$  at the  $t'$ -th iteration:

$$\begin{aligned} g_i^\mu(t) - g_i^\mu(t') &= \sum_{\substack{+1 \text{ entries flipped in } t \\ \text{but not in } t'}} J_{ij} \times (+2) + \sum_{\substack{+1 \text{ entries not flipped in } t \\ \text{but in } t'}} J_{ij} \times (-2) \\ &+ \sum_{\substack{-1 \text{ entries flipped in } t \\ \text{but not in } t'}} J_{ij} \times (-2) + \sum_{\substack{-1 \text{ entries not flipped in } t \\ \text{but in } t'}} J_{ij} \times (+2) \end{aligned}$$

Here the sign accompanying “2” will be absorbed into  $J_{ij}$  to simplify the calculation because  $J_{ij} \sim \mathcal{N}(0, 1)$ .

$$\begin{aligned} \langle [g_i^\mu(t) - g_i^\mu(t')]^2 \rangle &= \left\langle \left( \sum_{\substack{\text{entries flipped in } t \\ \text{but not in } t'}} J_{ij} \times 2 \right. \right. \\ &\quad \left. \left. + \sum_{\substack{\text{entries not flipped in } t \\ \text{but in } t'}} J_{ij} \times 2 \right)^2 \right\rangle \\ &= 4 \left\langle \sum_{\substack{M\nu(1-\nu) \text{ terms}}} J_{ij}^2 + \sum_{\substack{M\nu(1-\nu) \text{ terms}}} J_{ij}^2 \right\rangle \\ &= 8M\nu(1-\nu). \end{aligned}$$

Here  $\langle J_{ij}^2 \rangle = 1$  because  $J_{ij} \sim \mathcal{N}(0, 1)$ . On the other hand, let  $g_i^\mu(t) = \bar{g}_i^\mu + \delta g_i^\mu(t)$ , then we have

$$\begin{aligned} \langle [g_i^\mu(t) - g_i^\mu(t')]^2 \rangle &= \langle \delta g_i^\mu(t)^2 + \delta g_i^\mu(t')^2 - 2\delta g_i^\mu(t) \delta g_i^\mu(t') \rangle \\ &= \langle \delta g_i^\mu(t)^2 \rangle + \langle \delta g_i^\mu(t')^2 \rangle \\ &= 2 \langle \delta g_i^\mu(t)^2 \rangle. \end{aligned}$$

Hence, the variance of each DG unit is given by

$$\begin{aligned} \sigma_g^2 &= \langle \delta g_i^\mu(t)^2 \rangle \\ &= 4M\nu(1-\nu). \end{aligned} \quad (5)$$

Since the synaptic currents of the  $i$ -th DG unit for noisy instances are sum of many randomly altered numbers, those synaptic currents can be assumed to be Gaussian. The expected value of the activity of the  $i$ -th DG unit can be deduced by

$$\begin{aligned} \bar{S}_i^\mu &= -1 \times \int_{-\infty}^{\theta} dgf(g | \bar{g}_i^\mu, \sigma_g^2) + 1 \times \int_{\theta}^{+\infty} dgf(g | \bar{g}_i^\mu, \sigma_g^2) \\ &= -F(\theta | \bar{g}_i^\mu, \sigma_g^2) + 1 - F(\theta | \bar{g}_i^\mu, \sigma_g^2) \\ &= \text{erf}\left(\frac{\bar{g}_i^\mu - \theta}{\sqrt{2}\sigma_g}\right), \end{aligned} \quad (6)$$

where  $f(g | \bar{g}_i^\mu, \sigma_g^2)$  and  $F(g | \bar{g}_i^\mu, \sigma_g^2)$  are probability density function and cumulative density function, respectively, of a normal distribution with mean  $\bar{g}_i^\mu$  and variance  $\sigma_g^2$ . To arrive at the desired target output, e.g.,  $\eta = \text{sgn}(\sum_{i=1}^N W_i S_i^\mu)$  for all  $\mu$ , the cost function

$$E = \sum_{\mu=1}^P \left( \eta^\mu - W^T \bar{S}^\mu \right)^2$$

should be minimized. We then find the linear least squared error solution to  $W$ ,

$$W^T = \underset{\bar{W}}{\operatorname{argmin}} \left[ \sum_{\mu}^P (\eta^{\mu} - \bar{W}^T \bar{S}^{\mu})^2 \right], \quad (7)$$

by taking the Moore-Penrose pseudoinverse of the matrix  $\bar{S}$ ,

$$W^T = \eta \bar{S}^* \quad (8)$$

$$= \eta V \Sigma^* U^T, \quad (9)$$

where  $S^*$  is the pseudoinverse of  $S$ ,  $U$  and  $V$  are the matrices of left and right singular vectors, respectively, and  $\Sigma$  is the matrix of singular values. Here the Moore-Penrose pseudoinverse enables us to look for the best approximation using column vectors of  $\bar{S}^{\mu}$ . The approximation is also the best-fit solution minimizing the cost function. More explanation about Moore-Penrose pseudoinverse can be found in **Appendix A**.

To implement the synaptic competition underlying neurogenesis we compare three different models. In Model 1 (**Figures 1, 2, 3, 5**), at each time step we kill DG units corresponding to the bottom 30% of absolute values in vector  $W$ , i.e., the input weights to those units are re-randomized. In Model 2 we explore a multicontext case presented in **Figure 6** in which each DG unit projects to multiple CA3 units, therefore we take the sum of the absolute value of each DG units weight vector and compare this value across all DG units.

In Model 3 presented in **Figure 7**, rather than re-randomizing all input weights of selected DG units, we determine the probability of synaptic turnover of each DG unit from a linear transfer function of its DG-CA3 weight (**Figure S3A**). Results presented in **Figure 7** are from the mean of 100 simulations with a slope = 2.5 for the linear transfer function.

### 2.3. Analyzing Performance of the Network

To evaluate the performance of the network, the signal-to-noise ratio is introduced. The signal is defined by the square of the expectation of the difference between CA3 synaptic currents corresponding to (+) context, i.e.,  $\eta^{\mu} = +1$ , and (-) context, i.e.,  $\eta^{\mu} = -1$ , among all the patterns.

$$\text{Signal} = [\mathbf{E}(h^{\mu} | \eta^{\mu} = +1) - \mathbf{E}(h^{\mu} | \eta^{\mu} = -1)]^2. \quad (10)$$

To progress, we define the context-bias of a given DG unit  $i$ ,  $\Psi_i$ , as the difference between the fraction of (+) context patterns,  $f_i^+$ ,

and the fraction of (-) context patterns,  $f_i^-$ , to which it responds.

$$\Psi_i = f_i^+ - f_i^-, \quad (11)$$

where  $f_i^{\pm}$  is the fraction of ( $\pm$ ) context patterns activating DG unit  $i$ . On the other hand,

$$\begin{aligned} (S\eta^T)_i &= \sum_{\text{active for a (+) pattern}} (+1) \times (+1) \\ &+ \sum_{\text{inactive for a (+) pattern}} (-1) \times (+1) \\ &+ \sum_{\text{active for a (-) pattern}} (+1) \times (-1) \\ &+ \sum_{\text{inactive for a (-) pattern}} (-1) \times (-1) \end{aligned} \quad (12)$$

$$= Pf_i^+ - P(1 - f_i^+) - Pf_i^- + P(1 - f_i^-) \quad (13)$$

$$= 2P\Psi_i. \quad (14)$$

Note that  $S$  here is a matrix, whose column vectors are activities of DG neurons for different input patterns.  $\eta$  is a label vector, where entries are expected output (CA3) of the patterns. Then we can then express the context-bias in terms of  $S$  and  $\eta$  in a matrix-vector equation as:

$$\Psi = \frac{1}{2P} S\eta^T. \quad (15)$$

With this, the signal can be expressed as

$$\text{Signal} = \left[ 2 \sum_i W_i (f_i^+ - f_i^-) \right]^2 \quad (16)$$

$$= [2W^T\Psi]^2. \quad (17)$$

On the other hand, we define the noise as the sum of variances of this current for the (+) and (-) contexts respectively:

$$\text{Noise} = \mathbf{Var}(h^{\mu} | \eta^{\mu} = +1) + \mathbf{Var}(h^{\mu} | \eta^{\mu} = -1) \quad (18)$$

$$\stackrel{N \rightarrow \infty}{=} 4 \sum_i W_i^2 [f_i^+ - (f_i^+)^2 + f_i^- - (f_i^-)^2] \quad (19)$$

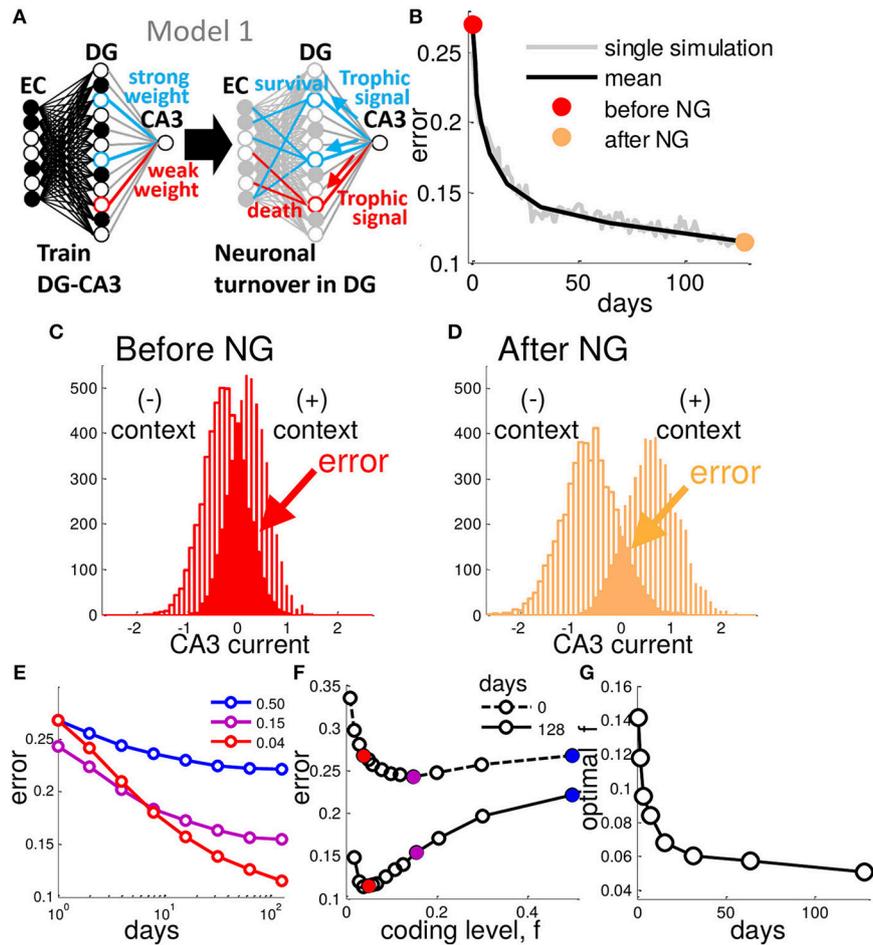
$$= 4 \sum_i W_i^2 [f_i^+ (1 - f_i^-) + f_i^- (1 - f_i^+) - (f_i^+ - f_i^-)^2] \quad (20)$$

From these expressions we derive the signal to noise ratio (SNR) in terms of  $f_i^{\pm}$  and  $W^T$ .

$$\frac{\text{Signal}}{\text{Noise}} = \frac{[2W^T\Psi]^2}{4 \sum_i W_i^2 [f_i^+ (1 - f_i^-) + f_i^- (1 - f_i^+) - (f_i^+ - f_i^-)^2]} \quad (21)$$

$$= \frac{\sum_i W_i^2 \Psi_i^2 + \sum_{i \neq j} W_i \Psi_i W_j \Psi_j}{\sum_i W_i^2 [f_i^+ (1 - f_i^-) + f_i^- (1 - f_i^+) - (f_i^+ - f_i^-)^2]} \quad (22)$$

$$\stackrel{N \rightarrow \infty}{=} \frac{\sum_i W_i^2 (f_i^+ - f_i^-)^2}{\sum_i W_i^2 f_i^+ (1 - f_i^-) + \sum_i W_i^2 f_i^- (1 - f_i^+) - \sum_i W_i^2 (f_i^+ - f_i^-)^2} \quad (23)$$



**FIGURE 1 |** Neurogenesis enhances generalization performance. **(A)** In Model 1, after a weight vector is assigned by training, DG units with weak weights to CA3 are replaced with new randomly connected units. **(B)** At each day of training the network is tested with randomly generated patterns belonging to one of the two contexts. This generalization error decreases as a function of the number of iterations of neural turnover. Single simulation (gray) and mean of many simulations (black), before (red point) and after (orange point) neurogenesis. **(C,D)** CA3 Synaptic current distribution for all test patterns representing the two contexts before **(C)** and after **(D)** 128 iterations (days) of neural turnover. Results are from a network of 200 EC, 500 DG neurons and a single CA3 readout. Each context consists of 50 EC patterns with input noise,  $v$ , fixed at 0.2, and theta is chosen to yield a coding level of  $f = 0.04$ , turnover rate is fixed at 0.30 (See Experimental Procedures). **(E)** Mean error is shown decreasing as a function of the number of iterations of neural turnover for three different coding levels. **(F)** Error is shown as a function of coding level before and after 128 iterations of neural turnover. After neurogenesis the performance is improved at all levels of sparseness (all coding levels,  $f$ ). **(G)** The coding level at which minimum error occurs (optimal  $f$ ) is plotted vs. the number of iterations of neural turnover. Neural turnover favor a sparser (reduced) coding level. Mean error is calculated as the mean of 20 simulations.

This expression allows us to observe the intuitive relationship between the context-bias of DG cells and the SNR. The second term in the numerator of Equation (22) should vanish as  $N \rightarrow \infty$ , as it sums random numbers centered at zero. For  $W^T$ ,

$$W^T = 2P\Psi^T U (\Sigma^*)^2 U^T \tag{24}$$

In the presentation of our results it is useful to let  $\hat{\Psi}_i^T = \Psi^T u_i u_i^T$  where  $u_i$  is the  $i$ -th column of  $U$ , so that

$$W^T = 2P \sum_i^N \sigma_i^{-2} \hat{\Psi}_i^T \tag{25}$$

observing the weight vector as a weighted sum of projected context-bias vectors. The derivation of this equation can be found in **Appendix B**.

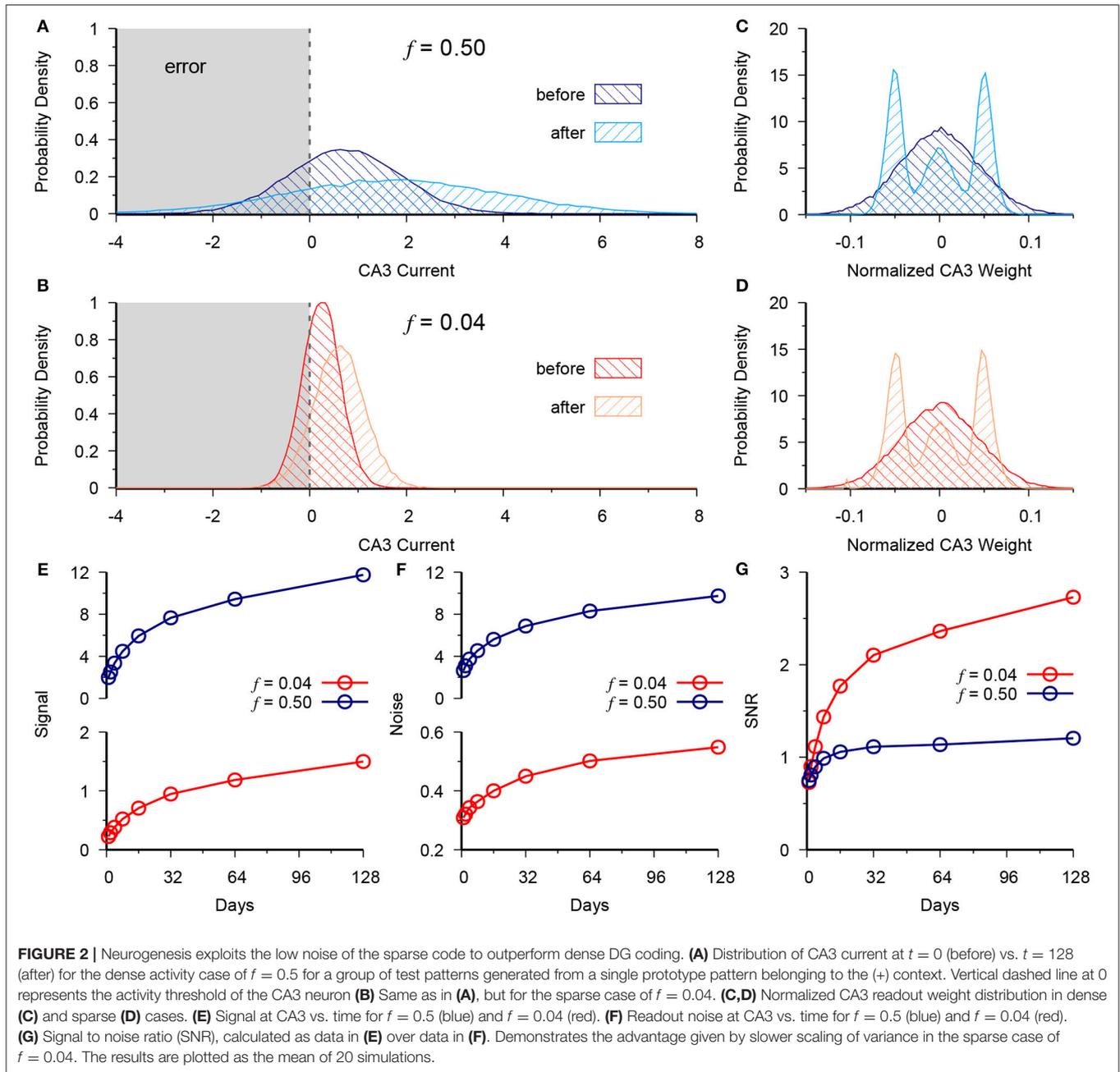
### 2.4. Dimensionality of DG Contextual Representation

From above, the weight vector is defined as:

$$W^T = \eta V \Sigma^* U^T \tag{26}$$

Permitting us to rewrite the weight vector as a linear sum of coefficients produced with their respective left singular vectors:

$$W = \sum_{i=1}^D \alpha_i u_i, \tag{27}$$



where  $D$  is the dimension of the square matrix  $U$ . The  $D$  dimensions are ranked from 1 to  $D$  according to their corresponding coefficients. We define a cumulative weight vector of a given dimensionality as:

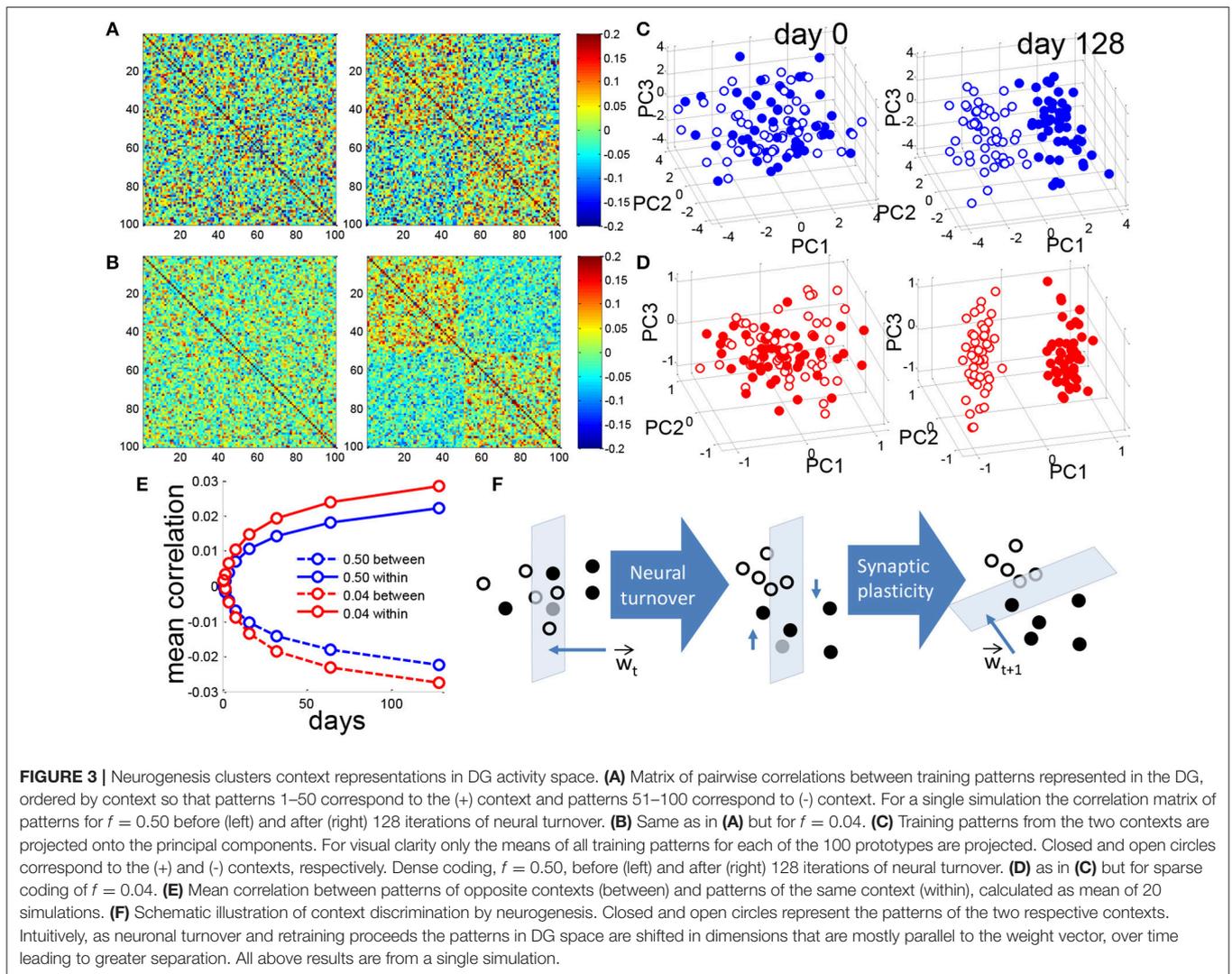
$$\hat{W}_d = \sum_i^d \alpha_i u_i, \quad (28)$$

where  $d$  takes a value from 1 to  $D$ , representing the number of dimensions chosen for a given cumulative weight vector. We then define the cumulative performance,  $\text{perf}_{\text{cum}} = \frac{(0.5 - \text{err})}{0.5}$ ,

where the error is calculated for every cumulative weight vector (**Figure S3C**).

## 2.5. Model Parameters

All results in **Figures 1–5** are from a network with 200 EC, 500 DG units and a single CA3 unit. Data in **Figure 6** are from the same size network except that the number of CA3 units is increased to 3 to allow for multicontext discrimination. In **Figures 1, 2, 3, 4, 7** the network was trained with the mean representation of each of 100 prototype patterns as described above. In **Figure 6** the network was trained with 8 groups of 12 prototypes, to represent 8 subcontexts, by calculating



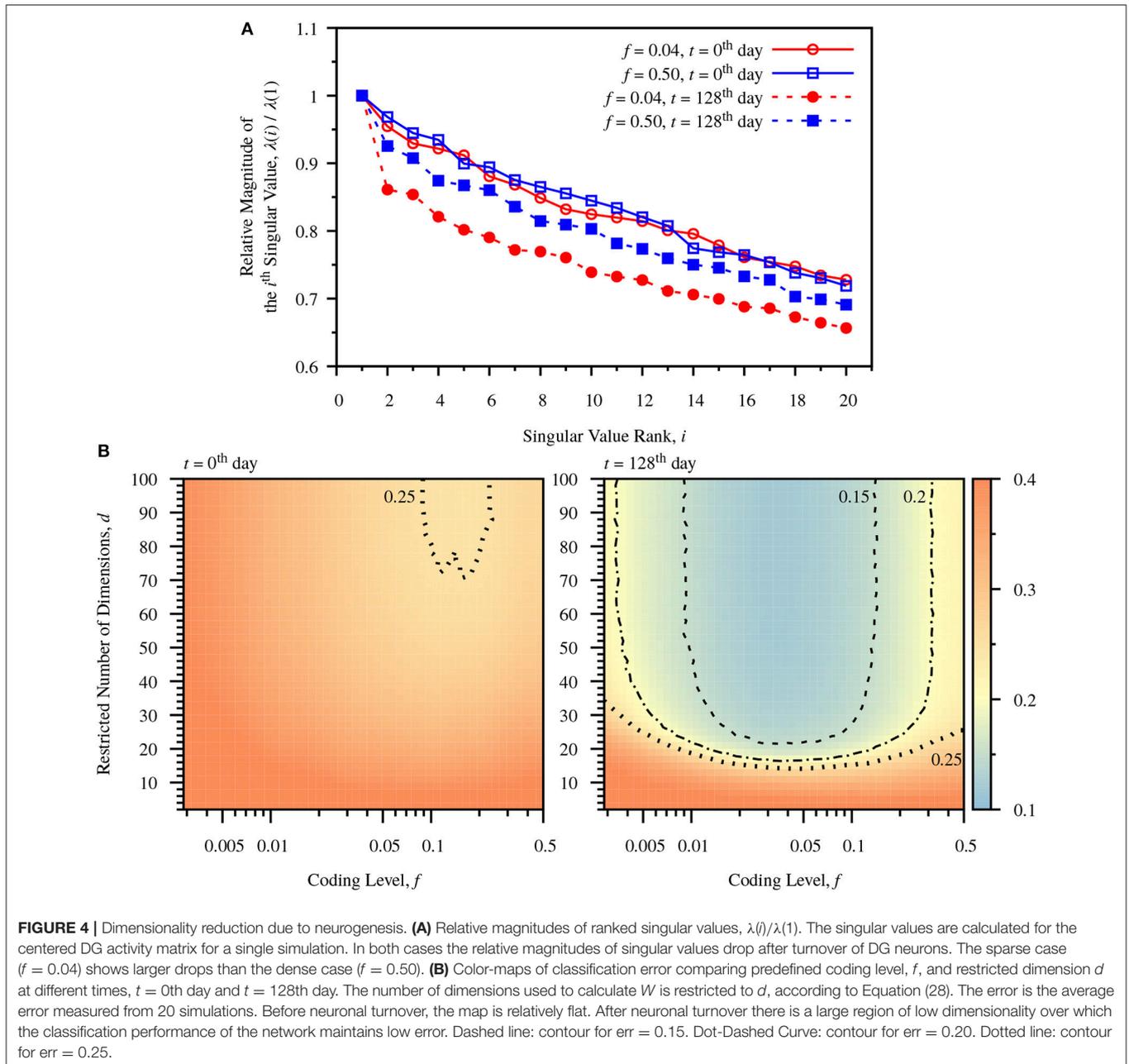
the mean representation of each prototype assuming some variability as described above. In **Figure 5** the network is the same size, however, training consisted of 100 noisy instances of 100 prototypes, rather than using the mean representation of each prototype. This is because we wished to relate the results of this training directly to the equations that we derived for the SNR from the SVD as above. Both types of training gave similar qualitative results, therefore they are not explicitly compared.

### 3. RESULTS

#### 3.1. Network Model for Adult Neurogenesis in the Formation of Associative Memories

We implement a feed-forward multilayer perceptron in which pattern discrimination (classification) is the readout of performance. The model consists of a three-layer network including entorhinal cortical inputs (EC), dentate gyrus (DG), and a CA3 output (**Figure 1A**). We assume that a given DG cell

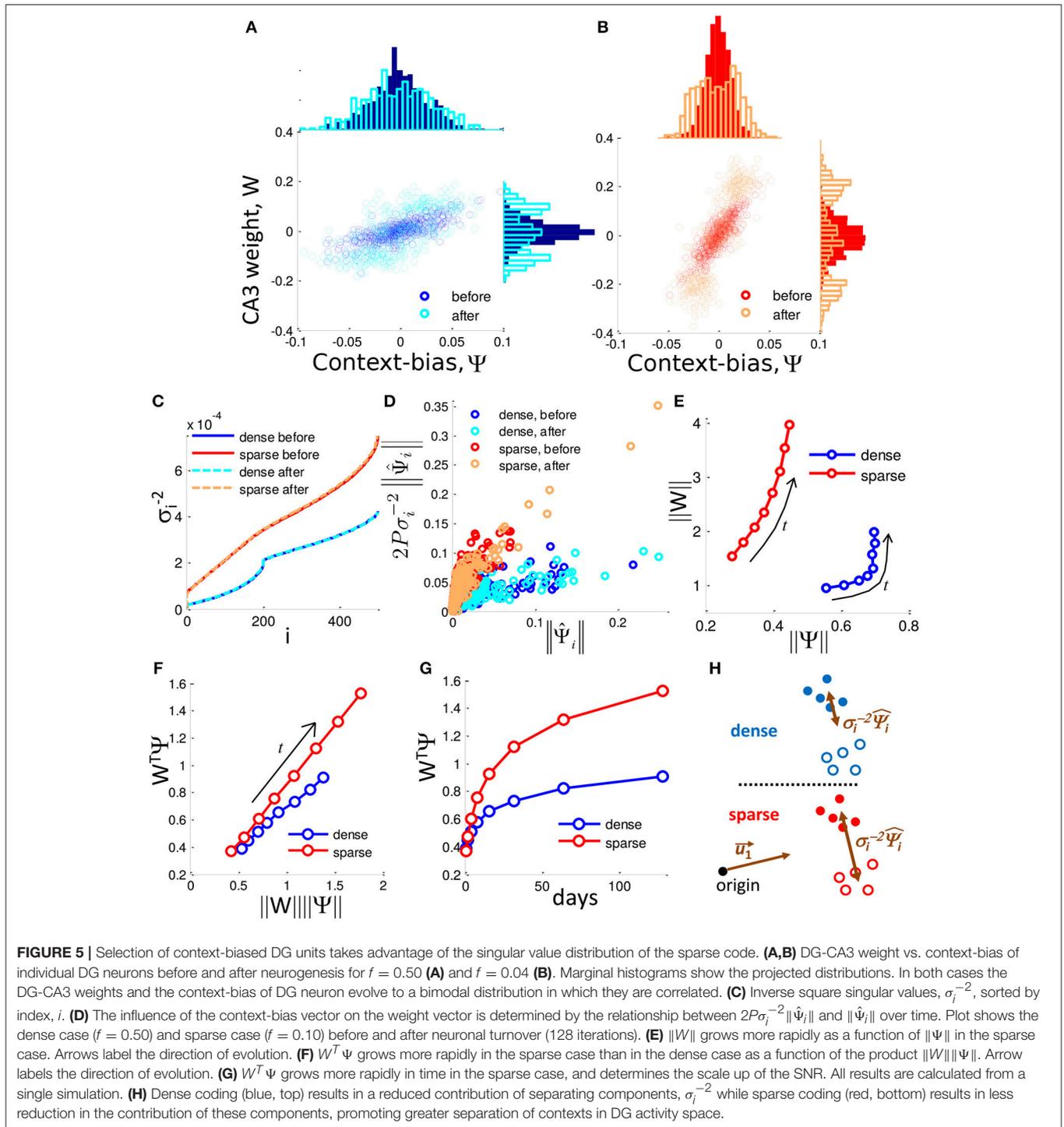
receives a weighted sum of its inputs from the EC. Thus the total current into the  $i$ 'th DG cell in response to the  $\mu$ 'th EC pattern,  $\xi^\mu$ , is given by  $g_i^\mu = \sum_j^M J_j \xi_j^\mu$  where the weights,  $J_j$ , are drawn randomly from a normal distribution,  $\mathcal{N}(0, 1)$ , and its activity is determined by the nonlinear function of this current,  $S_i^\mu = \text{sgn}(g_i^\mu - \theta)$ , where we refer to  $\theta$  as the activation threshold, which is a tunable parameter we use to control the coding level, i.e., the expected value of the fraction of patterns to which a given unit responds, defined as  $f = \frac{1}{2^p} \sum_{\mu=1}^P \frac{1}{N} \sum_{i=1}^N (S_i^\mu + 1)$ . We define a context as a group of prototypical activity patterns generated in the EC, where each pattern represents a stimulus that is present in the given context. We assume that there is random variability in the environment, or within the system such that among these patterns each binary element may be flipped with probability  $\nu$ . Averaging for each prototype over the input noise  $\nu$ , we obtain corresponding mean input currents for each DG cell for each prototype pattern,  $\bar{g}_i^\mu = g_i^\mu (1 - 2\nu)$ , with variance  $\sigma_g^2 = 4M\nu(1 - \nu)$  (See Materials and Methods). This gives us a set of mean prototype activity patterns in the



space of DG activity, where each neurons activity is defined as  $\bar{s} = \text{erf}[(\theta - \bar{g}_i^\mu)/(\sqrt{2}\sigma_g)]$ . The network is said to perform contextual discrimination when the CA3 output correctly reads out the DG patterns according to the target label for the EC context to which those patterns belong.

To train the network, we randomly assign to the  $\mu$ -th EC pattern, a CA3 target,  $\eta^\mu$ , taking the value +1 or -1. Thus, assuming that  $\theta$  is held constant during training, the task of the network is to find a input weight matrix,  $J$ , and an output weight vector,  $W$ , such that  $W^T \bar{S} = \eta$ , where  $\bar{S}$  is the matrix of DG prototype patterns, and  $\eta$  is the corresponding vector of context labels.

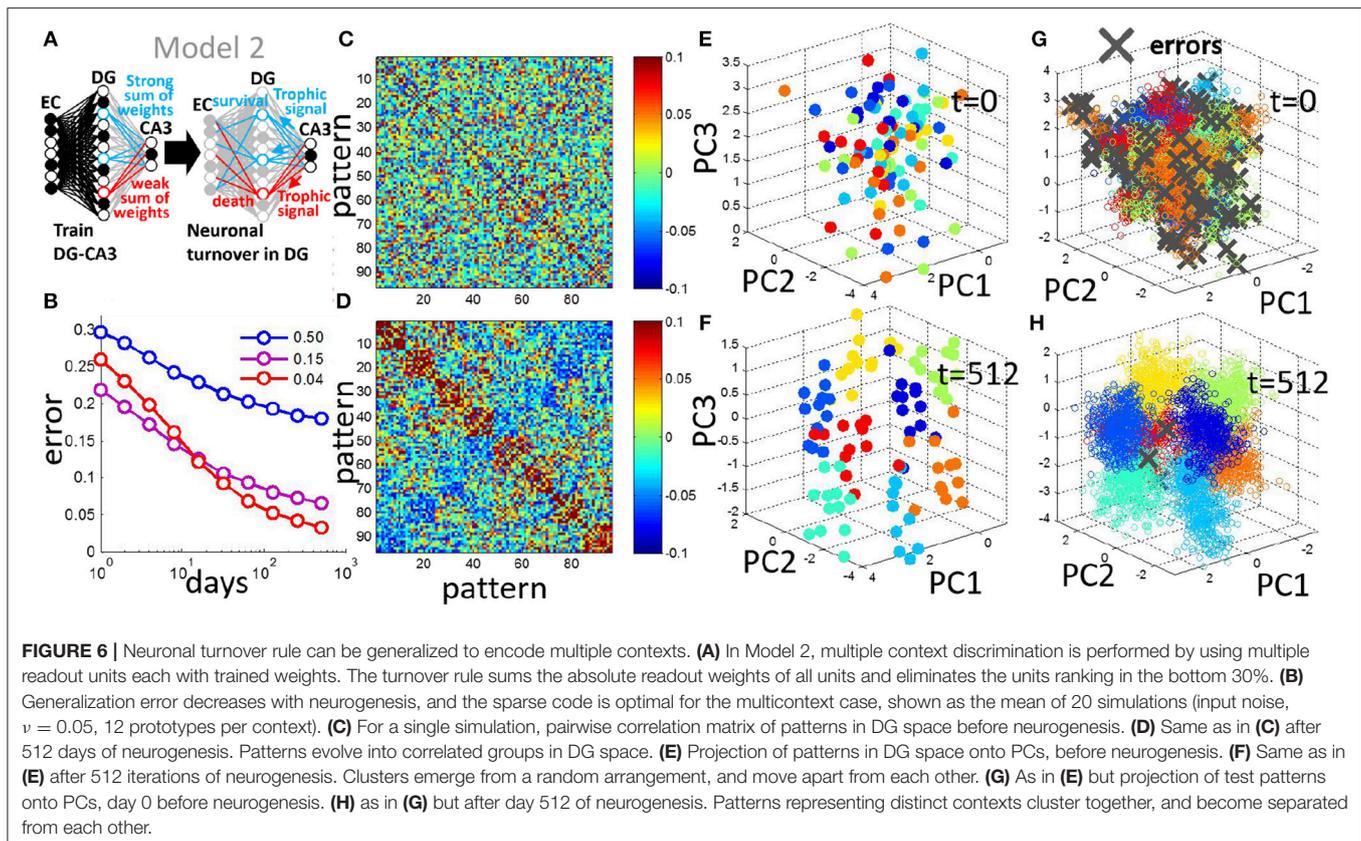
We hypothesize that neurogenesis provides a mechanism by which biology breaks this problem into two steps. We assume that, as in the brain, the time-scale of neurogenesis is much slower than that for synaptic plasticity, allowing us to train the output weights,  $W$ , independently of the input weights,  $J$ . Many learning rules could be used to train  $W$ , such as Hebb's rule, or Support Vector Machine (with a linear kernel), or Linear Discriminant Analysis. We obtained qualitatively similar results with all of these, therefore, to simplify later analysis, we use the pseudoinverse rule yielding  $W^T = \eta \bar{S}^*$ , where  $\bar{S}^*$  is the Moore-Penrose pseudoinverse of the matrix of DG prototype patterns, and  $W^T$  is the transpose of the output weight vector,



$W$ , whose elements are the DG-CA3 weights of the population of DG units (See Materials and Methods). Next we assume that DG neurons compete with each other for connection to CA3 such that the absolute value of  $W_i$  determines their probability of survival, i.e., neurons with large values will receive some trophic signal allowing them to survive, while those with values below some threshold will die, to be replaced by a new randomly

connected unit (**Figure 1A**). Thus training is summarized as follows:

1. Initialize the matrix of random EC-DG weights,  $J$ .
2. Calculate DG-CA3 weight vector,  $W$ , by  $W^T = \eta S^*$ .
3. Eliminate DG units with the weakest  $|W_i|$ s at a predefined percentage (to be stated in the following).



- Those DG units are replaced by new DG units. The EC-DG weights connecting to those new DG units are randomly drawn from a normal distribution  $\mathcal{N}(0, 1)$ .
- Repeat and start from Step 2.

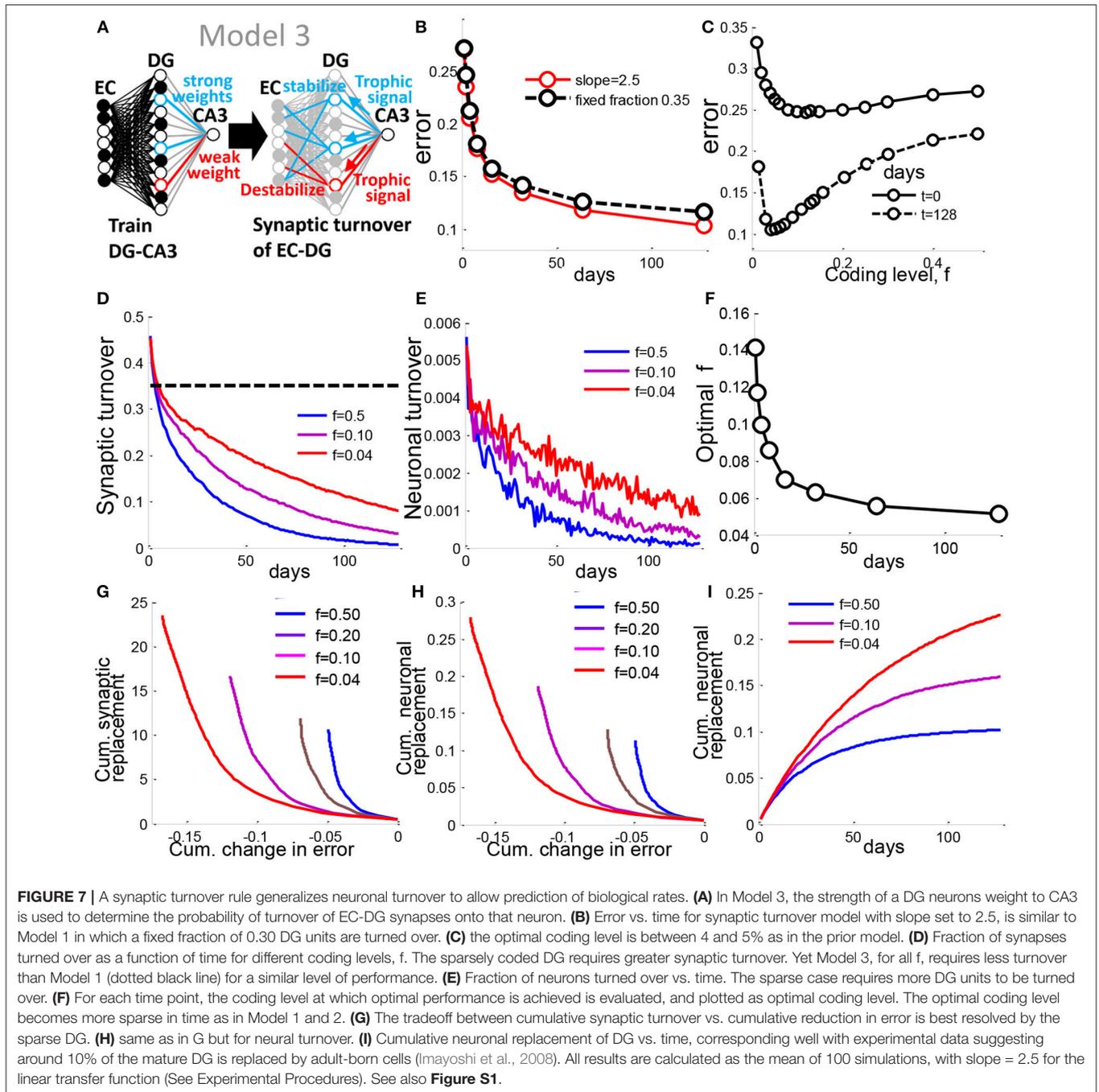
Since the cell cycle in biology corresponds to about 24 h, and each iteration of our model represents the death and birth of neurons, one iteration corresponds to roughly one biological day (the time axes is labeled “days”). One should note that the DG neurons considered in this model are those mature enough to emerge into the dentate gyrus and reach CA3. Those immature adult-born cells unable to reach CA3 are not considered in this model.

We test the network by presenting EC input patterns with a fraction of  $\nu$  bits flipped (corresponding to input noise, or variability) that belong to a known context, taking the CA3 output for the  $\mu$ -th test pattern as  $\hat{\eta}^\mu = \text{sgn}(\sum_i^N W_i S_i^\mu)$ , where  $N$  is the total number of DG units. Then we measure the error on a given test pattern,  $\text{err}^\mu = \begin{cases} 0, & \text{if } \hat{\eta}^\mu = \eta^\mu \\ 1, & \text{otherwise} \end{cases}$ , and mean over all test patterns,  $\langle \text{err}^\mu \rangle_\mu$ , yielding the generalization error. Neuronal turnover of the weakest 30% of DG neurons per day results in a steadily decreasing mean error as a function of the number of iterations (days) of contextual associative learning (**Figure 1B**), thus increasing the performance of this framework relative to the randomly initialized network corresponding to the case studied by Barak et al. (2013) and Babadi and Sompolinsky (2014). The choice of 30% may seem arbitrary, but further

clarification will follow. The error in **Figure 1B** is determined by the overlap between the two underlying distributions of total synaptic current into CA3 for the two contexts in the presence of variability on the input (**Figure 1C**). The sign of the CA3 readout should be opposite for each of the two possible associations, positive or negative for a given pattern belonging to the context with (+1) or (−1) context, respectively. After 128 days of neural turnover the spread between the distributions increases such that the overlap between them, is decreased (**Figure 1D**). From **Figure 1B** we see that the initial drop in error occurs rapidly, i.e., most of the performance gain from neurogenesis occurs within a week.

### 3.2. Neurogenesis Interacts With Sparse Activity to Enhance Contextual Discrimination

Sparseness of granule cell firing is likely induced via a combination of cell-intrinsic and extrinsic properties (Marin-Burgin et al., 2012). We control sparseness by adjusting  $\theta$  which represents the combination of these effects, determining the cells coding level,  $f$ . Neurogenesis increases performance at all coding levels (**Figures 1E,F**). The optimal code becomes more sparse and appears to plateau at around 4–5% of DG cells active (**Figure 1G**). Thus, in contrast to the initial optimal coding level of around 10–15% active, similar to previous reports in a similar framework (Barak et al., 2013; Babadi and Sompolinsky, 2014),



our best performance is achieved at a very sparse activity level that continues to sparsen with time (**Figure 1G**).

The error reduction depends on the turnover rate, i.e., the fraction of neurons targeted for turnover per day (**Figure S1A**), such that longer periods of learning (more iterations of neurogenesis) favored lower turnover rates (**Figure S1B**). On average, the optimal rate of turnover is a monotonically decreasing function of the number of days learning (**Figure S1C**), yielding an optimal turnover rate of around 0.3 at 128 days of learning.

We next analyzed the dynamics of the population of DG neurons. The survival rate of neurons during the time course of encoding the contexts depended on their age, i.e., those born more recently have a survival advantage (**Figure S1D**), indicating the gradual replacement of existing cells with those that are newly born. Neuronal replacement is highest at the beginning of learning, with a fraction of around 0.7 of 1-day old neurons surviving, but after 256 days of learning even 1-day old neurons survive at a very low rate of around 0.04. Whenever there is a sudden change of the contexts in the 2-class case, or addition of

a context in the multiclass case, we would indeed see a sudden jump in the survival rate of newborn neurons. Therefore the survival rate of newborn neurons scales with the learning rate, or the encoding of new information, consistent with experimental findings (Kempermann et al., 1997a,b; Gould et al., 1999).

To explore the relationship between connectivity and sparse coding, we tested the networks performance for varying degrees of connectivity from EC to DG (**Figure S2A**). The performance degrades as input connectivity is reduced, with the performance of sparsely coded models suffering more than that of more densely coded models (**Figures S2A,B**). Nevertheless the optimal coding level is a steeply monotonically decreasing function of the connectivity that is sparse above a connectivity of around 2.5% (**Figure S2C**), suggesting that sparse models perform well over a large range of connectivities.

We next analyzed the CA3 readout to determine why the memory performance scales up more quickly in the sparse vs. the dense coding case as a function of neurogenesis. We observe the total synaptic current coming into CA3 from the DG for a single test pattern that belongs to the (+) context. Accordingly we see that neurogenesis causes a positive shift in the distribution of total synaptic current into CA3 for both the dense (**Figure 2A**) and sparse (**Figure 2B**) cases with the normalized output weights shown in panels C and D, respectively. However, there is an accompanying increase in the spread of this distribution countering the performance gain given by the increased signal, since the tail of the distribution causes errors when it crosses the CA3 decision boundary (**Figure 2A**).

We define the Signal to Noise ratio (SNR) as:

$$\frac{\langle \text{Signal} \rangle}{\langle \text{Noise} \rangle} \stackrel{\text{def}}{=} \frac{\langle h_+ - h_- \rangle^2}{\langle \sigma_{h_+}^2 \rangle + \langle \sigma_{h_-}^2 \rangle}, \quad (29)$$

where  $h_+$  and  $h_-$  are the total synaptic current into CA3 from the patterns of the (+) and (-) contexts, respectively, and  $\sigma_+^2$  and  $\sigma_-^2$  are the respective variances of that current across patterns. In both the dense and sparse cases neurogenesis contributes to a scale-up of the signal (**Figure 2E**) and the noise (**Figure 2F**). Yet, in the signal-to-noise ratio (SNR) we see the superior performance of the sparse case (**Figure 2G**). Due to synaptic competition, the distribution of DG-CA3 weights gradually shifts to higher efficacy synapses for both the dense (**Figure 2C**) and the sparse case (**Figure 2D**).

### 3.3. Neurogenesis, Synaptic Plasticity, and Sparse Activity Cooperatively Facilitate Dimensionality Reduction

We then ask how the representation in the DG changes over time. Prior to neurogenesis there is no correlation among the patterns representing the two contexts for either the dense or sparse case (**Figures 3A,B**, left). After neurogenesis proceeds, for both the dense and sparse case, patterns that belong to a given context become correlated to each other, while those that belong to different contexts become anticorrelated (**Figures 3A,B**, right). Note that for the same amount of neural turnover, the sparse case always achieves a more correlated representation (**Figure 3E**). **Figure 3E** shows the mean correlations within the same context

and across different contexts shown in panels A and B. It suggests that the representations in DG for different patterns in the same context are similar, while representations for patterns in different contexts are more different after training.

For a closer look of the representations before and after the neurogenesis training, Principal Components Analysis (PCA) was used for presentations. Principal Components Analysis reveals that, initially the DG activity patterns are randomly distributed (**Figures 3C,D**, left) but after neurogenesis proceeds, patterns representing the two contexts become clustered, and separated, for both the dense and the sparse case (**Figures 3C,D**, right), while the sparse case clearly shows greater separation along PC1 (**Figure 3D**, right). Note that, though we do not show it here, the separation between clusters became observable after only 10–15 days. Since it becomes clearer with a long simulation time, we report the state at the 128th day for comparison.

We intuitively illustrate the effect of neuronal turnover (**Figure 3F**). Synaptic plasticity, between the DG and CA3, assigns a weight vector at a given time,  $t$ ,  $W_t$ . This weight vector defines a perpendicular hyperplane that separates the patterns defining the two contexts from each other in the space DG activity. Weak synapses, i.e., elements of the weight vector that are near zero, lie in dimensions that are almost perpendicular to the weight vector, and almost parallel to the hyperplane. By killing and replacing those DG units that have weak synapses to CA3 and mostly perpendicular to  $W$ , neuronal turnover randomly shifts the patterns in a direction that is mostly parallel to the hyperplane. On average, after this shift, the contexts are easier to separate when synaptic plasticity draws a new weight vector,  $W_{t+1}$ , and the cycle continues as such. Though step-to-step improvement on a single instantiation is noisy (**Figure 1B**, gray trace) the average performance appears to monotonically decrease (**Figure 1B**, black trace).

To observe the influence of neurogenesis and sparse coding on dimensionality, we observe the singular values,  $\lambda(i)$ , of the centered DG activity matrix,  $S$ , corresponding to the standard deviation of activity patterns in the  $i$ 'th dimension. The ratio of  $\lambda(i)/\lambda(1)$  decreases after neurogenesis for all components in both the dense and sparse case, but the decrease is more profound in the sparse case (**Figure 4A**). To see more clearly how the number of dimensions affects classification performance we observe the effect of restricting the number of components in the weight vector. Observing Equation (28), the weight vector can be decomposed into a sum of weighted components,  $\hat{W}_d = \sum_i^d \alpha_i u_i$ . We observe how the classification error varies as we incrementally add back components to the weight vector up to dimensionality  $d$ , plotting a color map of the number of restricted dimensions vs. coding level. Before neurogenesis this map is relatively flat (**Figure 4B**, left panel), indicating a weak dependence of dimensionality on coding level. In contrast, after neurogenesis the map exhibits a sharp drop in error after a only around 20 components, especially in the sparse coding range around  $f = 0.04$  (**Figure 4B**, right panel). This indicates that neurogenesis reduces the effective dimensionality required for maximal performance at a fixed coding level, and that sparse coding allows for a greater reduction in dimensionality.

### 3.4. The Separation Between Contexts Is Determined by the Context-Bias of Selected DG Neurons

To simplify analysis in the next two sections it is useful to consider training with the matrix of noisy prototype patterns,  $S$ , rather than the matrix of mean prototype patterns,  $\bar{S}$  (See Materials and Methods). We next observe how neuronal competition affects the organization of the DG neuronal population. We define the context-bias of a given DG cell,  $\Psi_i$ , as the fraction of patterns it responds to belonging to the (+) context minus those that belong to the (-) context:

$$\Psi_i \stackrel{\text{def}}{=} (f_i^+ - f_i^-) \quad (30)$$

Therefore, the context-bias takes a value between  $-1$  and  $+1$  and is equal to  $0$  in cases where a DG cell responds to the same number of (-) and (+) context patterns. For the entire DG population, this can be expressed as the context-bias vector,  $\Psi$ ,

$$\Psi = \frac{1}{2P} S \eta^T, \quad (31)$$

where each column of  $S$  is a pattern of DG activity, and  $\eta$  is the vector of target CA3 activities (either  $-1$  or  $+1$ ) for each respective input pattern and is the total number of patterns. The derivation can be found in Equation (15). Note that  $\Psi$  is equivalent to the separation between the means of the patterns representing the two respective contexts (See Materials and Methods). Neurogenesis selects for neurons that are biased for each of the two contexts (Figures 5A,B, Top histogram). Therefore, the distribution of  $\Psi_i$  partitions into 3 groups, those that are biased to respond to context (-), those that are biased toward context (+), and newborn randomly generated neurons whose context-bias is centered on zero (Figure 5B). The two biased groups of surviving neurons therefore form an ensemble that can be thought of as memory engrams for their respective contexts. Note that a DG cells context-bias is correlated with its weight to CA3 (Figures 5A,B, scatter plot). On average, the dense case (Figure 5A, top histogram) consists of DG cells that are more biased between the two contexts than the DG cells of the sparse case (Figure 5B, top histogram). This is because the maximum difference between a neurons responsiveness to the two contexts is limited by the total fraction of patterns to which a neuron can respond, i.e., the coding level. With neuronal turnover, in both cases, the average context-bias, and the average CA3 weight increases (Figures 5A,B, top histograms, right histograms, respectively).

We can express the SNR in these terms for a set of training patterns as (See Materials and Methods):

$$\frac{\text{Signal}}{\text{Noise}} = \frac{(W^T \Psi)^2}{\sum_i W_i^2 f_i^+ (1 - f_i^-) + \sum_i W_i^2 f_i^- (1 - f_i^+) - \sum_i W_i^2 (f_i^+ - f_i^-)^2} \quad (32)$$

The inner product between the DG-CA3 weight vector and the context-bias vector,  $W^T \Psi$ , determines the SNR between contexts. With neuronal turnover, the increase in absolute

weight (Figures 5A,B, side histograms), and absolute context-bias (Figures 5A,B, top histograms) results in increased inner product,  $W^T \Psi$ , for both the dense and sparse cases (Figure 5G), accounting for the increase in the SNR. However, the SNR grows more quickly in the sparse case (Figure 2G).

### 3.5. Extremely Sparse Coding Allows the Context-Bias of Individual Units to More Closely Determine the Output

We next address the dynamics with which the context-bias and weight vectors change as functions of each other. The purpose of this section is to give mathematical intuition for how neurogenesis takes advantage of sparse coding. In particular, we will discuss how the eigen-components of  $W$  and  $\Psi$  are interacting with each other in the dense coding case and sparse coding case. Note, as described above that the SNR is determined by the product of the weight vector,  $W$ , and the selectivity vector,  $\Psi$ . Furthermore, a DG cells synaptic weight determines its probability of survival. The weight vector is defined as:

$$W^T \stackrel{\text{def}}{=} \eta S^*, \quad (33)$$

where  $S^*$  is the pseudoinverse of the matrix of patterns in DG space. Using the Singular Value Decomposition (see Materials and Methods) we can re-express this in a way that allows us to intuitively understand the relationship between the context-bias vector and the weight vector. First we define  $\hat{\Psi}_i$  as the projection of the context-bias vector,  $\Psi$ , onto the respective  $i$ -th left singular projection matrix,  $u_i u_i^T$ .

$$\hat{\Psi}_i \stackrel{\text{def}}{=} u_i u_i^T \Psi \quad (34)$$

As noted above,  $\Psi$  is equivalent to the vector of mean separation between the contexts. Therefore, each vector  $\hat{\Psi}_i$  represents the separation between the context means in the direction of a given singular vector,  $u_i$ , which expresses the direction of the  $i$ th largest component of the activity patterns in DG space.  $\hat{\Psi}_i$  can be thought of as the contribution along the singular vector,  $u_i$ , to the mean separation between contexts,  $\Psi$ . Note that the singular vectors with large singular values represent the most important dimensions of the distribution of patterns in DG space.

Above, we noted that the two contexts separate from each other as neuronal turnover proceeds. Correspondingly,  $\|\Psi\|$ , the euclidean length of  $\Psi$ , increases over time (top histograms of Figures 5A,B, and a summary in Figure 5E). However the dense and sparse cases differ in the way dimensionality is reduced. To observe this we now express the weight vector in terms of  $\hat{\Psi}_i$  as:

$$W^T = 2P \sum_i^N \sigma_i^{-2} \hat{\Psi}_i^T, \quad (35)$$

where  $2P$  is a constant scale factor equal to twice the total number of patterns, and  $\sigma_i$  are the  $i$ -th singular values of the matrix  $S$ . The derivation can be found in Appendix B. We see that the weight vector is merely a weighted sum of  $\hat{\Psi}_i$ . Here a tradeoff emerges. Somewhat counterintuitively, the contributions of the

$\hat{\Psi}_i$  are scaled down by their respective  $\sigma_i^{-2}$ . Thus, though certain singular vectors may represent the mean separation between contexts, their contribution to the weight vector is limited by their singular values. In other words, the more a given  $\hat{\Psi}_i$  determines the mean separation, the more it is scaled down by its respective  $\sigma_i^{-2}$ .

In order to investigate the difference between dense coding and sparse coding cases, let us look into the distributions of singular values. In **Figure 5C**, the ranked reciprocals of the square of singular values,  $\sigma_i^{-2}$ , for different cases are presented. The scale-down effect in the dense case is more significant than in the sparse case for the ranks within a neighborhood of the rank 1. Thus, the weight vector in the dense case is subject to more shrinking of the  $\hat{\Psi}_i$  by their respective  $\sigma_i^{-2}$  and the elements of the weight vector have a narrower distribution in the dense case than in the sparse case prior to neurogenesis (**Figure 5D** for individual contributions and side histograms of **Figures 5A,B** for full distributions).

Due to the differences in scaling factors shown in panel C,  $\|W\|$  has a larger magnitude in the sparse case compared to the dense case despite that  $\|\Psi\|$  has smaller values, as shown in panel E. In addition to the difference in the magnitude, the normalized inner product  $(W^T\Psi) / (\|W\|\|\Psi\|)$  of the sparse case is larger than that of the dense case (**Figure 5F**), implying that the cosine distance between  $W$  and  $\Psi$  is smaller in the sparse case. In addition, neuronal turnover increases the inner product more rapidly in the sparse case (**Figure 5G**). Because  $W^T\Psi$  represents the degree of separation between the presentations of (+) context and (-) context, sparse coding is superior to dense coding in the context separation. This situation is schematically illustrated in **Figure 5H**.

### 3.6. The Neurogenesis Learning Rule Generalizes to Multiple Contexts

We next analyze patterns of activity in a model with multiple CA3 units to enable the encoding of an arbitrary number of distinct contexts. We use a similar neurogenesis rule in Model 2, in which the DG units compete for trophic signals, except now a DG neurons survival is determined by the sum of the absolute value of its output weights (**Figure 6A**, see section Materials and Methods) such that those neurons with a sum ranking in the bottom 30% of the population are turned over. In this case we have a weight matrix,  $W$  in which the elements of each column represents the DG-CA3 weights of a given output CA3 unit. We train the network with 8 contexts and test the network as before, by presenting a novel pattern,  $\mu$  but now we compare the pattern of CA3 activities represented in the vector  $\hat{\eta}^\mu = \text{sign}(W^T S)$  to the vector representing the target CA3 pattern specified by  $\eta^\mu$ . Requiring a match between these patterns for correct classification, we can then define the error for the  $\mu$ -th pattern at CA3 as  $\text{err}^\mu = \begin{cases} 0, & \text{if } \hat{\eta}^\mu = \eta^\mu \\ 1, & \text{otherwise} \end{cases}$ . The mean error

across test patterns decreases similarly to the generalization error of the two-context case, and again demonstrates the superiority of the sparse case with a coding level of  $f = 0.04$  (**Figure 6B**). One may notice that the less-sparse case with a coding level of

$f = 0.15$  has a similar performance level with  $f = 0.04$ . The setting with  $f = 0.15$  may be benefited from the increase in multiplicity in representations for this multiple-context case, *c.f.*, **Figure 1E**. However, the superiority of sparse coding still holds by comparing with the setting with  $f = 0.50$ .

Similar to the two-context case, the pairwise correlation of the training patterns in DG space demonstrates a clustering after neurogenesis (**Figures 6C,D**) in which patterns that are members of the same context tend to be correlated. PCA is used to observe the spread of the training patterns in DG space. The training patterns are initially randomly distributed in DG space (**Figure 6E**) but evolve into separated clusters with neuronal turnover (**Figure 6F**). To observe the effect of this separation on test patterns that the network has never seen before we project them onto the PCs of the DG representation of the training set, and mark any errors with a gray x (**Figures 6G,H**). Before neurogenesis, patterns of a given context are often misclassified due to the lack of separation between the contexts (**Figure 6G**). After neurogenesis, the separation between training patterns of the contexts (**Figure 6F**), reduces the probability of such errors on test patterns (**Figures 6B,H**).

### 3.7. A Model of Synaptic Turnover Achieves Similar Performance With Lower Material Cost

The models analyzed above assume that when a DG neuron has a weak connection to CA3, that neuron dies. However, the turnover rate that yields the best performance is about 30% of DG cells per day for 128 days of neuronal turnover (**Figures S1B,C**). We therefore explored a model assuming that biology seeks to conserve the material of synapses and neurons that might allow us to predict a realistic rate of neuronal turnover. In Model 3, as in the above models, the connections between DG and CA3 are trained with the pseudoinverse rule. Instead of neuronal turnover of units with weak DG-CA3 weights, we now implement synaptic turnover. A strong connection from the DG to CA3 results in a trophic signal that stabilizes that units EC-DG synapses, while a weak DG-CA3 weight is destabilizing (**Figure 7A**). We implement stability via the probability of EC-DG synaptic turnover. We assume a linear transfer function (See Materials and Methods, **Figure S3A**) between a DG units output weight to CA3 and the probability of that units input EC-DG weights being re-randomized, resulting in a random subset of that units EC-DG weights being chosen for re-randomization at each iteration. A slope of 2.5 was optimal in our simulations for the linear transfer function (**Figure S3B**). This rule results in similar improvement in performance to the prior rule that assumes that a fixed fraction of neurons turnover (**Figure 7B**). The same geometric intuition as the prior model applies (**Figure 7F**). The result as before is a reduced dimensionality of the contextual representations, such that reconstruction of an output weight vector that gives maximal cumulative performance (See Materials and Methods) can be achieved with far fewer dimensions (**Figure S3C**). Similar to Model 1 and 2, the optimal coding level becomes sparser with iterations of turnover (**Figures 7C,F**).

We then ask, what is the difference among different coding levels in terms of the cellular material turnover required to enable encoding? The total number of synapses turned over in this model is greatly reduced compared to the fixed turnover model, for all coding levels (**Figure 7D**). Since synaptic stability is thought to determine neuronal survival in several systems (Segal, 2010), including adult-born granule cells in the DG (Doengi et al., 2016), we made a similar assumption in the model to allow us to estimate the rate of neuronal turnover. We chose the conservative assumption that a neuron dies only if all of its synapses are targeted for turnover. With this assumption, the rate of neuronal turnover relative to the previous model drops by two orders of magnitude across all coding levels to range between 0.006 and 0.001 (**Figure 7E**), similar to the low rate of less than 1% that has been reported in rats (Cameron and McKay, 2001), 0.03–0.06% in the 2 month old mouse (Kempermann et al., 1997b), or 0.004% in humans (Spalding et al., 2013). Our results provide theoretical support to the findings that an extremely low rate of day-to-day neuronal turnover is sufficient to significantly alter memory performance.

The cumulative replacement of preexisting cells with newborn cells is also very low, ranging between 10–22% after 128 days of turnover across all coding levels (**Figure 7I**) similar to experimental results that have been previously reported in mice (Imayoshi et al., 2008). We see that for the same level of total synaptic or neural replacement, the cumulative error reduction is greater for the sparse case than for the dense case (**Figure 7G**), implying that sparse coding enables the learning rule to conserve on material turnover.

## 4. DISCUSSION

### 4.1. Neuronal Turnover in a Sparsely Active Dentate Gyrus

It is said to be paradoxical that the DG replenishes its neurons daily even though activity levels are very sparse on average (Piatti et al., 2013). Our results suggest that the sparseness of the DG is actually exploited by adult neurogenesis to find low-dimensional contextual representations that enhance memory encoding (**Figures 3C,D**). Placing synaptic turnover upstream of neuronal turnover performs similarly (**Figure 7**), suggesting that similar underlying processes could apply in other systems. As discussed below, such a model may help unify seemingly disparate findings in the neurogenesis literature.

Prior computational models of neurogenesis have implemented neuronal turnover by re-randomization (Chambers et al., 2004; Deisseroth et al., 2004; Becker, 2005; Crick and Miranker, 2006; Chambers and Conroy, 2007; Aimone et al., 2009; Finnegan and Becker, 2015), or by adding new neurons (Weisz and Argibay, 2012) with random synaptic weights. Here we contribute by explicitly addressing the interaction between sparseness and neurogenesis, and evaluating the consequences of a learning rule based on competition for target-derived stability.

The DG is significantly more sparse than most brain regions with a coding level estimated around 0.02–0.04 (Jung and McNaughton, 1993; Leutgeb et al., 2007; Danielson et al., 2016; Diamantaki et al., 2016). In our model, the optimal sparseness

for memory encoding evolves to a very sparse coding level as a function of the total amount of time over which the network has undergone encoding via neurogenesis (**Figure 1G**). This seems to suggest that the sparse code found in the DG may be tuned as such to make the best use of neuronal turnover in memory encoding - though we don't evaluate mechanisms of tuning sparseness, it could be accomplished on a multi-synaptic level such as by feedback inhibition, or by a homeostatic increase in firing threshold.

During neurogenesis, new neurons compete for synaptic contact (**Figure 1A**). As neurons compete and some replace others, the DG neuronal activities evolve to a low-dimensional representation of the two contexts that are to be learned (**Figure 3**). In this low-dimensional representation the activity-patterns representing the two contexts are grouped into distinct clusters representing the contexts (**Figures 3C,D, 6H**).

In a similar framework to ours it was known that there is a limit to how sparse a randomly connected network can be before a tradeoff emerges such that further sparseness actually impairs performance (Barak et al., 2013; Babadi and Sompolinsky, 2014). Babadi and Sompolinsky (2014) demonstrated analytically that the optimality of the sparse code is constrained by amplification of noise by random input weights that is mitigated when a hebbian learning rule is implemented on those weights. Given that hebbian learning structures the input weights to represent correlations among the inputs, they suggested that limitations on the effectiveness of sparse coding might emerge due to the unstructured nature of random weights. We first show that either neuronal (**Figure 1B**) or synaptic turnover (**Figure 7B**) improves the performance over the initial condition of random projection studied by Barak et al. (2013) and Babadi and Sompolinsky (2014). Furthermore, we demonstrate that a very sparse code can in fact be optimal even given random input weights (**Figures 1G, 7F**), implying that fine-tuning, such as the hebbian learning they employed (Babadi and Sompolinsky, 2014), is not always necessary at very sparse coding levels. Instead, via competition for target-derived stability, the sparse code facilitates the search for randomly connected neurons that collectively yield a low dimensional representation of the contextual inputs (**Figure 5H**).

Decomposing the CA3 weight vector allows us to see the higher correlation between the discriminative components,  $\hat{\Psi}_i$  and their contribution to the weight vector,  $\sigma_i^{-2}\hat{\Psi}_i$ , in the sparse case (**Figure 5F**). In other words, in the sparse case there exist discriminative components with singular values sufficiently small such that they can be strongly represented in the weight vector.

As a result, with each iteration (day), the synaptic strength of a DG neuron to CA3 can more readily grow in proportion to its contribution to the mean separation between contexts (**Figures 5E,F**). The overlap between these terms then scales up more quickly in the sparse case (**Figures 5E,G**). This greater coupling between the mean separation of contexts in the DG and the weights to CA3 (**Figure 5D**) thereby allows neurogenesis to more rapidly find separated contextual representations in the sparse case (**Figures 3C,D**). This greater separation allows the network to generalize better to new instances of the same context (**Figures 1E, 6B**).

## 4.2. Biological Predictions

The major prediction of this study is the dimensionality reduction of contextual codes in the dentate gyrus (DG). This prediction is in principle testable by recording the activity of a population of DG cells that includes both mature and immature neurons during contextual discrimination tasks. Then, analyses similar to those employed in the present study will be applicable to explore how the dimensionality of DG representation evolves during learning and how the dimensionality reduction is affected by the blockade of neurogenesis. Our results are also consistent with several experimental findings. Adult-born neurons are initially hyperexcitable, then gradually acquire the sparse firing characteristics of their mature counterparts (Schmidt-Hieber et al., 2004; Dieni et al., 2013). Correspondingly, input specificity increases with time (Marin-Burgin et al., 2012). This is consistent with the sparsening of the optimal coding level with time in our model (Figures 1G, 7F). Furthermore, if we assume that newborn DG cells initially have very few connections, greater hyperexcitability (higher coding level,  $f$ ) is necessary for optimal performance (Figures S2B,C).

The preference in our model for an average sparse coding level in the presence of neurogenesis (Figure 1F) is consistent with findings that neurogenesis induces a sparser code in the dentate gyrus (Ikhar et al., 2013) while blockade of neurogenesis results in increased average activity in the dentate gyrus (Burghardt et al., 2012; Lacefield et al., 2012). Meanwhile, increasing the excitability of the DG while neurogenesis is intact may impair contextual discrimination (Jinde et al., 2012).

The initial condition of our model, is equivalent to the encoding of novel contexts. As the contexts become familiar over time, the optimal neurogenesis rate decreases in the neuronal turnover model (Figure S1C), as does the predicted neuronal turnover in the synaptic turnover model (Figure 7E). This is consistent with experimental findings that novelty increases the neurogenesis rate (Kempermann et al., 1997b; Gould et al., 1999). Correspondingly, as the contextual encoding proceeds, their survival rate decreases with time, i.e., exceedingly few adult-born cells survive (Figure S1D). Therefore, relatively few mature cells are replaced and most of the cell death is replacement of immature cells by other immature cells. This is because a very old cell is already part of a favorable representation that enables discrimination and it is improbable to find a new cell that can better contribute. Thus newly adult-born cells have a survival advantage during novel encoding such as would occur during environmental enrichment, similar to what has been found experimentally (Kempermann et al., 1997b; Gould et al., 1999), while mature cells have the advantage under familiarity. Contextual novelty may explain why axonal retraction of mature DG cells results from a losing competition with adult-born cells in the juvenile rat (Yasuda et al., 2011), but not in adult mice in their home cage (Lopez et al., 2012). Since adults have already sufficiently encoded their environment, it is perhaps necessary to expose adults to enriched or novel environments (Kempermann et al., 1997b; Gould et al., 1999) to observe significant outcompeting of mature DG cells by new cells. However, this prediction in survival rate should not be confused

with the overall survival rate of all new-born dentate gyrus granule cells. The survival rate mentioned here considers only those dentate gyrus cells able to reach CA3 for competitions. For those newly generated dentate gyrus granule cells failed to emerge into the system, we consider that they are invisible in the model.

Our results are consistent with the presence of high-efficacy, so-called detonator synapses, at the Mossy Fiber (MF) terminals of DG axons to CA3 (McNaughton and Morris, 1987; Jonas et al., 1993; Treves and Rolls, 1994; Henze et al., 1997, 2002; Rollenhagen et al., 2007; Vyleta et al., 2016). The sparse activity of the DG causes the output weights to be larger than in less sparse systems, as the weights of sparse models are of greater magnitude for equivalent context-bias (Figure 5E). Furthermore, neuronal turnover during contextual learning leads to faster growth of the weights in the sparse model compared to those of the dense model (Figure 5E). This is consistent with the experimental finding that contextual learning increases the average synaptic efficacy of MF terminals of axons from the DG to CA3 (Galimberti et al., 2006).

## 4.3. Neuronal vs. Synaptic Turnover

It has been estimated that only around 0.03–0.09% of granule cells are turned over in the adult rodent DG (Kempermann et al., 1997b; Cameron and McKay, 2001), or 0.004% in humans (Spalding et al., 2013). These results have often raised the question - how can such a small number of cells significantly influence behavior (Piatti et al., 2013)? Indeed, there is a stark lack of consensus on whether adult hippocampal neurogenesis always positively correlates with DG-dependent learning (Frankland, 2013; Akers et al., 2014; Lipp and Bonfanti, 2016). Bats show no adult DG neurogenesis for the majority of species studied (Amrein, 2015), though bats clearly exhibit hippocampal place cells, and spatio-contextual reasoning that is attributed to the hippocampus (Finkelstein et al., 2016). Numerous comparative studies have demonstrated heterogeneous adult neurogenesis rates across mammalian species that does not seem to depend on their need for spatial reasoning (Cavegn et al., 2013; Amrein, 2015; van Dijk et al., 2016).

Experimental interventions that suggest a lack of positive correlation between neurogenesis rates and learning of DG-dependent tasks (Wood et al., 2001; Bartolomucci et al., 2002; Holmes et al., 2002; Bizon and Gallagher, 2003; Akirav et al., 2004; Leuner et al., 2004, 2006; Van der Borght et al., 2005), or that learning does not necessarily increase the number of new neurons (van Praag et al., 1999; Döbrössy et al., 2003; Ambrogini et al., 2004; Olariu et al., 2005; Pham et al., 2005; Snyder et al., 2005; Van der Borght et al., 2005), suggest that neuronal turnover is not always the relevant correlate of learning in the DG. Substantial evidence that depletion of neurogenesis does not impair such learning (Shors et al., 2001, 2002; Madsen et al., 2003; Raber et al., 2004; Snyder et al., 2005; Meshi et al., 2006; Frankland, 2013; Groves et al., 2013; Urbach et al., 2013) suggests that molecular mechanisms modulating DG synaptic processes can remain intact and support learning, without requiring neuronal turnover.

Placing synaptic turnover upstream of somatic turnover, as in Model 3 (Figure 7A), may help unify these findings. Synaptic turnover, rather than neuronal turnover may be the relevant measurement with which to correlate DG-dependent learning that is targetable by molecular and cellular interventions in the neurogenic niche. DG neurons compete for CA3 target factors, and those losing the competition have their input synapses destabilized (Figure 7A). If the amount of synaptic destabilization crosses a threshold (in our case, all input synapses destabilized) then the neuron dies. With these assumptions, we indeed find a very low optimal neurogenesis rate (Figure 7E) in the biologically reported range of a fraction of a percent (Kempermann et al., 1997b; Cameron and McKay, 2001; Spalding et al., 2013). This suggests that, via the same form of competition, en masse synaptic turnover could underlie learning, while only a minority of neurons actually turn over. Such a synaptic-turnover-driven neuronal turnover rule is consistent with evidence that activity-dependent competition among mature and immature DG granule cells for CA3 targets (Yasuda et al., 2011), and their input-synaptic stability (Tashiro et al., 2006; Doengi et al., 2016) appears to promote neuronal survival. Furthermore, there is a well-known overlap between factors that influence synaptic plasticity, and those that influence neurogenesis in the DG (Vivar et al., 2013), and many of these same factors influence synaptic stability more generally throughout the central nervous system (Vicario-Abejón et al., 2002). Future behavioral studies in animal models of modulated neurogenesis may benefit from measuring markers of synaptic stability, such as adhesion molecules required for synapse maintenance (Doengi et al., 2016), rather than somatic markers of neurogenesis.

#### 4.4. Concluding Remarks

Sparse coding is prevalent throughout many systems of the brain (Barak et al., 2013; Babadi and Sompolinsky, 2014). Our results suggest that neuronal or synaptic turnover in sparsely active regions of the brain may embody a novel learning rule that enhances the clustering of associated activity patterns, and thereby memory encoding and retrieval. Sparseness entails a lower metabolic cost since few neurons are active at any time, and our results further suggest that learning in a sparse layer via turnover conserves synaptic (Figure 7D) or

somatic material (Figure 7E), perhaps a previously unrecognized metabolic benefit to sparse coding. The learning curves of all implemented models suggest that differing degrees of sparseness across systems may be found to correspond to the timescale over which they are required to represent memories. Since the optimal sparseness of these models increases as a function of encoding time, we might think of the high sparseness of the DG as being tuned to enable retrieval of episodes that are encoded over long periods of time. Consistent with this timescale, amnesiac patient H.M. lost not only the ability to encode novel information, but also the ability to retrieve memories up to 11 years prior to the removal of his hippocampus (Corkin, 2002). Further investigation of the relationship between synaptic stability and neuronal survival (Doengi et al., 2016) may yield insight into how neuronal turnover and synaptic turnover are coupled. Our work, and that of others (Marin-Burgin et al., 2012; Bergami et al., 2015; Alvarez et al., 2016) suggests that local regulation of sparse activity in the DG may be critical during the addition of new synapses or new neurons that occurs during learning. Similar processes may regulate brain development in general.

#### AUTHOR CONTRIBUTIONS

AD conceived the project and performed mathematical analysis and numerical simulations, and co-wrote the manuscript. CF performed mathematical analysis and numerical simulations, and co-wrote the manuscript. TF supervised the project and co-wrote the manuscript.

#### FUNDING

This work was partly supported by Grants-in-Aid for Scientific Research, KAKENHI (No. 17H06030) from MEXT and CREST (JPMJCR13W1), Japan Science and Technology Agency.

#### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fncom.2018.00099/full#supplementary-material>

#### REFERENCES

- Aimone, J. B., Wiles, J., and Gage, F. H. (2009). Computational influence of adult neurogenesis on memory encoding. *Neuron* 61, 187–202. doi: 10.1016/j.neuron.2008.11.026
- Akers, K. G., Martinez-Canabal, A., Restivo, L., Yiu, A. P., De Cristofaro, A., Hsiang, H.-L. L., et al. (2014). Hippocampal neurogenesis regulates forgetting during adulthood and infancy. *Science (New York, N.Y.)* 344, 598–602. doi: 10.1126/science.1248903
- Akirav, I., Kozenicky, M., Tal, D., Sandi, C., Venero, C., and Richter-Levin, G. (2004). A facilitative role for corticosterone in the acquisition of a spatial task under moderate stress. *Learn. Mem. (Cold Spring Harbor, N.Y.)* 11, 188–195. doi: 10.1101/lm.61704
- Alvarez, D. D., Giacomini, D., Yang, S. M., Trincherro, M. F., Temprana, S. G., Buttner, K. A., et al. (2016). A disinaptic feedback network activated by experience promotes the integration of new granule cells. *Science* 354, 459–465. doi: 10.1126/science.aaf2156
- Ambrogini, P., Orsini, L., Mancini, C., Ferri, P., Ciaroni, S., and Cuppini, R. (2004). Learning may reduce neurogenesis in adult rat dentate gyrus. *Neurosci. Lett.* 359, 13–16. doi: 10.1016/j.neulet.2003.12.123
- Amit, D. J., and Fusi, S. (1994). Learning in neural networks with material synapses. *Neural Comput.* 6, 957–982. doi: 10.1162/neco.1994.6.5.957
- Amrein, I. (2015). Adult hippocampal neurogenesis in natural populations of mammals. *Cold Spring Harb. Perspect. Biol.* 7:a021295. doi: 10.1101/cshperspect.a021295
- Babadi, B., and Sompolinsky, H. (2014). Sparseness and expansion in sensory representations. *Neuron* 83, 1213–1226. doi: 10.1016/j.neuron.2014.07.035
- Barak, O., Rigotti, M., and Fusi, S. (2013). The sparseness of mixed selectivity neurons controls the generalization-discrimination trade-off. *J. Neurosci.* 33, 3844–3856. doi: 10.1523/JNEUROSCI.2753-12.2013

- Barlow, H. B. (1961). "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, ed W. A. Rosenblith (Cambridge, MA: MIT Press), 217–234.
- Bartolomucci, A., de Biurrun, G., Czéh, B., van Kampen, M., and Fuchs, E. (2002). Selective enhancement of spatial learning under chronic psychosocial stress. *Eur. J. Neurosci.* 15, 1863–1866. doi: 10.1046/j.1460-9568.2002.02043.x
- Becker, S. (2005). A computational principle for hippocampal learning and neurogenesis. *Hippocampus* 15, 722–738. doi: 10.1002/hipo.20095
- Bell, A. J., and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vis. Res.* 37, 3327–3338. doi: 10.1016/S0042-6989(97)00121-1
- Bergami, M., and Berninger, B. (2012). A fight for survival: the challenges faced by a newborn neuron integrating in the adult hippocampus. *Dev. Neurobiol.* 72, 1016–1031. doi: 10.1002/dneu.22025
- Bergami, M., Masserdotti, G., Temprana, S. G., Motori, E., Eriksson, T. M., Göbel, J., et al. (2015). A critical period for experience-dependent remodeling of adult-born neuron connectivity. *Neuron* 85, 710–717. doi: 10.1016/j.neuron.2015.01.001
- Bizon, J. L., and Gallagher, M. (2003). Production of new cells in the rat dentate gyrus over the lifespan: relation to cognitive decline. *Eur. J. Neurosci.* 18, 215–219. doi: 10.1046/j.1460-9568.2003.02733.x
- Burghardt, N. S., Park, E. H., Hen, R., and Fenton, A. A. (2012). Adult-born hippocampal neurons promote cognitive flexibility in mice. *Hippocampus* 22, 1795–1808. doi: 10.1002/hipo.22013
- Cameron, H. A., and McKay, R. D. (2001). Adult neurogenesis produces a large pool of new granule cells in the dentate gyrus. *J. Compar. Neurol.* 435, 406–417. doi: 10.1002/cne.1040
- Cavegn, N., van Dijk, R. M., Menges, D., Bretschneider, H., Phalanndwa, M., Chimimba, C. T., et al. (2013). Habitat-specific shaping of proliferation and neuronal differentiation in adult hippocampal neurogenesis of wild rodents. *Front. Neurosci.* 7:59. doi: 10.3389/fnins.2013.00059
- Chambers, R. A., and Conroy, S. K. (2007). Network modeling of adult neurogenesis: shifting rates of neuronal turnover optimally gears network learning according to novelty gradient. *J. Cogn. Neurosci.* 19, 1–12. doi: 10.1162/jocn.2007.19.1.1
- Chambers, R. A., Potenza, M. N., Hoffman, R. E., and Miranker, W. (2004). Simulated apoptosis/neurogenesis regulates learning and memory capabilities of adaptive neural networks. *Neuropsychopharmacology* 29, 747–758. doi: 10.1038/sj.npp.1300358
- Clelland, C. D., Choi, M., Romberg, C., Clemenson, G. D., Fragniere, A., Tyers, P., et al. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation. *Science (New York, N.Y.)* 325, 210–213. doi: 10.1126/science.1173215
- Corkin, S. (2002). What's new with the amnesic patient H.M.? *Nat. Rev. Neurosci.* 3, 153–160. doi: 10.1038/nrn726
- Crick, C., and Miranker, W. (2006). Apoptosis, neurogenesis, and information content in Hebbian networks. *Biol. Cybern.* 94, 9–19. doi: 10.1007/s00422-005-0026-8
- Danielson, N. B., Kaifosh, P., Zaremba, J. D., Lovett-Barron, M., Tsai, J., Denny, C. A., et al. (2016). Distinct contribution of adult-born hippocampal granule cells to context encoding. *Neuron* 90, 101–112. doi: 10.1016/j.neuron.2016.02.019
- Deisseroth, K., Singla, S., Toda, H., Monje, M., Palmer, T. D., and Malenka, R. C. (2004). Excitation-neurogenesis coupling in adult neural stem/progenitor cells. *Neuron* 42, 535–552. doi: 10.1016/S0896-6273(04)00266-1
- Diamantaki, M., Frey, M., Berens, P., Preston-Ferrer, P., and Burgalossi, A. (2016). Sparse activity of identified dentate granule cells during spatial exploration. *eLife* 5:e20252. doi: 10.7554/eLife.20252
- Dieni, C. V., Nietz, A. K., Panichi, R., Wadiche, J. I., and Overstreet-Wadiche, L. (2013). Distinct determinants of sparse activation during granule cell maturation. *J. Neurosci.* 33, 19131–19142. doi: 10.1523/JNEUROSCI.2289-13.2013
- Döbrössy, M. D., Drapeau, E., Aurousseau, C., Le Moal, M., Piazza, P. V., and Abrous, D. N. (2003). Differential effects of learning on neurogenesis: learning increases or decreases the number of newly born cells depending on their birth date. *Mol. Psychiatry* 8, 974–982. doi: 10.1038/sj.mp.4010419
- Doengi, M., Krupp, A. J., Körber, N., and Stein, V. (2016). SynCAM 1 improves survival of adult-born neurons by accelerating synapse maturation. *Hippocampus* 26, 319–328. doi: 10.1002/hipo.22524
- Du, J.-L., and Poo, M.-m. (2004). Rapid BDNF-induced retrograde synaptic modification in a developing retinotectal system. *Nature* 429, 878–883. doi: 10.1038/nature02618
- Finkelstein, A., Las, L., and Ulanovsky, N. (2016). 3-D maps and compasses in the brain. *Annu. Rev. Neurosci.* 39, 171–196. doi: 10.1146/annurev-neuro-070815-013831
- Finnegan, R., and Becker, S. (2015). Neurogenesis paradoxically decreases both pattern separation and memory interference. *Front. Syst. Neurosci.* 9:136. doi: 10.3389/fnsys.2015.00136
- Fitzsimonds, R. M., Song, H. J., and Poo, M. M. (1997). Propagation of activity-dependent synaptic depression in simple neural networks. *Nature* 388, 439–448. doi: 10.1038/41267
- Frankland, P. W. (2013). Neurogenic evangelism: comment on Urbach et al. (2013). *Behav. Neurosci.* 127, 126–129. doi: 10.1037/a0031512
- Galimberti, I., Gogolla, N., Alberi, S., Santos, A. F., Müller, D., and Caroni, P. (2006). Long-term rearrangements of hippocampal mossy fiber terminal connectivity in the adult regulated by experience. *Neuron* 50, 749–763. doi: 10.1016/j.neuron.2006.04.026
- Gonçalves, J. T., Schafer, S. T., and Gage, F. H. (2016). Adult neurogenesis in the hippocampus: from stem cells to behavior. *Cell* 167, 897–914. doi: 10.1016/j.cell.2016.10.021
- Gould, E., Beylin, A., Tanapat, P., Reeves, A., and Shors, T. J. (1999). Learning enhances adult neurogenesis in the hippocampal formation. *Nat. Neurosci.* 2, 260–265. doi: 10.1038/6365
- Groves, J. O., Leslie, I., Huang, G.-J., McHugh, S. B., Taylor, A., Mott, R., et al. (2013). Ablating adult neurogenesis in the rat has no effect on spatial processing: evidence from a novel pharmacogenetic model. *PLoS Genet.* 9:e1003718. doi: 10.1371/journal.pgen.1003718
- Gu, Y., Arruda-Carvalho, M., Wang, J., Janoschka, S. R., Josselyn, S. A., Frankland, P. W., et al. (2012). Optical controlling reveals time-dependent roles for adult-born dentate granule cells. *Nat. Neurosci.* 15, 1700–1706. doi: 10.1038/nn.3260
- Harris, K. D. (2008). Stability of the fittest: organizing learning through retroaxonal signals. *Trends Neurosci.* 31, 130–136. doi: 10.1016/j.tins.2007.12.002
- Henze, D. A., Card, J. P., Barrionuevo, G., and Ben-Ari, Y. (1997). Large amplitude miniature excitatory postsynaptic currents in hippocampal CA3 pyramidal neurons are of mossy fiber origin. *J. Neurophysiol.* 77, 1075–1086. doi: 10.1152/jn.1997.77.3.1075
- Henze, D. A., Wittner, L., and Buzsáki, G. (2002). Single granule cells reliably discharge targets in the hippocampal CA3 network *in vivo*. *Nat. Neurosci.* 5, 790–795. doi: 10.1038/nn887
- Holmes, M. M., Wide, J. K., and Galea, L. A. M. (2002). Low levels of estradiol facilitate, whereas high levels of estradiol impair, working memory performance on the radial arm maze. *Behav. Neurosci.* 116, 928–934. doi: 10.1037/0735-7044.116.5.928
- Ikrar, T., Guo, N., He, K., Besnard, A., Levinson, S., Hill, A., et al. (2013). Adult neurogenesis modifies excitability of the dentate gyrus. *Front. Neural Circuits* 7:204. doi: 10.3389/fncir.2013.00204
- Imayoshi, I., Sakamoto, M., Ohtsuka, T., Takao, K., Miyakawa, T., Yamaguchi, M., et al. (2008). Roles of continuous neurogenesis in the structural and functional integrity of the adult forebrain. *Nat. Neurosci.* 11, 1153–1161. doi: 10.1038/nn.2185
- Jinde, S., Zsiros, V., Jiang, Z., Nakao, K., Pickel, J., Kohno, K., et al. (2012). Hilar mossy cell degeneration causes transient dentate granule cell hyperexcitability and impaired pattern separation. *Neuron* 76, 1189–1200. doi: 10.1016/j.neuron.2012.10.036
- Jonas, P., Major, G., and Sakmann, B. (1993). Quantal components of unitary EPSCs at the mossy fibre synapse on CA3 pyramidal cells of rat hippocampus. *J. Physiol.* 472, 615–663. doi: 10.1113/jphysiol.1993.sp019965
- Josselyn, S. A., Köhler, S., and Frankland, P. W. (2015). Finding the engram. *Nat. Rev. Neurosci.* 16, 521–534. doi: 10.1038/nrn4000
- Jung, M. W., and McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer. *Hippocampus* 3, 165–182. doi: 10.1002/hipo.450030209

- Kempermann, G., Kuhn, H. G., and Gage, F. H. (1997a). Genetic influence on neurogenesis in the dentate gyrus of adult mice. *Proc. Natl. Acad. Sci. U.S.A.* 94, 10409–10414.
- Kempermann, G., Kuhn, H. G., and Gage, F. H. (1997b). More hippocampal neurons in adult mice living in an enriched environment. *Nature* 386, 493–495.
- Lacefield, C. O., Itskov, V., Reardon, T., Hen, R., and Gordon, J. A. (2012). Effects of adult-generated granule cells on coordinated network activity in the dentate gyrus. *Hippocampus* 22, 106–116. doi: 10.1002/hipo.20860
- Leuner, B., Gould, E., and Shors, T. J. (2006). Is there a link between adult neurogenesis and learning? *Hippocampus* 16, 216–224. doi: 10.1002/hipo.20153
- Leuner, B., Mendolia-Loffredo, S., and Shors, T. J. (2004). Males and females respond differently to controllability and antidepressant treatment. *Biol. Psychiatry* 56, 964–970. doi: 10.1016/j.biopsych.2004.09.018
- Leutgeb, J. K., Leutgeb, S., Moser, M.-B., and Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science (New York, N.Y.)* 315, 961–966. doi: 10.1126/science.1135801
- Lipp, H.-P., and Bonfanti, L. (2016). Adult neurogenesis in mammals: variations and confusions. *Brain Behav. Evol.* 87, 205–221. doi: 10.1159/000446905
- Lopez, C. M., Pelkey, K. A., Chittajallu, R., Nakashiba, T., Tóth, K., Tonegawa, S., et al. (2012). Competition from newborn granule cells does not drive axonal retraction of silenced old granule cells in the adult hippocampus. *Front. Neural Circuits* 6:85. doi: 10.3389/fncir.2012.00085
- Madsen, T. M., Kristjansen, P. E. G., Bolwig, T. G., and Wörtwein, G. (2003). Arrested neuronal proliferation and impaired hippocampal function following fractionated brain irradiation in the adult rat. *Neuroscience* 119, 635–642. doi: 10.1016/S0306-4522(03)00199-4
- Marin-Burgin, A., Mongiat, L. A., Pardi, M. B., and Schinder, A. F. (2012). Unique processing during a period of high excitation/inhibition balance in adult-born neurons. *Science* 335, 1238–1242. doi: 10.1126/science.1214956
- McNaughton, B., and Morris, R. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* 10, 408–415. doi: 10.1016/0166-2236(87)90011-7
- Meshi, D., Drew, M. R., Saxe, M., Ansorge, M. S., David, D., Santarelli, L., et al. (2006). Hippocampal neurogenesis is not required for behavioral effects of environmental enrichment. *Nat. Neurosci.* 9, 729–731. doi: 10.1038/nrn1696
- Nakashiba, T., Cushman, J. D., Pelkey, K. A., Renaudineau, S., Buhl, D. L., McHugh, T. J., et al. (2012). Young dentate granule cells mediate pattern separation, whereas old granule cells facilitate pattern completion. *Cell* 149, 188–201. doi: 10.1016/j.cell.2012.01.046
- Olariu, A., Cleaver, K. M., Shore, L. E., Brewer, M. D., and Cameron, H. A. (2005). A natural form of learning can increase and decrease the survival of new neurons in the dentate gyrus. *Hippocampus* 15, 750–762. doi: 10.1002/hipo.20097
- Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0
- Pham, K., McEwen, B., Ledoux, J., and Nader, K. (2005). Fear learning transiently impairs hippocampal cell proliferation. *Neuroscience* 130, 17–24. doi: 10.1016/j.neuroscience.2004.09.015
- Piatti, V. C., Ewell, L. A., and Leutgeb, J. K. (2013). Neurogenesis in the dentate gyrus: carrying the message or dictating the tone. *Front. Neurosci.* 7:50. doi: 10.3389/fnins.2013.00050
- Raber, J., Rola, R., LeFevour, A., Morhardt, D., Curley, J., Mizumatsu, S., et al. (2004). Radiation-induced cognitive impairments are associated with changes in indicators of hippocampal neurogenesis. *Radiat. Res.* 162, 39–47. doi: 10.1667/RR3206
- Rollenhagen, A., Sätzler, K., Rodríguez, E. P., Jonas, P., Frotscher, M., and Lübke, J. H. R. (2007). Structural determinants of transmission at large hippocampal mossy fiber synapses. *J. Neurosci.* 27, 10434–10444. doi: 10.1523/JNEUROSCI.1946-07.2007
- Sahay, A., Scobie, K. N., Hill, A. S., O'Carroll, C. M., Kheirbek, M. A., Burghardt, N. S., et al. (2011). Increasing adult hippocampal neurogenesis is sufficient to improve pattern separation. *Nature* 472, 466–470. doi: 10.1038/nature09817
- Schmidt-Hieber, C., Jonas, P., and Bischofberger, J. (2004). Enhanced synaptic plasticity in newly generated granule cells of the adult hippocampus. *Nature* 429, 184–187. doi: 10.1038/nature02553
- Segal, M. (2010). Dendritic spines, synaptic plasticity and neuronal survival: activity shapes dendritic spines to enhance neuronal viability. *Eur. J. Neurosci.* 31, 2178–2184. doi: 10.1111/j.1460-9568.2010.07270.x
- Sharma, N., Deppmann, C. D., Harrington, A. W., St. Hillaire, C., Chen, Z.-Y., Lee, F. S., et al. (2010). Long-distance control of synapse assembly by target-derived NGF. *Neuron* 67, 422–434. doi: 10.1016/j.neuron.2010.07.018
- Shors, T. J., Miesegaes, G., Beylin, A., Zhao, M., Rydel, T., and Gould, E. (2001). Neurogenesis in the adult is involved in the formation of trace memories. *Nature* 410, 372–376. doi: 10.1038/35066584
- Shors, T. J., Townsend, D. A., Zhao, M., Kozorovitskiy, Y., and Gould, E. (2002). Neurogenesis may relate to some but not all types of hippocampal-dependent learning. *Hippocampus* 12, 578–584. doi: 10.1002/hipo.10103
- Snyder, J., Hong, N., McDonald, R., and Wojtowicz, J. (2005). A role for adult neurogenesis in spatial long-term memory. *Neuroscience* 130, 843–852. doi: 10.1016/j.neuroscience.2004.10.009
- Spalding, K. L., Bergmann, O., Alkass, K., Bernard, S., Salehpour, M., Huttner, H. B., et al. (2013). Dynamics of hippocampal neurogenesis in adult humans. *Cell* 153, 1219–1227. doi: 10.1016/j.cell.2013.05.002
- Tao, H., Zhang, L. I., Bi, G., and Poo, M. (2000). Selective presynaptic propagation of long-term potentiation in defined neural networks. *J. Neurosci.* 20, 3233–3243. doi: 10.1523/JNEUROSCI.20-09-03233.2000
- Tashiro, A., Sandler, V. M., Toni, N., Zhao, C., and Gage, F. H. (2006). NMDA-receptor-mediated, cell-specific integration of new neurons in adult dentate gyrus. *Nature* 442, 929–933. doi: 10.1038/nature05028
- Tonegawa, S., Liu, X., Ramirez, S., and Redondo, R. (2015). Memory engram cells have come of age. *Neuron* 87, 918–931. doi: 10.1016/j.neuron.2015.08.002
- Toni, N., Laplagne, D. A., Zhao, C., Lombardi, G., Ribak, C. E., Gage, F. H., et al. (2008). Neurons born in the adult dentate gyrus form functional synapses with target cells. *Nat. Neurosci.* 11, 901–907. doi: 10.1038/nn.2156
- Treves, A., and Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus* 4, 374–391. doi: 10.1002/hipo.450040319
- Tsodyks, M. V., and Feigel'man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.* 6, 101–105. doi: 10.1209/0295-5075/6/2/002
- Urbach, A., Robakiewicz, I., Baum, E., Kaczmarek, L., Witte, O. W., and Filipkowski, R. K. (2013). Cyclin D2 knockout mice with depleted adult neurogenesis learn Barnes maze task. *Behav. Neurosci.* 127, 1–8. doi: 10.1037/a0031222
- Van der Borght, K., Wallinga, A. E., Luiten, P. G. M., Eggen, B. J. L., and Van der Zee, E. A. (2005). Morris water maze learning in two rat strains increases the expression of the polysialylated form of the neural cell adhesion molecule in the dentate gyrus but has no effect on hippocampal neurogenesis. *Behav. Neurosci.* 119, 926–932. doi: 10.1037/0735-7044.119.4.926
- van Dijk, R. M., Huang, S.-H., Slomianka, L., and Amrein, I. (2016). Taxonomic separation of hippocampal networks: principal cell populations and adult neurogenesis. *Front. Neuroanat.* 10:22. doi: 10.3389/fnana.2016.00022
- van Praag, H., Christie, B. R., Sejnowski, T. J., and Gage, F. H. (1999). Running enhances neurogenesis, learning, and long-term potentiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13427–13431. doi: 10.1073/pnas.96.23.13427
- Vicario-Abejón, C., Owens, D., McKay, R., and Segal, M. (2002). Role of neurotrophins in central synapse formation and stabilization. *Nat. Rev. Neurosci.* 3, 965–974. doi: 10.1038/nrn988
- Vivar, C., Potter, M. C., and van Praag, H. (2013). All about running: synaptic plasticity, growth factors and adult hippocampal neurogenesis. *Curr. Top. Behav. Neurosci.* 15, 189–210. doi: 10.1007/7854\_2012\_220
- Vyleta, N. P., Borges-Merjane, C., and Jonas, P. (2016). Plasticity-dependent, full detonation at hippocampal mossy fiber-CA3 pyramidal neuron synapses. *eLife* 5:e17977. doi: 10.7554/eLife.17977
- Weisz, V. I., and Argibay, P. F. (2012). Neurogenesis interferes with the retrieval of remote memories: forgetting in neurocomputational terms. *Cognition* 125, 13–25. doi: 10.1016/j.cognition.2012.07.002

- Wood, G. E., Beylin, A. V., and Shors, T. J. (2001). The contribution of adrenal and reproductive hormones to the opposing effects of stress on trace conditioning in males versus females. *Behav. Neurosci.* 115, 175–187. doi: 10.1037/0735-7044.115.1.175
- Wu, M. V., Sahay, A., Duman, R. S., and Hen, R. (2015). Functional differentiation of adult-born neurons along the septotemporal axis of the dentate gyrus: figure 1. *Cold Spring Harb. Perspect. Biol.* 7:a018978. doi: 10.1101/cshperspect.a018978
- Yasuda, M., Johnson-Venkatesh, E. M., Zhang, H., Parent, J. M., Sutton, M. A., and Umemori, H. (2011). Multiple forms of activity-dependent competition refine hippocampal circuits *in vivo*. *Neuron* 70, 1128–1142. doi: 10.1016/j.neuron.2011.04.027
- Zhou, B., Cai, Q., Xie, Y., and Sheng, Z.-H. (2012). Snapin recruits dynein to BDNF-TrkB signaling endosomes for retrograde axonal transport and is essential for dendrite growth of cortical neurons. *Cell Rep.* 2, 42–51. doi: 10.1016/j.celrep.2012.06.010
- Zhuo, J.-M., Tseng, H.-A., Desai, M., Bucklin, M. E., Mohammed, A. I., Robinson, N. T., et al. (2016). Young adult born neurons enhance hippocampal dependent performance via influences on bilateral networks. *eLife* 5:e22429. doi: 10.7554/eLife.22429

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 DeCostanzo, Fung and Fukai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDICES

### A. MOORE PENROSE PSEUDOINVERSE GUARANTEES A LEAST-SQUARE SOLUTION

Let us consider a linear system:

$$y = Ax. \quad (\text{A1})$$

For a given  $y$  and a given  $A$ , we would like to look for a solution to  $x$  minimizing the square residual given by

$$\|r\|^2 \equiv \|y - Ax\|^2 \quad (\text{A2})$$

$$= y^T y - y^T A x - x^T A^T y + x^T A^T A x. \quad (\text{A3})$$

The gradient of the residual is

$$\nabla_x \|r\|^2 = -2A^T y + 2A^T A x. \quad (\text{A4})$$

$\nabla_x \|r\|^2 = 0$  implies

$$A^T y = A^T A x. \quad (\text{A5})$$

Suppose  $A$  is a matrix with linearly independent columns, we have

$$x = (A^T A)^{-1} A^T y. \quad (\text{A6})$$

Here  $A^* \equiv (A^T A)^{-1} A^T$  is the Moore-Penrose pseudoinverse of  $A$ . The solution to  $x$  deduced by the Moore-Penrose pseudoinverse should be guaranteed to be the least-square solution.

In our model, the output weights is solved by Moore-Penrose inverse. One may consider replacing  $\eta_T$  with  $y$ ,  $\bar{S}^T$  with  $A$  and  $W$  with  $x$  in the linear system here. Since  $\bar{S}$  is generated from linearly independent input and random input weights, and  $\theta$  is chosen to match the coding level  $f$ , the rows of the matrix  $\bar{S}$  should also be linearly independent. The argument concerning Moore-Penrose pseudoinverse in this appendix follows.

### B. EXPRESSION OF $W$ IN TERMS OF $\hat{\Psi}$

Given that

$$W^T = \eta \bar{S}^* \quad (\text{A7})$$

$$= \eta V \Sigma^* U^T, \quad (\text{A8})$$

where matrices  $U$ ,  $V$  and  $\Sigma$  are the matrix given by singular-value decomposition. Note that singular-value decomposition is a generalized eigendecomposition. Diagonal entries of  $\Sigma$  stores eigenvalues of the matrix  $\bar{S}^*$ , while matrices  $V$  and  $U$  store left-singular and right-singular vectors. On the other hand,

$$\eta^T = 2P \bar{S}^* \Psi \quad (\text{A9})$$

$$\eta = 2P \Psi^T (\bar{S}^*)^T. \quad (\text{A10})$$

Therefore,

$$W^T = \left[ 2P \Psi^T (\bar{S}^*)^T \right] V \Sigma^* U^T \quad (\text{A11})$$

$$= 2P \Psi^T U (\Sigma^*)^2 U^T. \quad (\text{A12})$$

Let us consider the entries of  $W^T$ ,

$$(W^T)_j = 2P \sum_k (\Psi^T)_k \sum_l U_{kl} \sum_m (\Sigma^{-2})_{lm} (U^T)_{mj} \quad (\text{A13})$$

$$= 2P \sum_k (\Psi^T)_k \sum_l (\Sigma^{-2})_{ll} U_{kl} (U^T)_{lj} \quad (\text{A14})$$

$$= 2P \sum_l (\Sigma^{-2})_{ll} \sum_k (\Psi^T)_k U_{kl} (U^T)_{lj}. \quad (\text{A15})$$

Therefore, the matrix-vector form is given by

$$W^T = 2P \sum_i \sigma_i^{-2} \hat{\Psi}_i, \quad (\text{A16})$$

where  $\hat{\Psi} \equiv \Psi^T u_i u_i^T$  and  $u_i$  is the  $i$ th column of the matrix  $U$ .