CrossMark

Structural Plasticity, Effectual Connectivity, and Memory in Cortex

Andreas Knoblauch[1]* and Friedrich T. Sommer[2]

[1] Informatics Faculty, Albstadt-Sigmaringen University, Albstadt, Germany, [2] Redwood Center for Theoretical Neuroscience, University of California at Berkeley, Berkeley, CA, USA

Learning and memory is commonly attributed to the modification of synaptic strengths in neuronal networks. More recent experiments have also revealed a major role of structural plasticity including elimination and regeneration of synapses, growth and retraction of dendritic spines, and remodeling of axons and dendrites. Here we work out the idea that one likely function of structural plasticity is to increase "effectual connectivity" in order to improve the capacity of sparsely connected networks to store Hebbian cell assemblies that are supposed to represent memories. For this we define effectual connectivity as the fraction of synaptically linked neuron pairs within a cell assembly representing a memory. We show by theory and numerical simulation the close links between effectual connectivity and both information storage capacity of neural networks and effective connectivity as commonly employed in functional brain imaging and connectome analysis. Then, by applying our model to a recently proposed memory model, we can give improved estimates on the number of cell assemblies that can be stored in a cortical macrocolumn assuming realistic connectivity. Finally, we derive a simplified model of structural plasticity to enable large scale simulation of memory phenomena, and apply our model to link ongoing adult structural plasticity to recent behavioral data on the spacing effect of learning.

Keywords: synaptic plasticity, effective connectivity, transfer entropy, learning, potential synapse, memory consolidation, storage capacity, spacing effect

1. INTRODUCTION

Traditional theories attribute adult learning and memory to Hebbian modification of synaptic weights (Hebb, 1949; Bliss and Collingridge, 1993; Paulsen and Sejnowski, 2000; Song et al., 2000), whereas recent evidence suggests also a role for network rewiring by structural plasticity including generation of synapses, growth and retraction of spines, and remodeling of dendritic and axonal branches, both during development and adulthood (Raisman, 1969; Witte et al., 1996; Engert and Bonhoeffer, 1999; Chklovskii et al., 2004; Butz et al., 2009; Holtmaat and Svoboda, 2009; Xu et al., 2009; Yang et al., 2009; Fu and Zuo, 2011; Yu and Zuo, 2011). One possible function of structural plasticity is effective information storage, both in terms of space and energy requirements (Poirazi and Mel, 2001; Chklovskii et al., 2004; Knoblauch et al., 2010). Indeed, due to space and energy limitations, neural networks in the brain are only sparsely connected, even on a local scale (Abeles, 1991; Braitenberg and Schüz, 1991; Hellwig, 2000). Moreover, it is believed that the energy consumption of the brain is dominated by the number of postsynaptic potentials or, equivalently, the number of functional non-silent synapses (Attwell and Laughlin, 2001; Laughlin and Sejnowski, 2003; Lennie, 2003). Together this implies a pressure to minimize the number

and density of functional (non-silent) synapses. It has therefore been suggested that the function of structural plasticity "moves" the rare expensive synapses to the most useful locations, while keeping the mean number of synapses on a constant low level (Knoblauch et al., 2014). By this, sparsely connected networks can have computational abilities that are equivalent to densely connected networks. For example, it is known that memory storage capacity of neural associative networks scales with the synaptic density, such that networks with a high connectivity can store many more memories than networks with a low connectivity (Buckingham and Willshaw, 1993; Bosch and Kurfess, 1998; Knoblauch, 2011). For modeling structural plasticity it is therefore necessary to define different types of "connectivity," for example, to be able to distinguish between the actual number of anatomical synapses per neuron and the "potential" or "effectual" synapse number in an equivalent network with a fixed structure (Stepanyants et al., 2002; Knoblauch et al., 2014).

In this work we develop substantial new analytical results and insights focusing on the relation between network connectivity, structural plasticity, and memory. First, we work out the relation between "effectual connectivity" in structurally plastic networks and functional measures of brain connectivity such as "effective connectivity" and "transfer entropy." Assuming a simple model of activity propagation between two cortical columns or areas, we argue that effectual connectivity is basically equivalent to the functional measures, while maintaining a precise anatomical interpretation. Second, we give improved estimates on the information storage capacity of a cortical macrocolumn as a function of effectual connectivity (cf., Stepanyants et al., 2002; Knoblauch et al., 2010, 2014). For this we develop exact methods (Knoblauch, 2008) to analyze associative memory in sparsely connected cortical networks storing random activity patterns by structural plasticity. Moreover, we generalize our analyses that are reasonable only for very sparse neural activity, to a recently proposed model of associative memory with structural plasticity (Knoblauch, 2009b, 2016) that is much more appropriate for moderately sparse activity deemed necessary to stabilize cell assemblies or synfire chains in networks with sparse connectivity (Latham and Nirenberg, 2004; Aviel et al., 2005). Third, we point out in more detail how effectual connectivity may relate to cognitive phenomena such as the spacing effect that learning improves if rehearsal is distributed to multiple sessions (Ebbinghaus, 1885; Crowder, 1976; Greene, 1989). For this, we analyze the temporal evolution of effectual connectivity and optimize the time gap between learning sessions to compare the results to recent behavioral data on the spacing effect (Cepeda et al., 2008).

2. MODELING

2.1. Memory, Cell Assemblies and Synapse Ensembles

Memories are commonly identified with patterns of neural activity that can be revisited, evoked and/or stabilized by appropriately modified synaptic connections (Hebb, 1949; Bliss and Collingridge, 1993; Martin et al., 2000; Paulsen and Sejnowski, 2000; for alternative views see Arshavsky, 2006). In the simplest case such a memory corresponds to a group of neurons that fire at the same time and, according to the Hebbian hypothesis that "what fires together wires together" (Hebb, 1949) develop strong mutual synaptic connections (Caporale and Dan, 2008; Clopath et al., 2010; Knoblauch et al., 2012). Such groups of strongly connected neurons are called *cell assemblies* (Hebb, 1949; Palm et al., 2014) and have a number of properties that suggest a function for associative memory (Willshaw et al., 1969; Marr, 1971; Palm, 1980; Hopfield, 1982; Knoblauch, 2011): For example, if a stimulus activates a subset of the cells, the mutual synaptic connections will quickly activate the whole cell assembly which is thought to correspond to the retrieval or completion of a memory. In a similar way, a cell assembly in one brain area u can activate an associated cell assembly in another brain area v. We call the set of synapses that supports retrieval of a given set of memories their *synapse ensemble S*. Memory consolidation is then the process of consolidating the synapses S.

Formally, networks of cell assemblies can be modeled as associative networks, that is, single layer neural networks employing Hebbian-type learning. **Figure 1** illustrates a simple associative network with clipped Hebbian learning (Willshaw et al., 1969; Palm, 1980; Knoblauch et al., 2010; Knoblauch, 2016) that associates binary activity patterns $u^1, u^2, \ldots$ and $v^1, v^2, \ldots$ within neuron populations u and v having size $m = 7$ and $n = 8$, respectively: Here synapses are binary, where a weight W_{ij} may increase from 0 to 1 if both presynaptic neuron u_i and postsynaptic neuron v_j have been synchronously activated for at least θ_{ij} times,

$$W_{ij} = \begin{cases} 1, & \omega_{ij} := \sum_{\mu=1}^{M} R(u_i^\mu, v_j^\mu) \geq \theta_{ij} \\ 0, & \text{otherwise} \end{cases}. \quad (1)$$

where M is the number of stored memories, ω_{ij} is called the synaptic potential, R defines a local learning rule, and θ_{ij} is the threshold of the synapse. In the following we will consider the special case of Equation (1) with Hebbian learning, $R(u_i^\mu, v_j^\mu) = u_i^\mu \cdot v_j^\mu$, and minimal synaptic thresholds $\theta_{ij} = 1$, which corresponds to the well-known Steinbuch or Willshaw model (**Figure 1**; cf., Steinbuch, 1961; Willshaw et al., 1969). Further, we will also investigate the recently proposed general "zip net" model, where both the learning rule R and synaptic thresholds θ_{ij} may be optimized for memory performance (Knoblauch, 2016): For R we assume the optimal homosynaptic or covariance rules, whereas synaptic thresholds θ_{ij} are chosen large enough such that the chance $p_1 := \text{pr}[W_{ij} = 1]$ of potentiating a given synapse is 0.5 to maximize entropy of synaptic weights (see Appendix A.3 for further details). In general, we can identify the synapse ensemble S that supports storage of a memory set $\mathfrak{M}$ by those neuron pairs ij with a sufficiently large synaptic potential $\omega_{ij} \geq \theta_{ij}$ where θ_{ij} may depend on $\mathfrak{M}$. For convenience we may represent S as a binary matrix (with $S_{ij} = 1$ if $ij \in S$ and $S_{ij} = 0$ if $ij \notin S$) similar as the weight matrix W_{ij}.

After learning a memory association $u^\mu \rightarrow v^\mu$, a noisy input $\tilde{u}$ can retrieve an associated memory content $\hat{v}$ in a single
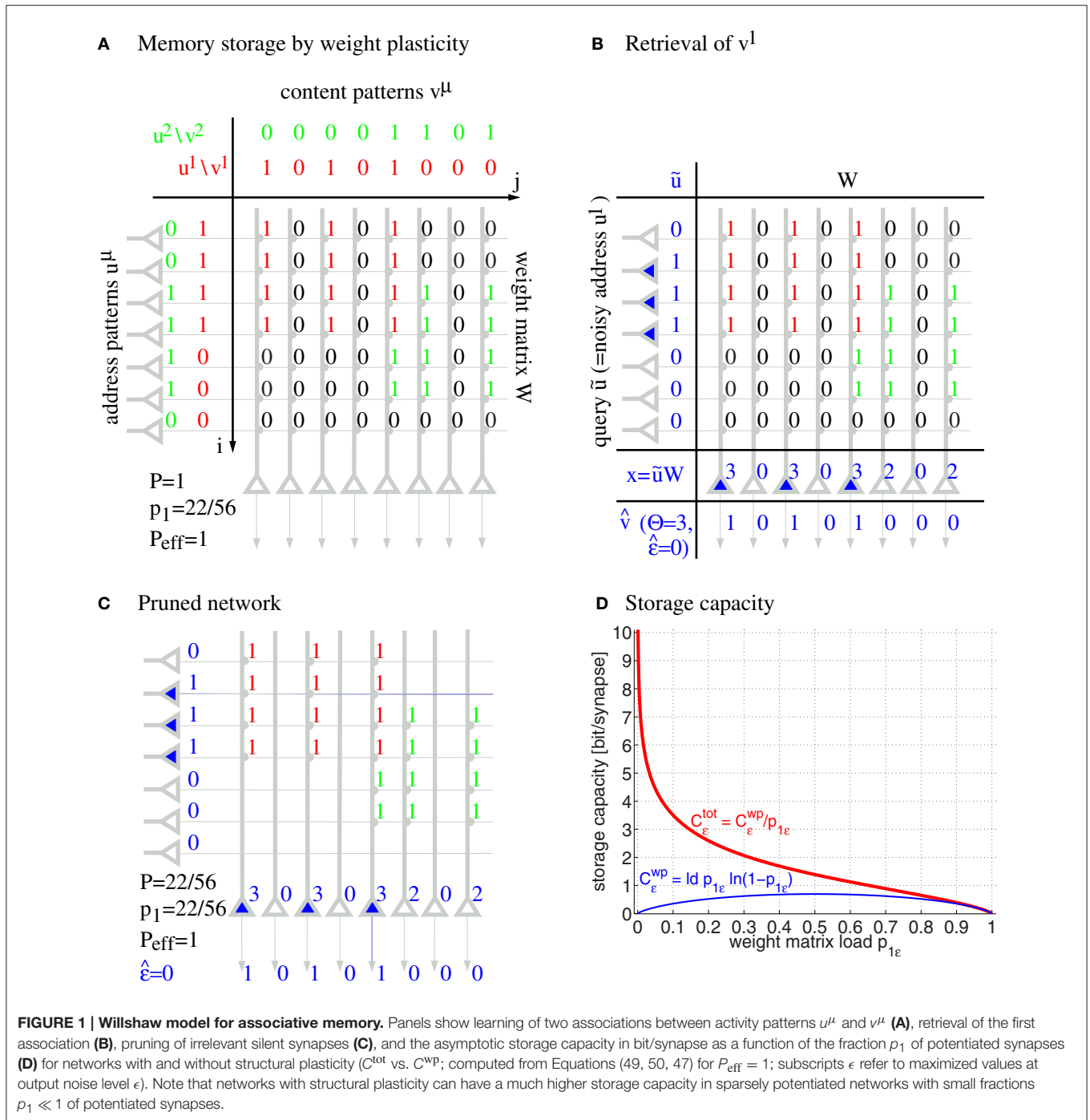
FIGURE 1 | Willshaw model for associative memory. Panels show learning of two associations between activity patterns u^μ and v^μ **(A)**, retrieval of the first association **(B)**, pruning of irrelevant silent synapses **(C)**, and the asymptotic storage capacity in bit/synapse as a function of the fraction p_1 of potentiated synapses **(D)** for networks with and without structural plasticity (C^{tot} vs. C^{wp}; computed from Equations (49, 50, 47) for $P_{eff} = 1$; subscripts ϵ refer to maximized values at output noise level ϵ). Note that networks with structural plasticity can have a much higher storage capacity in sparsely potentiated networks with small fractions $p_1 \ll 1$ of potentiated synapses.

processing step by

$$\hat{v}_j = \begin{cases} 1, & x_j = \left(\sum_{i=1}^{m} \tilde{u}_i W_{ij} + \mathcal{N}_j\right) \geq \Theta_j \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

for appropriately chosen neural firing thresholds Θ_j. The model may include random variables $\mathcal{N}_j$ to account for additional synaptic inputs and further noise sources, but for most analyses and simulations (except Section 3.1) we assume $\mathcal{N}_j = 0$ such that

retrieval depends deterministically on the input $\tilde{u}$. In **Figure 1B**, stimulating with a noisy input pattern $\tilde{u} \approx u^1$ perfectly retrieves the corresponding output pattern $\hat{v} = v^1$ for thresholds $\Theta_j = 2$. In the literature, input and output patterns are also called address and content patterns, and the (noisy) input pattern used for retrieval is called query pattern. In the illustrated completely connected network, the thresholds can simply be chosen according to the number of active units in the query pattern, whereas in biologically more realistic models, firing

thresholds are thought to be controlled by recurrent inhibition, for example, regulating the number of active units to a desired level l being the mean activity of a content pattern (Knoblauch and Palm, 2001). Thus, a common threshold strategy in the more abstract models is to simply select the l most activated "winner" neurons having the largest dendritic potentials x_j. In general, the retrieval outputs may have errors and the retrieval quality can then be judged by the output noise

$$\hat{\epsilon} = \frac{\sum_{j=1}^{n} |\hat{v}_j - v_j^\mu|}{l} \qquad (3)$$

defined as the Hamming distance between $\hat{v}$ and v^μ normalized to the mean number l of active units in an output pattern. Similarly, we can define input noise $\tilde{\epsilon}$ as the Hamming distance between $\tilde{u}$ and u^μ normalized to the mean number k of active units in an input pattern.

In the illustrated network u and v are different neuron populations corresponding to hetero-association. However, all arguments will also apply to auto-association when u and v are identical (with $m = n$, $k = l$), and cell assemblies correspond to cliques of interconnected neurons. In that case output activity can be fed back to the input layer iteratively to improve retrieval results (Schwenker et al., 1996). Stable activation of a cell assembly can then expected if output noise $\hat{\epsilon}$ after the first retrieval step is lower than input noise $\tilde{\epsilon}$.

Capacity analyses show that each synapse can store a large amount of information. For example, even without any structural plasticity, the Willshaw model can store $C^{\text{wp}} = 0.69$ bit per synapse by weight plasticity (wp) corresponding to a large number of about $n^2/\log^2 n$ small cell assemblies, quite close to the theoretical maximum of binary synapses (Willshaw et al., 1969; Palm, 1980). However, unlike in the illustration, real networks will not be fully connected, but, on a local scale of macrocolumns, the chance that two neurons are connected is only about 10% (Braitenberg and Schüz, 1991; Hellwig, 2000). In this case it is still possible to store a considerable number of memories, although maximal M scales with the number of synapses per neuron, and cell assemblies need to be relatively large in this case (Buckingham and Willshaw, 1993; Bosch and Kurfess, 1998; Knoblauch, 2011).

By including structural plasticity, for example, through pruning the unused silent synapses after learning in a network with high connectivity (**Figure 1C**), the total synaptic capacity of the Willshaw model can even increase to $C^{\text{tot}} \sim \log n \gg 1$ bit per (non-silent) synapse, depending on the fraction p_1 of potentiated synapses (**Figure 1D**; see Knoblauch et al., 2010). Moreover, the same high capacity can be achieved for networks that are sparsely connected at any time, if the model includes ongoing structural plasticity and repeated memory rehearsal or additional consolidation mechanisms involving memory replay (Knoblauch et al., 2014).

In Section 3.2 we precisely compute the maximal number of cell assemblies that can be stored in a Willshaw-type cortical macrocolumn. As the Willshaw model is optimal only for extremely small cell assemblies with $k \sim \log n$ (Knoblauch, 2011), we will extend these results also for the general "zip model" of

Equation (1) that performs close to optimal Bayesian learning even for much larger cell assemblies (Knoblauch, 2016).

2.2. Anatomical, Potential, and Effectual Connectivity

As argued in the introduction, connectivity is an important parameter to judge performance. However, network models with structural plasticity need to consider different types of connectivity, in particular, anatomical connectivity P, potential connectivity P_{pot}, effectual connectivity P_{eff}, and target connectivity as measured by consolidation load P_{1S} (see **Figure 2**; cf., Krone et al., 1986; Braitenberg and Schüz, 1991; Hellwig, 2000; Stepanyants et al., 2002; Knoblauch et al., 2014),

$$P := \frac{\text{\#actual synaptic connections}}{mn}, \qquad (4)$$

$$P_{\text{pot}} := \frac{\text{\#potential synaptic connections}}{mn}, \qquad (5)$$

$$P_{\text{eff}} := \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} H(W_{ij}S_{ij})}{\sum_{i=1}^{m} \sum_{j=1}^{n} H(S_{ij}^2)}, \qquad (6)$$

$$P_{1S} := \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} H(S_{ij}^2)}{mn}, \qquad (7)$$

where H is the Heaviside function (with $H(x) = 1$ if $x > 0$ and 0 otherwise) to include the general case of non-binary weights and synapse ensembles ($W_{ij}, S_{ij} \in \mathbb{R}$).

First, *anatomical connectivity P* is defined as the chance that there is an actual synaptic connection between two randomly chosen neurons (**Figure 2A**)[1]. However, for example in the pruned network of **Figure 1C**, the anatomical connectivity P equals the fraction p_1 of potentiated synapses (before pruning) and, thus, conveys only little information about the true (full) connectivity within a cell assembly. Instead, it is more adequate to consider potential and effectual connectivity (**Figures 2B,C**).

Second, *potential connectivity P_{pot}* is defined as the chance that there is a potential synapse between two randomly chosen neurons, where a potential synapse is defined as a cortical location ij where pre- and postsynaptic fibers are close enough such that a synapse could potentially be generated or has already been generated (Stepanyants et al., 2002).

Third, *effectual connectivity P_{eff}* defined as the fraction of "required synapses" that have already been realized is most interesting to judge the functional state of memories or cell assemblies during ongoing learning or consolidation with structural plasticity. Here we call the synapse ensemble S_{ij} required for stable storage of a given memory set also the *consolidation signal*. If ij corresponds to an actual synapse, we

---

[1]More precisely, this means the presence of *at least one synapse* connecting the first to the second neuron. This definition is motivated by simplifications employed by many theories for judging how many memories can be stored. These simplifications include, in particular, (1) point neurons neglecting dendritic compartments and non-linearities, and (2) ideal weight plasticity such that any desired synaptic strength can be realized. Then having two synapses with strength 1 would be equivalent to a single synapse with strength 2. The definition is further justified by experimental findings that the number of actual synapses per connection is narrowly distributed around small positive values (Fares and Stepanyants, 2009; Deger et al., 2012).

may identify the case $S_{ij} > 0$ with tagging synapse ij for consolidation (Frey and Morris, 1997). In case of simple binary network models such as the Willshaw or zip net models, the S_{ij} simply equal the optimal synaptic weights in a fully connected network after storing the whole memory set (Equation 1). Intuitively, if a set of cell assemblies or memories has a certain effectual connectivity P_{eff}, then retrieval performance will be as if these memories would have been stored in a structurally static network with anatomical connectivity P_{eff}, whereas true P in the structurally plastic network may be much lower than P_{eff}.

Last, *target connectivity* or *consolidation load* P_{1S} is the fraction of neuron pairs ij that require a consolidated synapse as specified by S_{ij}. This means that P_{1S} is a measure of the learning load of a consolidation task.
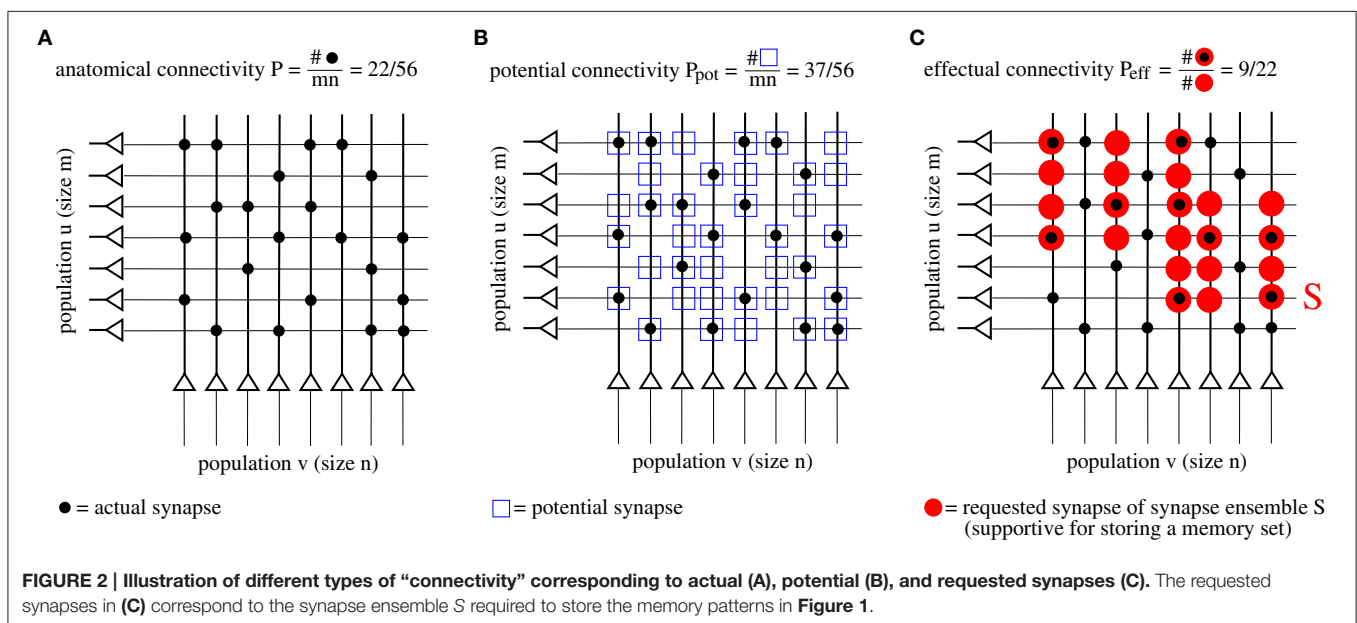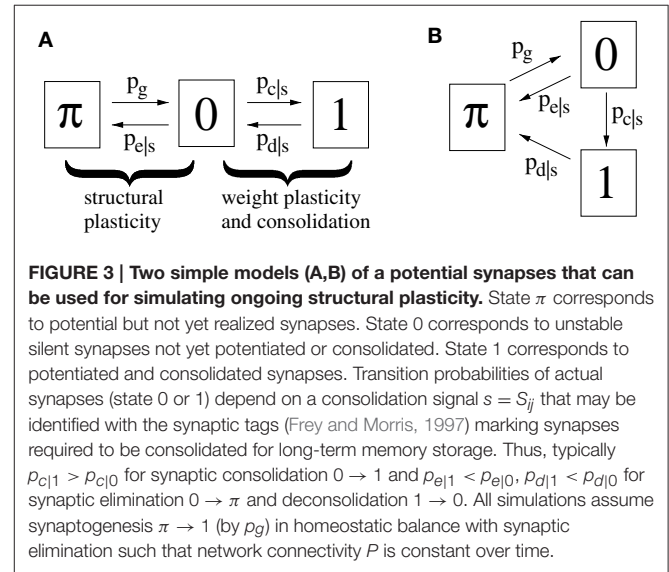
Note that our definitions of P_{eff} and P_{1S} apply as well to network models with gradual synapses ($W_{ij}, S_{ij} \in \mathbb{R}$). More generally, by means of the consolidation signal S_{ij}, we can abstract from any particular network model or application domain. Our theory is therefore not restricted to models of associative memory, but may be applied as well to other connectionist domains, given that the "required" synapse ensembles $\{ij|S_{ij} \neq 0\}$ and their weights can be defined properly by S_{ij}. The following provides a minimal model to simulate the dynamics of effectual connectivity during consolidation.

2.3. Modeling and Efficient Simulation of Structural Plasticity

Figure 3A illustrates a minimal model of a "potential" synapse that can be used to simulate the dynamics of ongoing structural plasticity (Knoblauch, 2009a; Deger et al., 2012; Knoblauch et al., 2014). Here a potential synapse ij^ν is the possible location of a real synapse connecting neuron i to neuron j, for example, a cortical location where axonal and dendritic branches of neurons i and j are close enough to allow the formation of a novel connection by

spine growth and synaptogenesis (Krone et al., 1986; Stepanyants et al., 2002). Note that there may be multiple potential synapses per neuron pair, $\nu = 1, 2, \ldots$. The model assumes that a synapse can be either potential but not yet realized (state π), realized but still silent (state and weight 0), or realized and consolidated (state and weight 1).

For real synapses, state transitions are modulated by the consolidation signal S_{ij} specifying synapses to be potentiated and consolidated Then *structural plasticity* means the transition processes between states π and 0 described by transition probabilities $p_g := \text{pr}[\text{state}(t + 1) = 0|\text{state}(t) = \pi]$ and $p_{e|s} := \text{pr}[\text{state}(t + 1) = \pi|\text{state}(t) = 0, S_{ij} = s]$. Similarly, *weight plasticity* means the transitions between states 0 and 1 described by probabilities $p_{c|s} := \text{pr}[\text{state}(t + 1) = 1|\text{state}(t) =$

FIGURE 3 | Two simple models (A,B) of a potential synapses that can be used for simulating ongoing structural plasticity. State π corresponds to potential but not yet realized synapses. State 0 corresponds to unstable silent synapses not yet potentiated or consolidated. State 1 corresponds to potentiated and consolidated synapses. Transition probabilities of actual synapses (state 0 or 1) depend on a consolidation signal $s = S_{ij}$ that may be identified with the synaptic tags (Frey and Morris, 1997) marking synapses required to be consolidated for long-term memory storage. Thus, typically $p_{c|1} > p_{c|0}$ for synaptic consolidation $0 \rightarrow 1$ and $p_{e|1} < p_{e|0}$, $p_{d|1} < p_{d|0}$ for synaptic elimination $0 \rightarrow \pi$ and deconsolidation $1 \rightarrow 0$. All simulations assume synaptogenesis $\pi \rightarrow 1$ (by p_g) in homeostatic balance with synaptic elimination such that network connectivity P is constant over time.

FIGURE 2 | Illustration of different types of "connectivity" corresponding to actual (A), potential (B), and requested synapses (C). The requested synapses in **(C)** correspond to the synapse ensemble S required to store the memory patterns in **Figure 1**.

$0, S_{ij} = s]$ and $p_{d|s} := \text{pr}[\text{state}(t + 1) = 0|\text{state}(t) = 1, S_{ij} = s]$. For simplicity, we do not distinguish between long-term potentiation (LTP) and synaptic consolidation (or L-LTP), both corresponding to the transition from state 0 to 1. In accordance with the state diagram of **Figure 3A**, the evolution of synaptic states can then be described by probabilities $p_{\text{state}}^{(s)}(t)$ that a given potential synapse receiving $S_{ij} = s$ is in a certain state $\in \{\pi, 0, 1\}$ at time step $t = 0, 1, 2, \ldots$,

$$p_1^{(s)}(t) = (1 - p_{d|s(t)})p_1^{(s)}(t - 1) + p_{c|s(t)}p_0^{(s)}(t - 1) \quad (8)$$

$$p_0^{(s)}(t) = (1 - p_{c|s(t)} - p_{e|s(t)})p_0^{(s)}(t - 1) + p_{d|s(t)}p_1^{(s)}(t - 1) + \\ p_g p_\pi^{(s)}(t - 1) \quad (9)$$

$$p_\pi^{(s)}(t) = (1 - p_g)p_\pi^{(s)}(t - 1) + p_{e|s(t)}p_0^{(s)}(t - 1) \\ = 1 - p_1^{(s)}(t) - p_0^{(s)}(t), \quad (10)$$

where the consolidation signal $s(t) = S_{ij}(t)$ may depend on time.

The second model variant (**Figure 3B**) can be described in a similar way except that $p_{d|s}$ describes the transition from state 1 to state π. Model B is more convenient to analyze the spacing effect. We will see that, in relevant parameter ranges, both model variants behave qualitatively and quantitatively very similar. However, in most simulations we have used model A.

Note that a binary synapse in the original Willshaw model (Equation 1, **Figures 1A,B**) is a special case of the described potential synapse ($p_g = p_{e|s} = p_{d|s} = 0, p_{c|s} = s \in \{0, 1\}, S_{ij} = W_{ij}$ as in Equation 1). Then pruning following a (developmental) learning phase (**Figure 1C**) can be modeled by the same parameters except increasing $p_{e|s} > 0$ to positive values. Finally, adult learning with ongoing structural plasticity can be modeled by introducing a homeostatic constraint to keep P constant (cf., Equation 69 in Appendix B.1; cf., Knoblauch et al., 2014), such that in each step the number of generated and eliminated synapses are about the same. **Figure 4** illustrates such a simulation for $p_{e|s} = 1 - s$ and a fixed consolidation signal S_{ij} corresponding to the same memories as in **Figure 1**. Here the instable silent (state 0) synapses take part in synaptic turnover until they grow at a tagged location ij with $S_{ij} = 1$ where they get consolidated (state 1) and escape further turnover. This process of increasing effectual connectivity (see Equation 70 in Appendix B.2) continues until all potential synapses with $S_{ij} = 1$ have been realized and consolidated (**Figure 4**, $t = 4$) or synaptic turnover comes to an end if all silent synapses have been depleted.

Microscopic simulation of large networks of potential synapses can be expensive. We have therefore developed a method for efficient simulation of structural plasticity on a macroscopic level: Instead of the lower case probabilities (Equations 8–10) we consider additional memory-specific upper-case connectivity variables $P_{\text{state}}^{(s)}$ defined as the fractions of neuron pairs ij that receive a certain consolidation signal $s(t) = S_{ij}(t)$ and are in a certain state $\in \{\emptyset, \pi, 0, 1\}$ (where $\emptyset$ denotes neuron pairs without any potential synapses). In general it is

$$P_1^{(s)}(t) = P_{\text{pot}}^{(s)} \sum_{\mathfrak{n}=1}^{\infty} \mathfrak{p}(\mathfrak{n}) \left(1 - (1 - p_1^{(s)}(t))^{\mathfrak{n}}\right) \quad (11)$$

$$P_\pi^{(s)}(t) = P_{\text{pot}}^{(s)} \sum_{\mathfrak{n}=1}^{\infty} \mathfrak{p}(\mathfrak{n}) \left(p_\pi^{(s)}(t)\right)^{\mathfrak{n}} \quad (12)$$

$$P_0^{(s)}(t) = P_{\text{pot}}^{(s)} - P_1^{(s)}(t) - P_\pi^{(s)}(t) \quad (13)$$

where $p_1^{(s)}$ and $p_\pi^{(s)}$ are as in Equations (8, 10); $P_{\text{pot}}^{(s)}$ is the fraction of neuron pairs receiving s that have at least one potential synapse; and $\mathfrak{p}(\mathfrak{n})$ is the conditional distribution of potential synapse number $\mathfrak{n}$ per neuron pair having at least one potential synapse. Thus, we define a pre-/postsynaptic neuron pair ij to be in state 1 iff it has at least one state-1 synapse; in state 0 iff it does not have a state-1 synapse but at least one state-0 synapse; and in state π if it is neither in state 1 nor state 0 but has at least one potential synapse. See Fares and Stepanyants (2009) for neuroanatomical estimates of $\mathfrak{p}(\mathfrak{n})$ in various cortical areas.

Summing over s we obtain further connectivity variables P_1, P_0, P_π from which we can finally determine the familiar network connectivities defined in the previous section,

$$P_{\text{state}}(t) = \sum_s P_{\text{state}}^{(s)}(t) \quad \text{for state} \in \{\emptyset, \pi, 0, 1\} \quad (14)$$

$$P(t) = P_0(t) + P_1(t) \quad (15)$$

$$P_{\text{pot}}(t) = P_\pi(t) + P_0(t) + P_1(t) \quad (16)$$

$$P_{1S} = \sum_{s \neq 0} \sum_{\text{state} \in \{\emptyset, \pi, 0, 1\}} P_{\text{state}}^{(s)}(t) \quad (17)$$

$$P_{\text{eff}}(t) = \frac{\sum_{s \neq 0} P_1^{(s)}(t)}{P_{1S}}. \quad (18)$$

In general, the consolidation signal $s = s(t) = S_{ij}(t)$ will not be constant but may be a time-varying signal (e.g., if different memory sets are consolidated at different times). To efficiently simulate a large network of many potential synapses, we can partition the set of potential synapses in groups that receive the same signal $s(t)$. For each group we can calculate the temporal evolution of state probabilities $p_\pi^{(s)}(t), p_0^{(s)}(t), p_1^{(s)}(t)$ of individual synapses from Equations (8–10). From this we can then compute from Equations (11–13) the group-specific macroscopic connectivity variables $P_\pi^{(s)}(t), P_0^{(s)}(t), P_1^{(s)}(t)$, and finally from Equations (14–18) the temporal evolution of the various network connectivities $P_\pi(t), P_0(t), P_1(t), P(t)$ as well as effectual connectivity $P_{\text{eff}}(t)$ for certain memory sets. For such an approach the computational cost of simulating structural plasticity scales only with the number of different groups corresponding to different consolidation signals $s(t)$ (instead of the number of potential synapses as for the microscopic simulations).

Moreover, this approach is the basis for further simplifications and the analysis of cognitive phenomena like the spacing effect described in Appendix B. For example, for simplicity, the following simulations and analyses assume that each neuron pair ij can have at most a single potential synapse [i.e., $\mathfrak{p}(1) = 1$]. In previous works we have simulated also a model variant
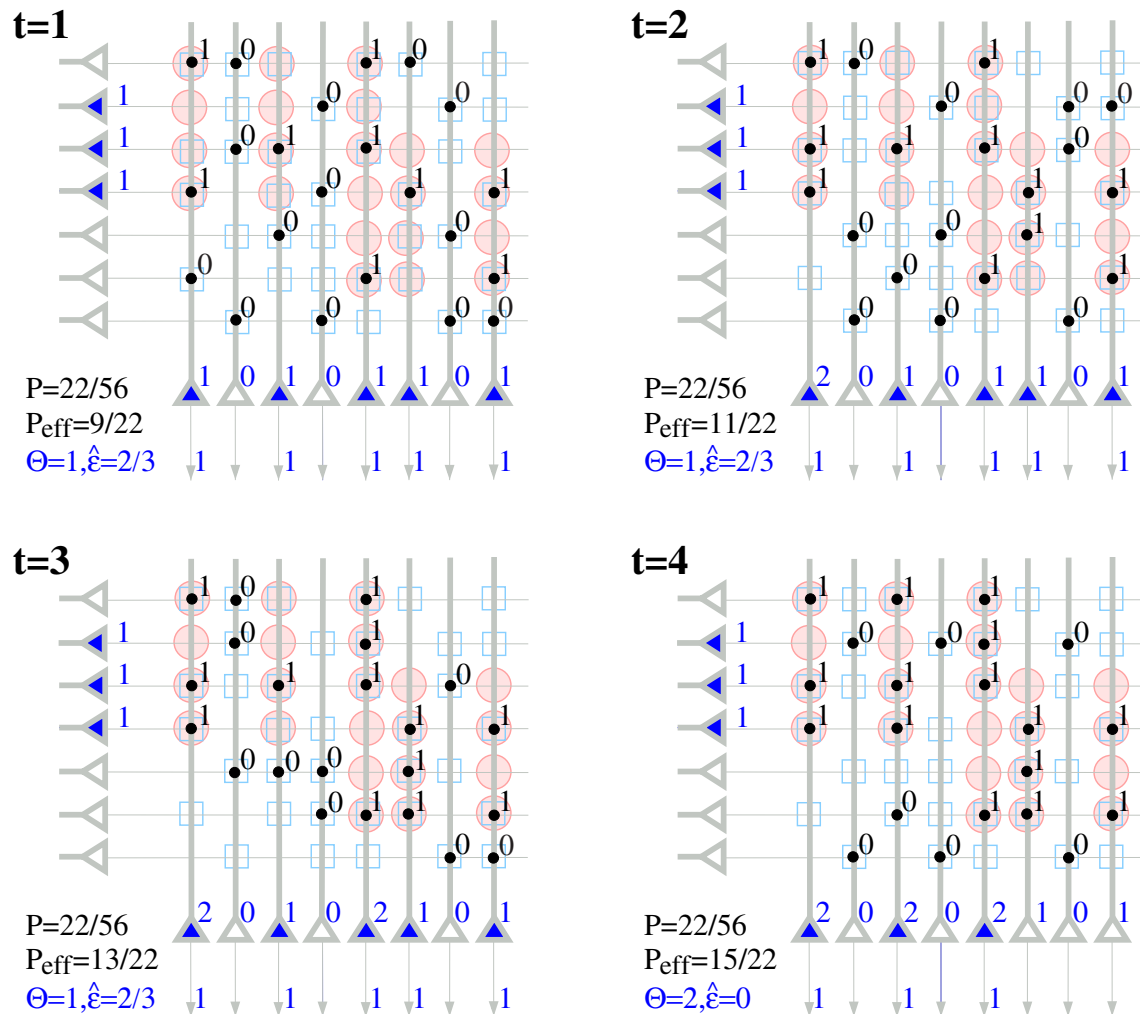
FIGURE 4 | Ongoing structural plasticity maintaining a constant anatomical connectivity $P = 22/56$ for the memory patterns of Figure 1 with actual, potential and requested synapses as in Figure 2 and assuming only single potential synapses per neuron pair ($\mathfrak{p}(1) = 1$, $p_{e|s} = 1 - s$, $p_{c|s} = s$, $p_{d|s} = 0$). Note that P_{eff} increases with time from the anatomical level $P_{eff} = 9/22 \approx P$ at $t = 1$ toward the level of potential connectivity with $P_{eff} = 15/22 \approx P_{pot}$ at $t = 4$. Correspondingly, output noise $\hat{\epsilon}$ decreases with increasing P_{eff}. At each time firing threshold Θ is chosen maximally to activate at least $l = 3$ neurons corresponding to the mean cell assembly size in the output population.

allowing multiple synapses per neuron pair, where we observed very similar results as for single synapses (Knoblauch et al., 2014). As synapse number per connected neuron pair has sometimes been reported to be narrowly distributed around a small number (e.g., $\mathfrak{n} = 4$; cf., Fares and Stepanyants, 2009), one may also identify each single synapse in our model with a group of about 4 real cortical synapses (see Section 4).

This assumption is actually justified by evidence that $\mathfrak{n}$ is narrowly distributed around a small number, e.g., $\mathfrak{n} = 4$ (Fares and Stepanyants, 2009). This means that two neurons are either unconnected or connected by a group of about four synapses (which is actually a very surprising finding as it is unclear how the neurons can regulate $\mathfrak{n}$; cf., Deger et al., 2012). This situation is well consistent with our modeling assumption $\mathfrak{p}(1) = 1$ if we

identify each model synapse with such a group of about 4 real synapses.

3. RESULTS

3.1. Information Storage Capacity, Effectual Connectivity and its Relation to Functional Measures of Brain Connectivity

For an information-theoretic evaluation, associative memories are typically viewed as memory channels that transmit the original content patterns ν^μ and retrieve corresponding retrieval output pattern $\hat{\nu}^\mu$ (see **Figure 5A**). Thus, the absolute amount of transmitted or stored information C_{abs} of all M memories

FIGURE 5 | Relation between effectual connectivity P_{eff}, information storage capacity C, and output noise $\hat{\epsilon}$. (A) Processing model for computing storage capacity $C := C_{abs}/Pmn$ for M given memory associations between input patterns u^μ and output patterns v^μ stored in the synaptic weights (Equation 1; $p := \mathrm{pr}[u^\mu = 1]$, $q := \mathrm{pr}[v^\mu = 1]$; k and l are mean cell assembly sizes in neuron populations u and v). During retrieval noisy address inputs $\tilde{u}^\mu$ with component errors $p_{ab} := \mathrm{pr}[\tilde{u}_i^\mu = b | u_i^\mu = a]$ and input noise $\tilde{\epsilon} := p_{10} + (1/q - 1)p_{01}$ are propagated through the network (Equation 2) yielding output patterns $\hat{v}^\mu$ with component errors $q_{ab} := \mathrm{pr}[\hat{v}_j^\mu = b | v_j^\mu = a]$ and output noise $\hat{\epsilon} = q_{10} + (1/q - 1)q_{01}$. The retrieved information is then the transinformation between v^μ and $\hat{v}^\mu$. To simplify analysis, we assume independent transmission of individual (i.i.d.) memory bits v_j^μ over a binary channel with transmission errors q_{01}, q_{10}. **(B)** Information storage capacity $C(P_{eff})$ (blue curve), and output noise $\hat{\epsilon}(P_{eff})$ (red curve) as functions of effectual connectivity P_{eff} for a structurally plastic Willshaw network (similar to **Figure 4**) of $m = n = 100,000$ neurons storing $M = 10^6$ cell assemblies of sizes $k = l = 50$ and having anatomical connectivity $P = 0.1$ assuming zero input noise ($\tilde{\epsilon} = 0$). Data have been computed similar to Equation (37) using Equations (44–46) for $0 \leq P_{eff} \leq P/p_1$.

equals the transinformation or mutual information (Shannon and Weaver, 1949; Cover and Thomas, 1991)

$$C_{abs} := T(\hat{V}; V) := \sum p(\hat{V}, V) \log_2 \frac{p(\hat{V}, V)}{p(\hat{V}) \cdot p(V)} \quad (19)$$

where $V := (v^1, v^2, \ldots v^M)$ and $\hat{V} := (\hat{v}^1, \hat{v}^2, \ldots, \hat{v}^M)$ correspond to the sets of original and retrieved content patterns, and $p(.)$ to their probability distributions. If all M memories and n neurons have independent and identically distributed (i.i.d) activities (e.g., same fraction q of active units per pattern and component transmission error probabilities q_{01}, q_{10}), we can approximate this memory channel by a simple binary channel transmitting $M \cdot n$ memory bits $v_j^\mu \mapsto \hat{v}_j^\mu$ as assumed in Appendix A. Then

$$C_{abs} \approx M \cdot T(\hat{v}^\mu; v^\mu) \approx M \cdot n \cdot T(q, q_{01}, q_{10}) \quad (20)$$

where $T(\hat{v}^\mu; v^\mu)$ is the transinformation for single memory patterns and $T(q, q_{01}, q_{10})$ is the transinformation of a single bit (see Equation 38). From this we obtain the normalized information storage capacity C per synapse after dividing C_{abs} by the number of synapses Pmn (similar to Equation 37).

In our first experiment we have investigated the relation between information storage capacity and effectual connectivity P_{eff} during ongoing structural plasticity. For this we have assumed a larger network of size $m = n = 100000$ with anatomical connectivity $P = 0.1$ and larger cell assemblies with sizes $k = l = 50$, but otherwise a similar setting as for the toy example illustrated by **Figure 4**. **Figure 5B** shows output noise $\hat{\epsilon}$ and normalized capacity C as functions of effectual connectivity P_{eff} for a given number of $M = 10^6$ random

memories. Interestingly, both $\hat{\epsilon}$ and C turn out to be monotonic functions of P_{eff} because output errors decrease with increasing P_{eff} (see Equations 45, 46). Therefore, also output noise $\hat{\epsilon}(P_{eff})$ decreases with increasing P_{eff} whereas, correspondingly, stored information per synapse $C(P_{eff})$ increases with P_{eff}. Because monotonic functions are invertible, we can thus conclude that effectual connectivity P_{eff} is an equivalent measure of information storage capacity or the transinformation (=mutual information) between the activity patterns of two neuron populations u and v. As can be seen from our data, $C(P_{eff})$ tends to be even linear over a large range, $C \sim P_{eff}$, until saturation occurs if $\hat{\epsilon} \to 0$ approaches zero corresponding to high-fidelity retrieval outputs.

Next, based on the this equivalence between P_{eff} and C, we work out the close relationship between P_{eff} and commonly used functional measures of brain connectivity. Recall that we have introduced "effectual connectivity" as a measure of memory related synaptic connectivity (**Figure 2C**) that shares with other definitions of connectivity (such as anatomical and potential connectivity) the idea that any "connectivity" measure should correspond to the chance of finding a connection element (such as an actual or potential synapse) between two cells. By contrast, in brain imaging and connectome analysis (Friston, 1994; Sporns, 2007) the term "connectivity" has a more heterogeneous meaning ranging from patterns of synaptic connections (anatomical connectivity) and correlations between neural activity (functional connectivity) to causal interactions between brain areas. The latter is also referred to as "effective connectivity" although usually measured in information theoretic terms (bits) such as delayed mutual information or transfer entropy (Schreiber, 2000). For example, in the simplest case the transfer entropy between activities $u(t)$ and $v(t)$ measured in two brain areas u

and v is defined as

$$T_{u \to v} := \sum p(v(t+1), u(t), v(t)) \log_2 \frac{p(v(t+1)|u(t), v(t))}{p(v(t+1)|v(t))} \quad (21)$$

where $p(.)$ denotes the distribution of activity patterns (see Equation 4 in Schreiber, 2000)[2] . Such ideas of effective connectivity come from the desire to extract directions of information flow between two brain areas from measured neural activity, contrasting with (symmetric) correlation measures that can neither detect processing directions nor distinguish between causal interactions and correlated activity due to a common cause.

To see the relation between these functional measures of "effective connectivity" and P_{eff}, first, note that transfer entropy equals the well-known conditional transinformation or conditional mutual information between $v(t+1)$ and $u(t)$ given $v(t)$ (Dobrushin, 1959; Wyner, 1978),

$$T(v(t+1); u(t)|v(t)) :=$$

$$\sum p(v(t+1), u(t), v(t)) \log_2 \frac{p(v(t+1), u(t)|v(t))}{p(v(t+1)|v(t)) \cdot p(u(t)|v(t))} \quad (22)$$

$$= \sum p(v(t+1), u(t), v(t)) \log_2$$

$$\frac{p(v(t+1)|u(t), v(t)) \cdot p(u(t)|v(t))}{p(v(t+1)|v(t)) \cdot p(u(t)|v(t))} = T_{u \to v} . \quad (23)$$

Second, we may apply this to one-step retrieval in an associative memory (Equation 2). Then $u(t) = \tilde{u}^\mu$ is a noisy input, and the update $v(t+1) = F(u(t)) = \hat{v}^\mu$ produces the corresponding output pattern, where the mapping F corresponds to activity propagation through the associative network. As here the update does not depend on the old state $v(t)$, we may approximate transfer entropy by the regular transinformation or mutual information

$$T_{u \to v} = T(v(t+1); u(t)|v(t)) \approx T(F(u(t)); u(t)) \quad (24)$$

$$= I(u(t)) - I(u(t)|F(u(t))) \quad (25)$$

$$= I(F(u(t))) - I(F(u(t))|u(t)) \quad (26)$$

where $I(X) := -\sum_x p(x) \log p(x)$ is the Shannon information of a random variable X, and $I(X|Y) := -\sum_{x,y} p(x, y) \log p(x|y)$ the conditional information of X given Y (Shannon and Weaver, 1949; Cover and Thomas, 1991). Thus, up to normalization, transfer entropy $T_{u \to v} \approx T(F(u(t)); u(t)) = T(\hat{v}^\mu; \tilde{u}^\mu)$ has a very similar form as storage capacity C_{abs} in Equation (20). If $F(u)$ is deterministic, the second term in Equation (26) vanishes and transfer entropy equals the output information $I(F(u(t))) \leq I(u(t))$. If $F(u)$ is also invertible, the second term in Equation (25) would vanish and $T_{u \to v} = I(u(t)) = I(F(u(t))) = C_{\text{abs}}/M$. However, in the associative memory application many

---

[2] The general case considers delay vectors $(u(t), u(t-1), \ldots, u(t-K+1))$ and $(v(t), v(t-1), \ldots, v(t-L+1))$ instead of $u(t)$ and $v(t)$.

input patterns are (ideally) mapped to one memory and $F(u)$ is noninvertible and thus $T_{u \to v} = I(F(u(t))) < I(u(t))$. Moreover, in more realistic cortex models F is also nondeterministic as $v(t+1)$ will depend not only on activity $u(t)$ from a single input area, but also on inputs from further cortical and subcortical areas as well as on numerous additional noise sources. Thus, in fact it will be $T_{u \to v} < I(F(u(t)))$.

Third, we can compare $T_{u \to v}$ to information storage capacity (Equation 20) by normalizing to single memory patterns,

$$\frac{C_{\text{abs}}}{M} = \frac{CPmn}{M} = T(\hat{v}^\mu; v^\mu) = T(F(u(t); v^{\mu(u(t))}) \quad (27)$$

$$= I(F(u(t))) - I(F(u(t))|v^{\mu(u(t))}) \quad (28)$$

where $\mu(u(t))$ is a function determining the memory index of the input pattern $u^{\mu(u(t))}$ best matching the current input $\tilde{u} = u(t)$. Thus, comparing Equation (26) to Equation (28) yields generally
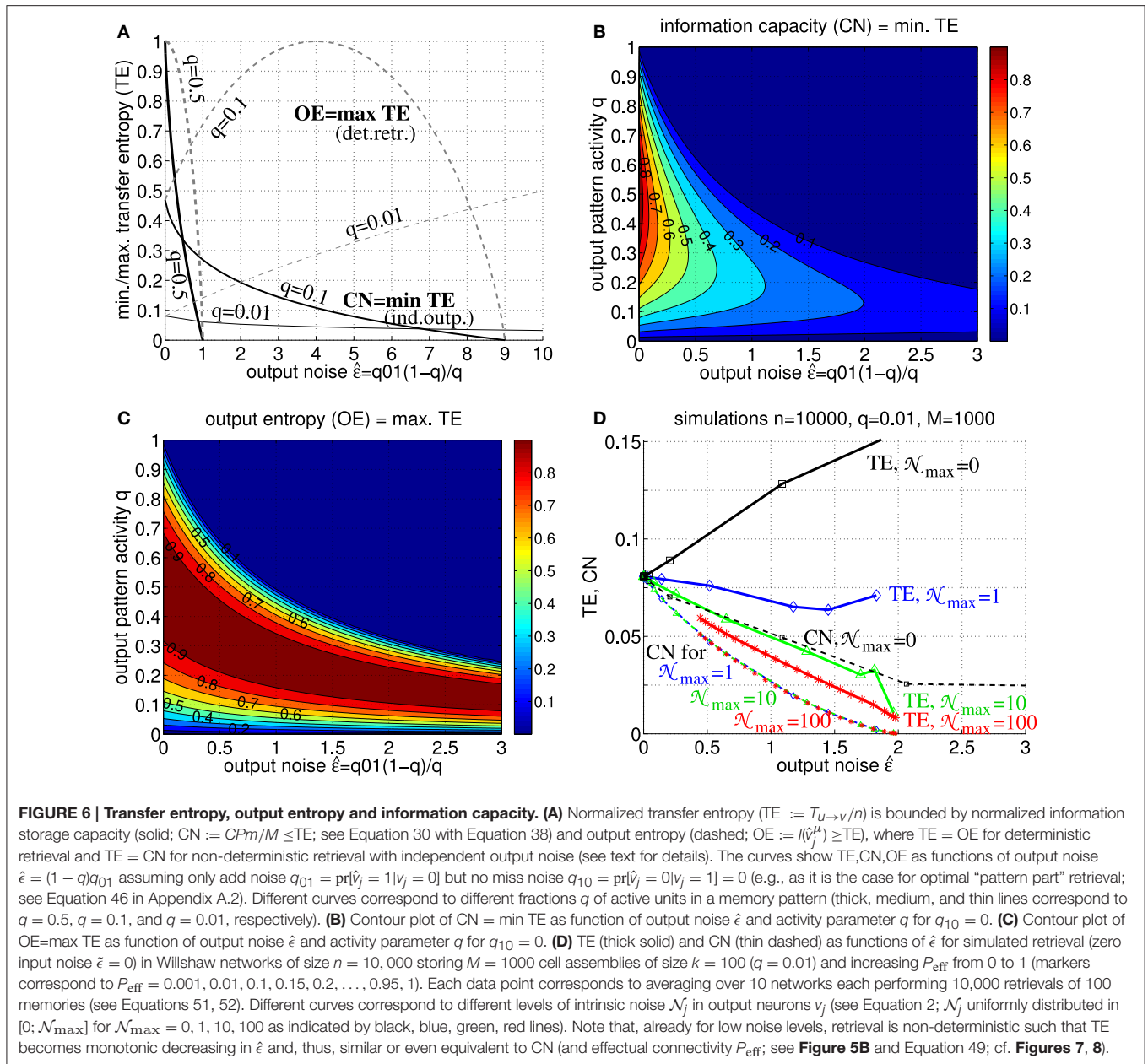
$$T_{u \to v} - \frac{C_{\text{abs}}}{M} = I(F(u(t))|v^{\mu(u(t))}) - I(F(u(t))|u(t)) \geq 0 . \quad (29)$$

where the bound is true as $v^{\mu(u(t))}$ is a deterministic function of $u(t)$. In particular, for deterministic F, transfer entropy $T_{u \to v} = \frac{C_{\text{abs}}}{M} + I(F(u(t))|v^{\mu(u(t))})$ typically exceeds normalized capacity $\frac{C_{\text{abs}}}{M}$, whereas equality follows for $I(F(u(t))|v^{\mu(u(t))}) = I(F(u(t))|u(t))$, for example, error-free retrieval with $F(u(t)) = v^{\mu(u(t))}$. Appendix A.4 shows that equality holds generally as well for nondeterministic propagation of activity (e.g., Equation 2 with $\mathcal{N}_j \neq 0$) if we assume that component retrieval errors occur independently with probabilities $q_{01} := \text{pr}[\hat{v}_j^\mu = 1|v_j^\mu = 0] \approx \text{pr}[\hat{v}_j^\mu = 1|v_j^\mu = 0, \tilde{u}] = \text{pr}[\hat{v}_j^\mu = 1|\tilde{u}]$ and $q_{10} := \text{pr}[\hat{v}_j^\mu = 0|v_j^\mu = 1] \approx \text{pr}[\hat{v}_j^\mu = 0|v_j^\mu = 1, \tilde{u}] = \text{pr}[\hat{v}_j^\mu = 0|\tilde{u}]$ corresponding to the same (nondeterministic, i.i.d.) processing model as we have presumed in our capacity analysis (**Figure 5A**; see also Appendix A, Equations 42–43 or Equations 45–46 for Willshaw networks). Then normalizing transfer entropy TE and information capacity CN per output unit yields (see Equations 53, 38)

$$TE := \frac{T_{u \to v}}{n} \overset{>}{\approx} T(q, q_{01}, q_{10}) \approx CN := \frac{C_{\text{abs}}}{Mn} = \frac{CPm}{M} . \quad (30)$$

Thus, "effective connectivity" as measured by transfer entropy becomes (up to normalization) equivalent to the information storage capacity C of associative networks (see Equation 37 with Equation 38).

Figure 6 shows upper bounds $TE \leq OE := I(v_j^\mu)$ and lower bounds TE≥CN of transfer entropy as functions of output noise level $\hat{\epsilon} = qq_{10} + (1-q)q_{01}$ for different activities q of output patterns (cf., Equations 26, 29, 30). For low output noise ($\hat{\epsilon} \to 0$) both $T_{u \to v}$ and C approach the full information content of the stored memory set. In general both TE and CN are monotonic functions of $\hat{\epsilon}$ for relevant (sufficiently low) noise levels $\hat{\epsilon}$. While TE increases with $\hat{\epsilon}$ for deterministic retrieval ($\mathcal{N}_j = 0$; cf. Equation 2), TE becomes a decreasing function of $\hat{\epsilon}$ already for low levels of intrinsic noise ($\mathcal{N}_j$ on the order of single synaptic inputs; see panel D). Similar decreases are obtained even without

FIGURE 6 | Transfer entropy, output entropy and information capacity. (A) Normalized transfer entropy (TE $:= T_{u \to v}/n$) is bounded by normalized information storage capacity (solid; CN $:= CPm/M \leq$ TE; see Equation 30 with Equation 38) and output entropy (dashed; OE $:= I(\hat{v}_j^\mu) \geq$ TE), where TE = OE for deterministic retrieval and TE = CN for non-deterministic retrieval with independent output noise (see text for details). The curves show TE,CN,OE as functions of output noise $\hat{\epsilon} = (1-q)q_{01}$ assuming only add noise $q_{01} = \text{pr}[\hat{v}_j = 1|v_j = 0]$ but no miss noise $q_{10} = \text{pr}[\hat{v}_j = 0|v_j = 1] = 0$ (e.g., as it is the case for optimal "pattern part" retrieval; see Equation 46 in Appendix A.2). Different curves correspond to different fractions q of active units in a memory pattern (thick, medium, and thin lines correspond to $q = 0.5$, $q = 0.1$, and $q = 0.01$, respectively). **(B)** Contour plot of CN = min TE as function of output noise $\hat{\epsilon}$ and activity parameter q for $q_{10} = 0$. **(C)** Contour plot of OE=max TE as function of output noise $\hat{\epsilon}$ and activity parameter q for $q_{10} = 0$. **(D)** TE (thick solid) and CN (thin dashed) as functions of $\hat{\epsilon}$ for simulated retrieval (zero input noise $\tilde{\epsilon} = 0$) in Willshaw networks of size $n = 10,000$ storing $M = 1000$ cell assemblies of size $k = 100$ ($q = 0.01$) and increasing P_{eff} from 0 to 1 (markers correspond to $P_{\text{eff}} = 0.001, 0.01, 0.1, 0.15, 0.2, \ldots, 0.95, 1$). Each data point corresponds to averaging over 10 networks each performing 10,000 retrievals of 100 memories (see Equations 51, 52). Different curves correspond to different levels of intrinsic noise $\mathcal{N}_j$ in output neurons v_j (see Equation 2; $\mathcal{N}_j$ uniformly distributed in $[0; \mathcal{N}_{\max}]$ for $\mathcal{N}_{\max} = 0, 1, 10, 100$ as indicated by black, blue, green, red lines). Note that, already for low noise levels, retrieval is non-deterministic such that TE becomes monotonic decreasing in $\hat{\epsilon}$ and, thus, similar or even equivalent to CN (and effectual connectivity P_{eff}; see **Figure 5B** and Equation 49; cf. **Figures 7**, **8**).
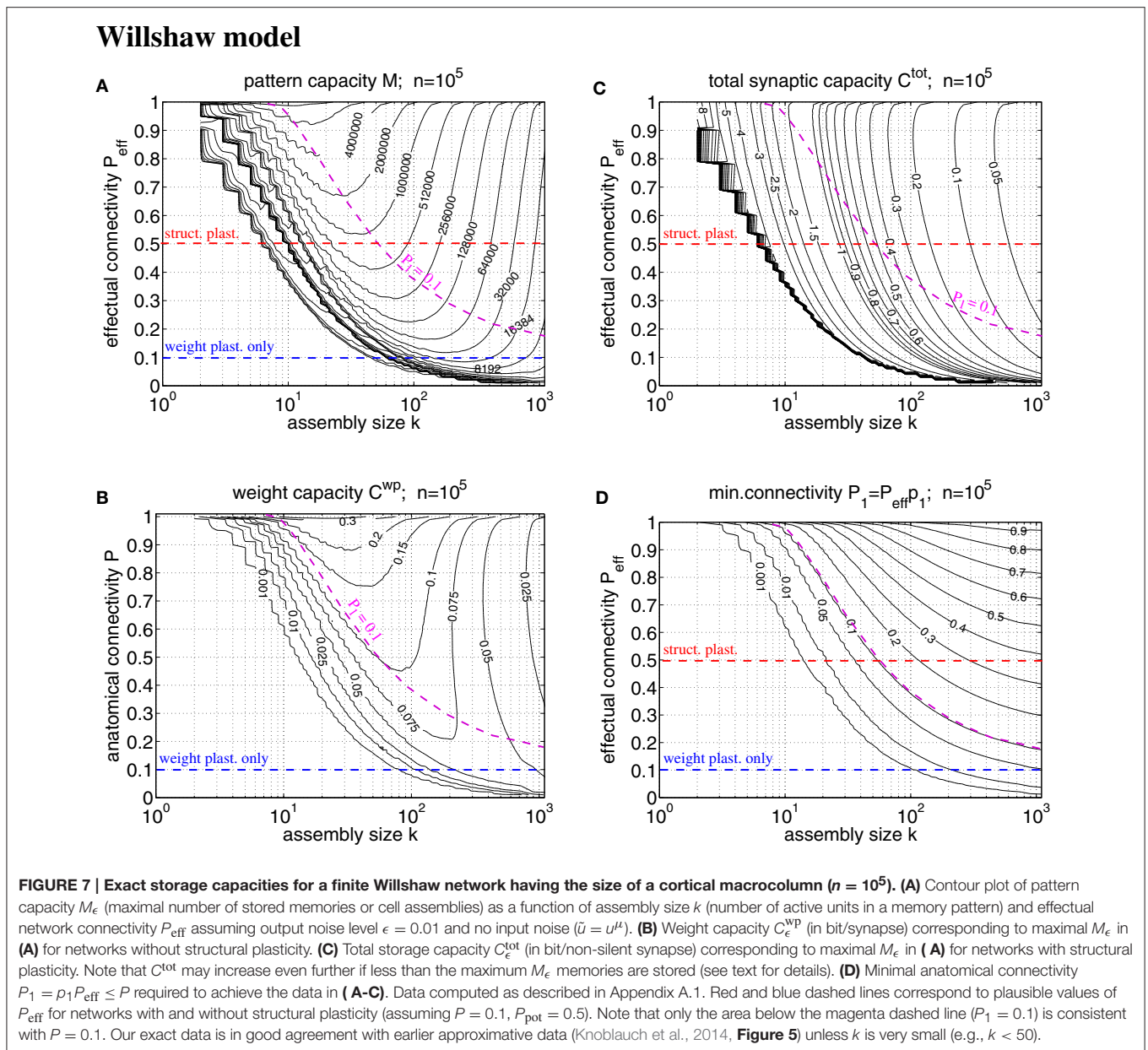
intrinsic noise, $\mathcal{N}_j = 0$, if the target assembly v^μ receives (noisy) synaptic inputs from multiple cortical populations (data not shown; cf. Braitenberg and Schüz, 1991).

Our results thus show that, at least for realistic intrinsic noise and/or inter-columnar synaptic connectivity, transfer entropy $T_{u \to v}$ becomes equivalent to information capacity C. Because of the monotonic (or even linear) dependence of C on P_{eff} (see **Figure 5B** and Equation 49; cf. **Figures 7**, **8**), transfer entropy is equivalent also to effectual connectivity P_{eff}. Thus, we may interpret effectual connectivity P_{eff} as an essentially equivalent measure of "effective connectivity" as previously defined for functional brain imaging. Still, due to its anatomical definition, P_{eff} can only measure a *potential* causal interaction. For example, if both the synaptic connections from brain area u to v and

the reverse connections from v to u have high P_{eff}, we will not be able to infer the direction of information flow in a certain memory task unless we measure the actual neural activity.
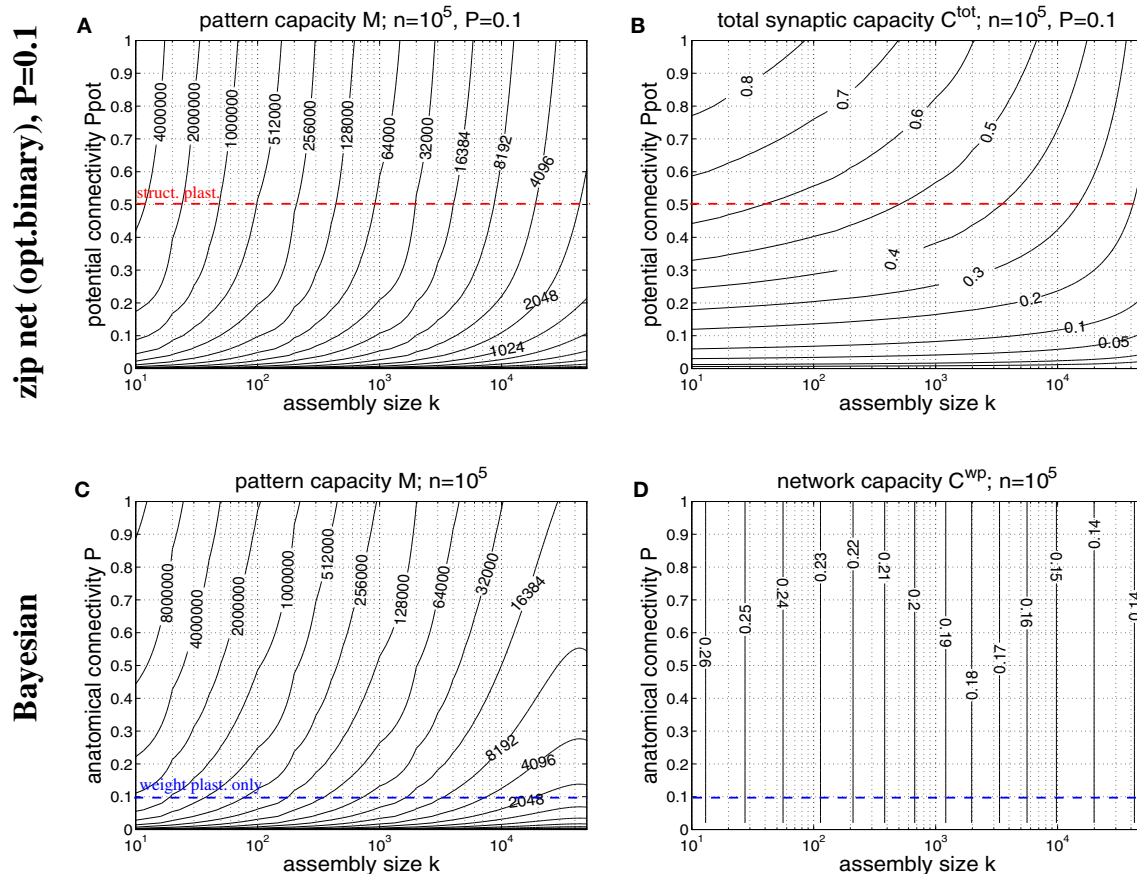
3.2. Storage Capacity of a Macrocolumnar Cortical Network

A typical cortical macrocolumn comprises on the order of $n = 10^5$ neurons below about 1 mm^2 cortex surface, where the anatomical connectivity is about $P = 0.1$ and the potential connectivity about $P_{\text{pot}} = 0.5$ corresponding to a filling fraction $f := P/P_{\text{pot}} = 0.2$ (Braitenberg and Schüz, 1991; Hellwig, 2000; Stepanyants et al., 2002). Sizes of cell assemblies have been estimated to be somewhere between 50 and 500 in entorhinal

Willshaw model

FIGURE 7 | Exact storage capacities for a finite Willshaw network having the size of a cortical macrocolumn ($n = 10^5$). (A) Contour plot of pattern capacity M_ϵ (maximal number of stored memories or cell assemblies) as a function of assembly size k (number of active units in a memory pattern) and effectual network connectivity P_{eff} assuming output noise level $\epsilon = 0.01$ and no input noise ($\tilde{u} = u^\mu$). **(B)** Weight capacity C_ϵ^{wp} (in bit/synapse) corresponding to maximal M_ϵ in **(A)** for networks without structural plasticity. **(C)** Total storage capacity C_ϵ^{tot} (in bit/non-silent synapse) corresponding to maximal M_ϵ in **(A)** for networks with structural plasticity. Note that C^{tot} may increase even further if less than the maximum M_ϵ memories are stored (see text for details). **(D)** Minimal anatomical connectivity $P_1 = p_1 P_{eff} \leq P$ required to achieve the data in **(A-C)**. Data computed as described in Appendix A.1. Red and blue dashed lines correspond to plausible values of P_{eff} for networks with and without structural plasticity (assuming $P = 0.1$, $P_{pot} = 0.5$). Note that only the area below the magenta dashed line ($P_1 = 0.1$) is consistent with $P = 0.1$. Our exact data is in good agreement with earlier approximative data (Knoblauch et al., 2014, **Figure 5**) unless k is very small (e.g., $k < 50$).

cortex (Waydo et al., 2006). Given these data we can try to estimate the number M of local cell assemblies or memories that can be stored in a macrocolumn (Sommer, 2000). In a previous work (Knoblauch et al., 2014, **Figure 5**) we have estimated the storage capacity for the Willshaw model (**Figures 1**, **4**) by approximating dendritic potential distributions by Gaussians. However, this approximation can be off as, in particular, for sparse activity dendritic potentials can strongly deviate from Gaussians. We have therefore developed a method to compute the exact storage capacity for the Willshaw model storing random memories (see Appendix A). **Figure 7** shows corresponding contour plots of pattern capacity M_ϵ, weight capacity C_ϵ^{wp}, total synaptic capacity C_ϵ^{tot}, and the required minimal anatomical connectivity P_1 (assuming that all silent synapses have been

pruned in the end). We can make several observations: First, the exact results can significantly deviate from the approximations (cf., Knoblauch et al., 2014, **Figure 5**). In particular, for extremely sparse activity ($k < 10$) the Gaussian assumption seems violated and the true capacities are significantly lower than estimated previously. Still, for larger more realistic $50 < k < 500$ the new data is in good agreement with the previous Gaussian estimates, and for even larger $k > 500$ the true capacities even slightly exceed the previous estimates. Second, the previous conclusions, therefore, largely hold: Without structural plasticity ($P_{eff} = P = 0.1$) the storage capacity would be generally very low and only a small number of memories could be stored. For very sparse $k \approx 50$ not even a single memory could be stored and thus, the cell assembly hypothesis would be inconsistent with experimental

FIGURE 8 | Storage capacities for binary zip nets (A,B) and Bayesian neural networks (C,D) having the size of a cortical macrocolumn ($n = 10^5$). (A) Contour plot of the pattern capacity M_ϵ of an optimal binary zip net (employing the optimal covariance or homosynaptic learning rule; see Knoblauch, 2016) with $P = P_1 = 0.1$ as a function of cell assembly size k and potential network connectivity P_{pot} (which is here an upper bound on the achievable effectual connectivity P_{eff}). **(B)** Total storage capacity C_ϵ^{tot} for zip nets including structural plasticity for the setting of **(A)**. **(C)** Contour plot of the pattern capacity M_ϵ of an optimal Bayesian associative network (Knoblauch, 2011) without structural plasticity as a function of cell assembly size k and anatomical network connectivity P. **(D)** Weight capacity C_ϵ^{wp} for the Bayesian net for the setting of **(C)**. Other parameters are as assumed for **Figure 7** ($\epsilon = 0.01$, $\tilde{u} = u^\mu$). Data computed as described in Appendix A.3. Red and blue dashed lines correspond to plausible values for P_{pot} and P, respectively.

estimates of k. Third, by contrast, networks including structural plasticity increasing P_{eff} from $P = 0.1$ to $P_{\text{pot}} \approx 0.5$ can store many more memories: For example, for $k = 50$, the pattern capacity increases from $M \approx 0$ to about $M \approx 800,000$. For $k = 500$, there is still an increase from $M \approx 13,000$ to $M \approx 45,000$. Fourth, correspondingly, networks without structural plasticity would have only a very small weight capacity C^{wp}: For example, at $P_{\text{eff}} = P = 0.1$ it is $C^{\text{wp}} \approx 0$ bps for $k \leq 50$ and still $C^{\text{wp}} < 0.07$ bps for $k = 500$. Fifth, by contrast, networks with structural plasticity have a much higher total synaptic capacity C^{tot}, i.e., they can store much more information per actual synapse and are therefore also much more energy-efficient, in particular for sparse activity: Although the very high values $C^{\text{tot}} \rightarrow \log n$ are approached only for unrealistically low k and high P_{eff}, they can still store $C^{\text{tot}} \approx 0.5$ bps for realistic $P_{\text{eff}} = 0.5$ and $k = 50$. This high value appears to decrease, however, to only $C^{\text{tot}} \approx 0.06$ bps for $k = 500$ which would suggest that, for relatively large cell assemblies with $k = 500$, a

network without structural plasticity (at $P = 0.1$) would be more efficient than a network with structural plasticity (at $P_{\text{eff}} = 0.5$). However, as the Willshaw model is known to be sub-optimal for relatively large $k \gg \log n$, we will re-discuss this issue below for a more general network model. Sixth, another weakness of the Willshaw model is that the fraction $p_1 := 1 - (1 - \frac{k^2}{n^2})^M$ of 1-synapses is coupled both to cell assembly size k and number of stored memories M (due to the fixed synaptic threshold $\theta = 1$, cf., Equation 1). Therefore, the residual (minimal) anatomical connectivity of a pruned network $P_1 = p_1 P_{\text{eff}}$ depends also on k,M, and we can obtain $P_1 \approx P = 0.1$ consistent with physiology only in a limited range of the k-P_{eff}-planes of **Figure 7**. At least, physiological $k \approx 50$ and $P_{\text{eff}} \approx 0.5$ match physiological $P_1 = 0.1$, whereas larger $k \gg 50$ would require P_1 being larger than the anatomical connectivity $P = 0.1$. As many cortical areas comprise significant fractions $P_0 > 0$ of silent synapses we may as well allow for smaller $P_1 < P = 0.1$ satisfying $P_0 + P_1 = P$ (where C^{tot} would become a measure only of energy efficiency,

but no longer of space efficiency), but the very high values of $C^{tot} \gg 1$ can generally be reached only for tiny fractions of 1-synapses.

In order to overcome some weaknesses of the Willshaw model we have recently proposed a novel network model (so called binary "zip nets") where the fraction p_1 of potentiated 1-synapses is no longer coupled to cell assembly size k and number M (Knoblauch, 2009b, 2010b, 2016). Instead, the model assumes that synaptic thresholds θ_{ij} (see Equation 1) are under homeostatic control to maintain a constant fractions p_1 (or P_1) of potentiated 1-synapses. We have shown for the limit $Mpq \to \infty$ that this model can reach for $p_1 = 0.5$ up to a "zip" factor $\zeta \approx 0.64$ almost the same high storage capacities M_ϵ and C_ϵ^{wp} as the optimal Bayesian neural network (Kononenko, 1989; Lansner and Ekeberg, 1989; Knoblauch, 2011), although requiring only binary synapses. Moreover, if compressed by structural plasticity, zip nets can also reach $C_\epsilon^{tot} \to \log n$ for $p_1 \to 0$, similar to the Willshaw model. As the Willshaw model is optimal only for extremely sparse activity ($k \leq \log n$) it is thus interesting to evaluate the performance gain of structural plasticity for physiological k using the zip net instead of the Willshaw model. **Figure 8** shows data from evaluating storage capacity of a cortical macrocolumn of size $n = 10^5$ both for the zip net model (upper panels) and the Bayesian model (lower panels), the latter being a benchmark for the optimal network without structural plasticity (Knoblauch, 2011). In order to compute the capacity of the zip net we have assumed physiological anatomical connectivity $P = P_1 = 0.1$ where structural plasticity "moves" the $P_1 n^2$ relevant 1-synapses to the most useful locations within the limits given by potential connectivity P_{pot} (as P_1 is fixed, unlike to the Willshaw model, final P_{eff} after learning may be lower than P_{pot} in zip nets; see Appendix A.3 for methodological details). We can make the following observations: First, as expected, for high connectivity and very sparse activity (e.g., $k \ll 100$) the zip nets may perform worse than the Willshaw model (because the Willshaw model then performs close to the optimal Bayesian net). Second, for more physiological parameters $P_{pot} \leq 0.5$, $k \geq 50$ the zip net can store significantly more memories than the Willshaw model, for example, for $P_{pot} = 0.5$ the zip net reaches $M \approx 1000000$ for $k = 50$ and still $M \approx 120,000$ for $k = 500$. Third, also the total synaptic capacity C^{tot} is higher than for the Willshaw network, for example for $P_{pot} = 0.5$, it is $C^{tot} \approx 0.6$ for $k = 50$ and still $C^{tot} \approx 0.5$ for $k = 500$ (remember that the corresponding value for the Willshaw model required unphysiological $P_1 > 0.1$). Fourth, although the Bayesian network can store significantly more memories M it has only a moderate storage capacity below $C^{wp} = 0.25$. In fact, for plausible cell assembly sizes, the binary synapses of the zip net with structural plasticity at $P = 0.1$ and $P_{pot} = 0.5$ achieve more than double the capacity of the optimal (but biologically implausible) Bayesian network with real-valued synapses at $P = 0.1$.

In summary, the new data confirms our previous conclusion that structural plasticity strongly increases space and energy efficiency of associative memory storage in neural networks under physiological conditions (Knoblauch et al., 2014).

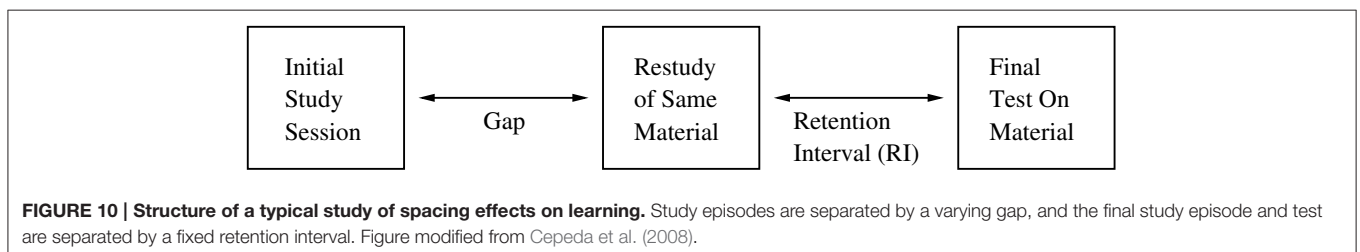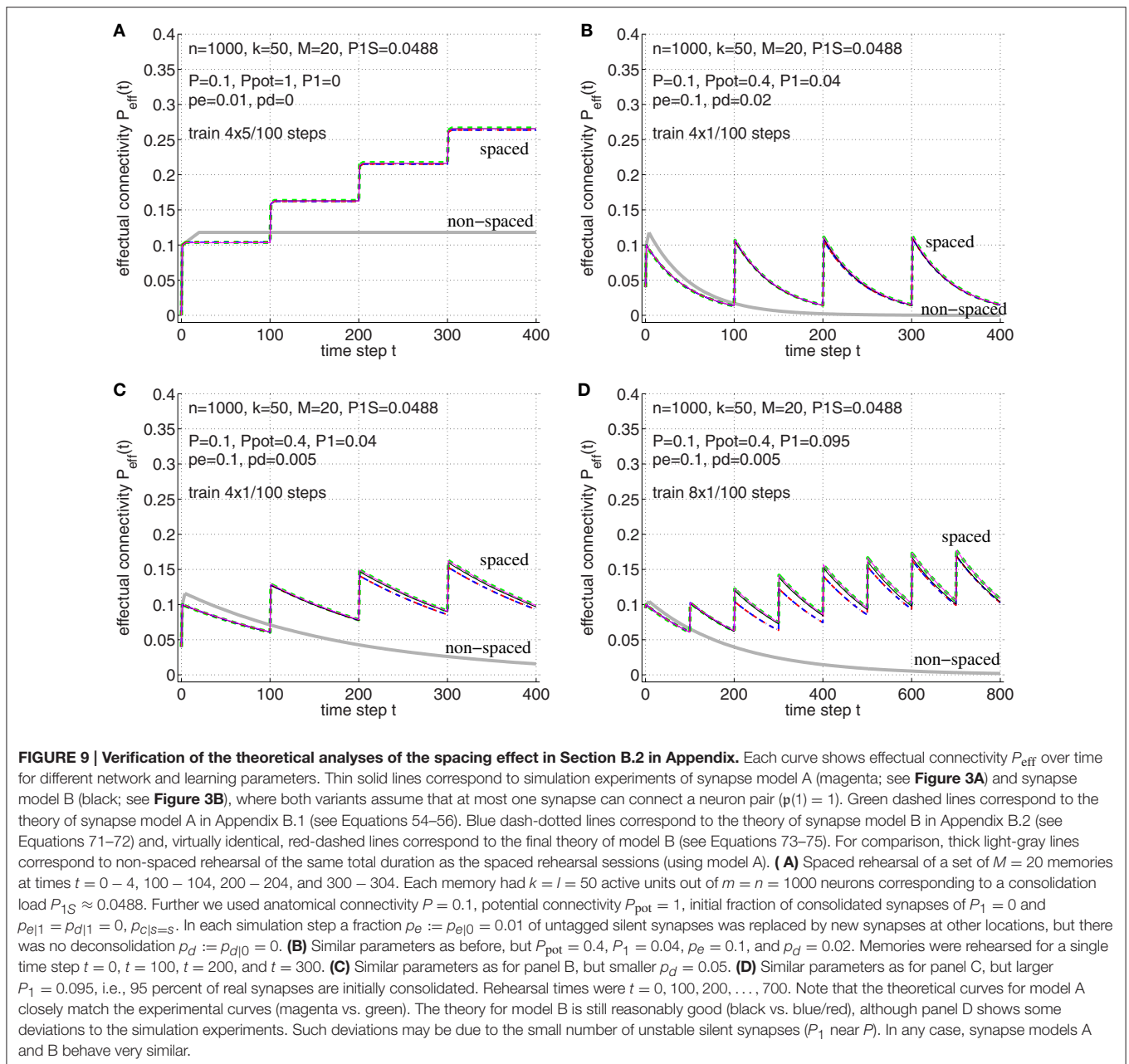## 3.3. Structural Plasticity and the Spacing Effect

In previous works we have linked structural plasticity and cognitive effects like retrograde amnesia, absence of catastrophic forgetting, and the spacing effect (Knoblauch, 2009a; Knoblauch et al., 2014). Here we focus on a more detailed analysis of the spacing effect that learning is most efficient if learning is distributed in time (Ebbinghaus, 1885; Crowder, 1976; Greene, 1989). For example, learning a list of vocabularies in two sessions each lasting 10 min is more efficient than learning in a single session of 20 min. We have explained this effect by slow ongoing structural plasticity and fast synaptic weight plasticity: Thus, spaced learning is useful because during the (long) time gaps between two (or more) learning sessions structural plasticity can grow many novel synapses that are potentially useful for storing new memories and that can quickly be potentiated and consolidated by synaptic weight plasticity during the (brief) learning sessions (Knoblauch et al., 2014, Section 7.3).

Appendix B.2 develops a simplified theory of the spacing effect that is based on model variant B of a potential synapse (which can more easily be analyzed than model A; see **Figure 3**) and the concept and methods proposed in Section 2.3. In particular, with (Equations 73–75) we can easily compute the temporal evolution of effectual connectivity $P_{eff}(t)$ for arbitrary rehearsal sequences of a novel set of memories to be learned. As output noise $\hat{\epsilon}$ is a decreasing function of P_{eff} (see **Figure 5B**), we can use P_{eff} as a measure of retrieval performance.

To illustrate the effect of spaced vs. non-spaced rehearsal (or consolidation) on P_{eff}, and to verify the theory in Appendix B.2, **Figure 9** shows the temporal evolution of $P_{eff}(t)$ for different models and synapse parameters. It can be seen that for high potential connectivity $P_{pot} \approx 1$ and low deconsolidation probability $p_{d|s} \approx 0$ the spacing effect is most pronounced and the network easily realizes high-performance long-term memory (with high P_{eff}; see panel A). Larger $p_{d|0} > 0$ is plausible to model short-term memory, whereas realizing long-term memory would then require repeated consolidation steps (panels B–D). Significant spacing effects are visible for any parameter set. Comparing the microscopic simulations of both synapse models from **Figure 3** to the macroscopic simulations using the methods of Section 2.3 and Appendix B.2, it can be seen that all model and simulation variants behave qualitatively and quantitatively very similar. This justifies to use the theory of Appendix B.2 in the following analysis of recent psychological experiments exploring the spacing effect.

For example, Cepeda et al. (2008) describe an internet-based learning experiment investigating the spacing effect over longer time intervals of more than a year (up to 455 days). The structure of the experiment followed **Figure 10**. The subjects had to learn a set of facts in an initial study session. After a gap interval (0–105 days) without any learning the subjects restudied the same material. After a retention interval (RI; 7–350 days) there was the final test.

These experiments showed that the final recall performance depends both on the gap and the RI showing the following characteristics: First, for any gap duration, recall performance

FIGURE 9 | Verification of the theoretical analyses of the spacing effect in Section B.2 in Appendix. Each curve shows effectual connectivity P_{eff} over time for different network and learning parameters. Thin solid lines correspond to simulation experiments of synapse model A (magenta; see **Figure 3A**) and synapse model B (black; see **Figure 3B**), where both variants assume that at most one synapse can connect a neuron pair ($\mathbf{p}(1) = 1$). Green dashed lines correspond to the theory of synapse model A in Appendix B.1 (see Equations 54–56). Blue dash-dotted lines correspond to the theory of synapse model B in Appendix B.2 (see Equations 71–72) and, virtually identical, red-dashed lines correspond to the final theory of model B (see Equations 73–75). For comparison, thick light-gray lines correspond to non-spaced rehearsal of the same total duration as the spaced rehearsal sessions (using model A). (**A**) Spaced rehearsal of a set of $M = 20$ memories at times $t = 0 - 4$, $100 - 104$, $200 - 204$, and $300 - 304$. Each memory had $k = l = 50$ active units out of $m = n = 1000$ neurons corresponding to a consolidation load $P_{1S} \approx 0.0488$. Further we used anatomical connectivity $P = 0.1$, potential connectivity $P_{\text{pot}} = 1$, initial fraction of consolidated synapses of $P_1 = 0$ and $p_{e|1} = p_{d|1} = 0$, $p_{c|s=s}$. In each simulation step a fraction $p_e := p_{e|0} = 0.01$ of untagged silent synapses was replaced by new synapses at other locations, but there was no deconsolidation $p_d = p_{d|0} = 0$. (**B**) Similar parameters as before, but $P_{\text{pot}} = 0.4$, $P_1 = 0.04$, $p_e = 0.1$, and $p_d = 0.02$. Memories were rehearsed for a single time step $t = 0$, $t = 100$, $t = 200$, and $t = 300$. (**C**) Similar parameters as for panel B, but smaller $p_d = 0.05$. (**D**) Similar parameters as for panel C, but larger $P_1 = 0.095$, i.e., 95 percent of real synapses are initially consolidated. Rehearsal times were $t = 0, 100, 200, \ldots, 700$. Note that the theoretical curves for model A closely match the experimental curves (magenta vs. green). The theory for model B is still reasonably good (black vs. blue/red), although panel D shows some deviations to the simulation experiments. Such deviations may be due to the small number of unstable silent synapses (P_1 near P). In any case, synapse models A and B behave very similar.

FIGURE 10 | Structure of a typical study of spacing effects on learning. Study episodes are separated by a varying gap, and the final study episode and test are separated by a fixed retention interval. Figure modified from Cepeda et al. (2008).

decline as a function of RI in a negatively accelerated fashion, which corresponds to the familiar "forgetting curve." Second, for any RI greater than zero, an increase in study gap causes

recall to first increase and then decrease. Third, as RI increases, the optimal gap increases, whereas that ratio of optimal gap to RI declines. The following shows that our simple associative

memory model based on structural plasticity can explain most of these characteristics.

It is straight-forward to model the experiments of Cepeda et al. (2008) by applying our model of structural plasticity and synaptic consolidation. **Figure 11** illustrates $P_{\text{eff}}(t)$ for a learning protocol as employed in the experiments: In an initial study session facts are learned until time $t^{(1)}$ when some desired performance level $P_{\text{eff}}^{(1)}$ is reached. After a gap the facts are rehearsed briefly at time $t^{(2)}$ reaching a performance equivalent to $P_{\text{eff}}^{(2)}$. After the retention interval at time $t^{(3)}$ performance still corresponds to an effectual connectivity $P_{\text{eff}}^{(3)}$.

Similar to Cepeda et al. (2008), we want to optimize the gap duration in order to maximize $P_{\text{eff}}^{(3)}$ for a given retention interval RI. After the second rehearsal at time $t^{(2)}$, P_{eff} decays exponentially by a fixed factor $1 - p_{d|0}$ per time step (Equation 74). Therefore, $P_{\text{eff}}^{(3)} = P_{\text{eff}}^{(2)}(1 - p_{d|0})^{t^{(3)} - t^{(2)}}$ is a function of $P_{\text{eff}}^{(2)}$ that decreases with the retention interval length $t^{(3)} - t^{(2)}$. We can therefore equivalently maximize $P_{\text{eff}}^{(2)}$ with respect to the gap length $\Delta t := t^{(2)} - t^{(1)}$. For $p_{c|s} = s$, $p_{e|1} = p_{d|1} = 0$, a good approximation of $P_{\text{eff}}^{(2)}$ follows from Equation (73),

$$P_{\text{eff}}^{(2)} \approx \frac{\begin{array}{l} PP_{\text{pot}} + [(P_{\text{pot}} - P)P_{\text{eff}}^{(1)} - P_{\text{pot}}P_1^{(t1)}](1 - p_{d|0})^{\Delta t} \\ -P_{\text{pot}}(P - P_1^{(t1)})(1 - p_{e|0})^{\Delta t} \end{array}}{P_{\text{pot}} - P_1^{(t1)}(1 - p_{d|0})^{\Delta t} - [P - P_1^{(t1)}](1 - p_{e|0})^{\Delta t}} , \quad (31)$$

where $P_1^{(t1)} := P_1^{(t0)}(1 - P_{1S})(1 - p_{d|0})^{t^{(1)}} + P_{1S}P_{\text{eff}}^{(1)}$ with $P_1^{(t0)}$ denoting the initial fraction of consolidated synapses at time 0.[3] Since $P_{\text{eff}}^{(2)}$ does not depend on the RI we can already see that the optimal gap interval Δt depends on the RI neither (which contrasts with the experiments reporting that optimal Δt increases with RI). Optimizing Δt yields the optimality criterion (see Appendix B.3)

$$\frac{P_{\text{eff}}^{(1)} - P_1^{(t1)}}{P - P_1^{(t1)}} + (\alpha - 1)\frac{P_{\text{eff}}^{(1)}}{P_{\text{pot}}}x^\alpha - \alpha x^{\alpha - 1} = 0 . \quad (32)$$
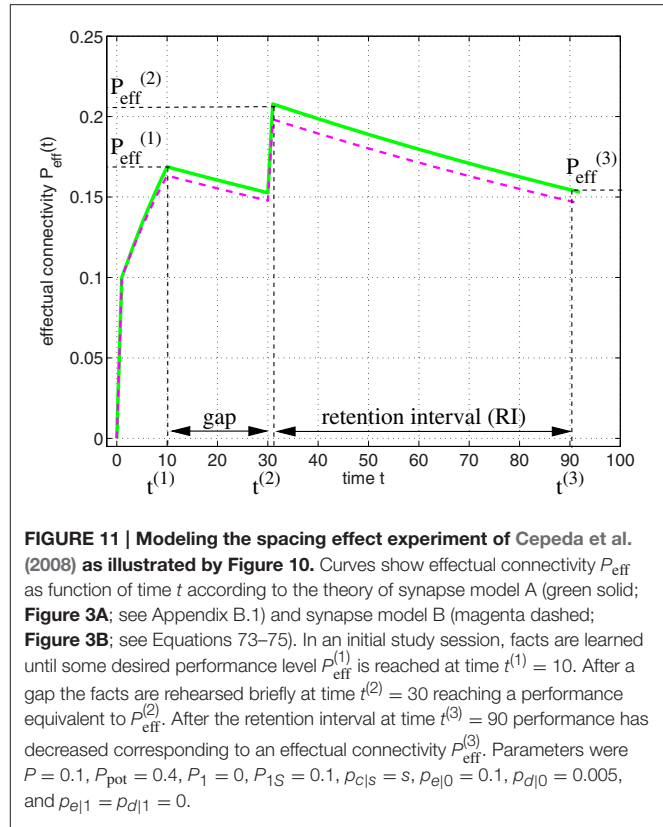
with

$$x := (1 - p_{d|0})^{\Delta t} = e^{\Delta t \ln(1 - p_{d|0})} \Leftrightarrow \Delta t = \frac{\ln x}{\ln(1 - p_{d|0})} \quad (33)$$

$$\alpha := \frac{\ln(1 - p_{e|0})}{\ln(1 - p_{d|0})} , \quad (34)$$

which can easily be evaluated using standard Newton-type numerical methods. Note that Equation (32) can be used to link neuroanatomical and neurophysiological to psychological data.

FIGURE 11 | Modeling the spacing effect experiment of Cepeda et al. (2008) **as illustrated by Figure 10.** Curves show effectual connectivity P_{eff} as function of time t according to the theory of synapse model A (green solid; **Figure 3A**; see Appendix B.1) and synapse model B (magenta dashed; **Figure 3B**; see Equations 73–75). In an initial study session, facts are learned until some desired performance level $P_{\text{eff}}^{(1)}$ is reached at time $t^{(1)} = 10$. After a gap the facts are rehearsed briefly at time $t^{(2)} = 30$ reaching a performance equivalent to $P_{\text{eff}}^{(2)}$. After the retention interval at time $t^{(3)} = 90$ performance has decreased corresponding to an effectual connectivity $P_{\text{eff}}^{(3)}$. Parameters were $P = 0.1$, $P_{\text{pot}} = 0.4$, $P_1 = 0$, $P_{1S} = 0.1$, $p_{c|s} = s$, $p_{e|0} = 0.1$, $p_{d|0} = 0.005$, and $p_{e|1} = p_{d|1} = 0$.

For example, given the optimal gap Δt_{opt} from psychological experiments, Equation (32) gives a constraint on the remaining network and learning parameters. Alternatively, we can solve Equation (32) to determine the optimal gap Δt_{opt} given the remaining parameters.

We have verified Equation (32) by simulations illustrated in **Figure 12** (compare simulation data to Cepeda et al., 2008, **Figure 3**). For these simulations we chose physiologically plausible model parameters: Similarly as before we used $P_{\text{pot}} = 0.4$ (Stepanyants et al., 2002; DePaola et al., 2006), $P = 0.1$ (Braitenberg and Schüz, 1991; Hellwig, 2000). Further, we used $P_1^{(t0)} = 0.02$ as neurophysiological experiments investigating two-state properties of synapses suggest that about 20% of synapses are in the "up" state (Petersen et al., 1998; O'Connor et al., 2005)[4] . Then we chose a small consolidation load $P_{1S} = 0.001$ assuming that the small set of novel facts is negligible compared to the presumably large set of older memories. As before, we assumed p_g in homeostatic balance to maintain a constant anatomical connectivity $P(t)$ (Equation 69) and binary consolidation signals $s = S_{ij} \in \{0, 1\}$ with $p_{c|s} = s$ and $p_{d|1} = p_{e|1} = 0$ for any synapse ij. For the remaining learning

---

[3] Note that a constant (instead of decaying) "background" consolidation P_1 can be modeled, for example, by using $P_1^{(t0)} = 0$ and then excluding the initially consolidated synapses from further simulation. This means to simulate a network with anatomical connectivity $P' = P - P_1$, potential connectivity $P'_{\text{pot}} = P_{\text{pot}} - P_1$, no initial consolidation with $P'_1 = 0$, and otherwise same parameters as the original network. Then the effectual connectivity can be computed from $P_1^{(1)} = P_1^{(1)'} + P_{1S}P_1$ using Equation (18) where $P_1^{(1)'} mn$ is obtained from the simulation.
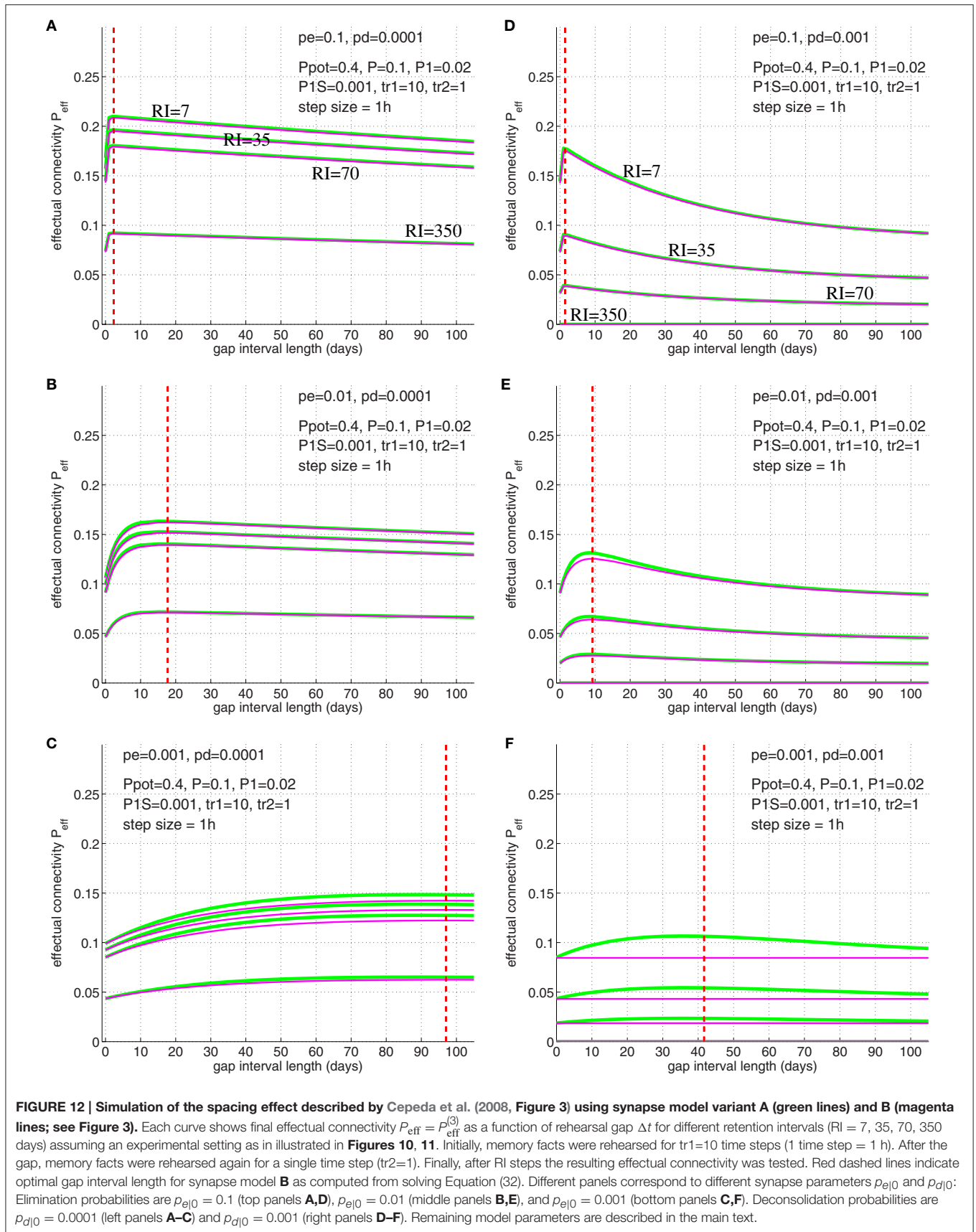
[4] It may be more realistic that the total number of "up"-synapses is kept constant by homeostatic processes (i.e., $P_1/P = 0.2$). However, here we were more interested in verifying our theory which assumes exponential decay of "up"-synapses. To account for homeostasis with constant P_1 one may proceed as described in footnote 3. Nevertheless, the qualitative behavior of the model does not strongly depend on P_1 or $P_1^{(t0)}$ unless their values being close to P which would strongly impair learning.

FIGURE 12 | Simulation of the spacing effect described by Cepeda et al. (2008, Figure 3) using synapse model variant A (green lines) and B (magenta lines; see Figure 3). Each curve shows final effectual connectivity $P_{\text{eff}} = P_{\text{eff}}^{(3)}$ as a function of rehearsal gap Δt for different retention intervals (RI = 7, 35, 70, 350 days) assuming an experimental setting as in illustrated in **Figures 10**, **11**. Initially, memory facts were rehearsed for tr1=10 time steps (1 time step = 1 h). After the gap, memory facts were rehearsed again for a single time step (tr2=1). Finally, after RI steps the resulting effectual connectivity was tested. Red dashed lines indicate optimal gap interval length for synapse model **B** as computed from solving Equation (32). Different panels correspond to different synapse parameters $p_{e|0}$ and $p_{d|0}$: Elimination probabilities are $p_{e|0} = 0.1$ (top panels **A,D**), $p_{e|0} = 0.01$ (middle panels **B,E**), and $p_{e|0} = 0.001$ (bottom panels **C,F**). Deconsolidation probabilities are $p_{d|0} = 0.0001$ (left panels **A–C**) and $p_{d|0} = 0.001$ (right panels **D–F**). Remaining model parameters are described in the main text.

parameters $p_{e|0}$ and $p_{d|0}$ we have chosen several combinations to test their relevance for fitting the model to the observed data.

The simulation results of **Figure 12** imply the following conclusions: First, the simulations show that the optimal gap determined by Equation (32) closely matches the simulation results, for both synapse models (**Figure 3**). Second, for fixed deconsolidation $p_{d|0}$, larger $p_{e|0}$ implies smaller optimal gaps Δt_{opt}. Thus, faster synaptic turnover implies smaller optimal gaps. Third, for fixed turnover $p_{e|0}$, larger $p_{d|0}$ implies smaller Δt_{opt}. Thus, faster deconsolidation implies also smaller optimal gaps. Fourth, together this means that faster (weight and structural) plasticity implies smaller optimal gaps. Fifth, although model variants A and B (**Figure 3**) behave very similar for most parameters settings, they can differ significantly for some parameter combinations. For example, for $p_{e|0} = p_{d|0} = 0.001$ (panel F) the peak in P_{eff} of model A is more than a third larger than the peak of model B. In fact, there the curve of model B is almost flat. Still, even here, the optimal gap interval length is very similar for the two models. An obvious reason why model A sometimes performs better than model B is that deconsolidation of a synapse in model A does not necessarily imply elimination as in model B (see **Figure 3**). Sixth, our simple model already satisfies two of the three characteristics of the spacing effect mentioned above: Both the forgetting effect and the existence of an optimal time gap can be observed in a wide parameter range. Best fits to the experimental data occurred for $p_{e|0} = 0.01$ and $p_{d|0} = 0.0002$ (between parameters of panels B,C; data not shown). Last, however, our simple model cannot reproduce the third characteristic: As argued above, the optimal gap interval length Δt_{opt} does not depend on the retention interval RI. This is in contrast to the experiments of Cepeda et al. (2008) reporting that Δt_{opt} increases with RI.

Nevertheless, we have shown in some preliminary simulations that a slight extension of the model can easily resolve the latter discrepancy (Knoblauch, 2010a): By mixing two populations of synapses having different plasticity parameters corresponding to a small and large optimal gap (or fast and slow plasticity), respectively, it is possible to obtain a dependence of optimal spacing as in the experiments.

4. DISCUSSION

In this theoretical work we have identified roles of structural plasticity and effectual connectivity P_{eff} for network performance, measuring brain connectivity, and optimizing learning protocols. Analyzing how many cell assemblies or memories can be stored in a cortical macrocolumn (of size $1\,mm^3$), we find a strong dependence of storage capacity on P_{eff} and cell assembly size k (see **Figures 7**, **8**). We find that, without structural plasticity, when cell assemblies would have a connectivity close to the low anatomical connectivity $P \approx 0.1$, only a small number of relatively large cell assemblies could be stably stored (Latham and Nirenberg, 2004; Aviel et al., 2005) and, correspondingly, retrieval would not be energy efficient (Attwell and Laughlin, 2001; Laughlin and Sejnowski, 2003; Lennie, 2003; Knoblauch et al., 2010; Knoblauch, 2016). It thus appears that storing and

efficiently retrieving a large number of small cell assemblies as observed in some areas of the medial temporal lobe (Waydo et al., 2006) would require structural plasticity increasing P_{eff} from the low anatomical level toward the much larger level of potential connectivity $P_{pot} \approx 0.5$ (Stepanyants et al., 2002). Similarly, our model predicts ongoing structural plasticity for any cortical area that exhibits sparse neural activity and high capacity.

Moreover, we have shown a close relation between our definition of effectual connectivity P_{eff} and previous measures of functional brain connectivity. While the latter, for example transfer entropy, are solely based on correlations between neural activity in cortical areas (Schreiber, 2000), our definition of P_{eff} as the fraction of realized required synapses has also a clear anatomical basis (**Figure 2**). Via the link of memory channel capacity $C(P_{eff})$ used to measure storage capacity of a neural network, we have shown that P_{eff} is basically an equivalent measure of functional connectivity as transfer entropy. By this, it may become possible to establish an anatomically grounded link between structural plasticity and functional connectivity. For example, this could enable predictions on which cortical areas exhibit strong ongoing structural plasticity during certain cognitive tasks.

Further, as one example linking cognitive phenomena to its potential anatomical basis, we have more closely investigated the spacing effect that learning becomes more efficient if rehearsal is distributed to multiple sessions (Crowder, 1976; Greene, 1989; Cepeda et al., 2008). In previous works we have already shown that the spacing effect can easily be explained by structural plasticity and that, therefore, structural plasticity may be the common physiological basis of various forms of the spacing effect (Knoblauch, 2009a; Knoblauch et al., 2014). Here we have extended these results to explain some recent long-term memory experiments investigating the optimal time gap between two learning sessions (Cepeda et al., 2008). For a given retention interval, our model, if fitted to neuroanatomical data, can easily explain the profile of the psychological data, in particular, the existence of an optimal gap that maximizes memory retention. It is even possible to analyze this profile, linking the optimal gap to parameters of the synapse model, in particular, the rate of deconsolidation $p_{d|0}$ and elimination $p_{e|0}$. Our results show that small optimal gaps correspond to fast structural and weight plasticity with a high synaptic turnover rate $p_{e|0}$ and relative large $p_{d|0}$ with a high forgetting rate, whereas large gaps correspond to slow plasticity processes. This result has two implications: First, it may be used to explain the remaining discrepancy that in the psychological data the time gap depends on the retention interval, whereas in our model it does not: As preliminary simulations indicate, the experimental data could be reproduced by mixing (at least) two synapse populations with different sets of parameters, where they could be both within the same cortical area (stable vs. unstable synapses; cf., Holtmaat and Svoboda, 2009) or distributed to different areas (e.g., fast plasticity in the medial temporal lobe, and slower plasticity in neocortical areas). Moreover, as the temporal profile of optimal learning depends on parameters of structural plasticity, it may become possible in future experiments to link behavioral data

on memory performance to physiological data on structural plasticity in cortical areas where these memories are finally stored.

Although we have concentrated on analyzing one-step retrieval in feed-forward networks, our results apply as well to recurrent networks and iterative retrieval (Hopfield, 1982; Schwenker et al., 1996; Sommer and Palm, 1999): Obviously, all results on the temporal evolution of P_{eff} (including the results on the spacing effect) depend only on synapses having proper access to consolidation signals S_{ij} by either repeated rehearsal or memory replay, and therefore hold independently of network and retrieval type. However, linking P_{eff} to output noise (Equation 3) needs to assume a particular retrieval procedure. At least one-step retrieval is known to be almost equivalent for both feedforward and recurrent networks yielding almost identical output noise and pattern capacity M_ϵ (Knoblauch, 2008). Estimating retrieved information for pattern completion in auto-associative recurrent networks, however, requires to subtract the information already provided by the input patterns $\tilde{u}^\mu$. Here information storage capacity C is maximal if $\tilde{u}^\mu$ contains half of the one-entries (or information) of the original pattern u^μ, which leads to factor $1/2$ and $1/4$ decreases of M and C compared to hetero-association (cf., Equations 48, 49 for $\lambda = 1/2$; Palm and Sommer, 1992). Nevertheless, up to such scaling, our results demonstrating C increasing with P_{eff} are still valid. Similarly, our capacity analyses of M_ϵ and C_ϵ can also be applied to iterative retrieval by requiring that the one-step output noise level ϵ is smaller than the initial input noise $\tilde{\epsilon}$. As typically output noise $\hat{\epsilon}$ steeply decreases with input noise $\tilde{\epsilon}$ (cf. Equation 45), additional retrieval steps will drive $\hat{\epsilon}$ toward zero, with activity quickly converging to the memory attractor.

Our theory depends on the assumption that potential connectivity P_{pot} is significantly larger than anatomical connectivity P. This assumption may be challenged by experimental findings suggesting that cortical neuron pairs are either unconnected or have multiple (e.g., 4 or 5) instead of single synapses (Fares and Stepanyants, 2009) and the corresponding theoretical works to explain these findings (Deger et al., 2012; Fauth et al., 2015b). For example, Fauth et al. (2015a) predict that narrow distributions of synapse numbers around 4 or 5 follow from a regulatory interaction between synaptic

and structural plasticity, where connections having a smaller synapse number cannot stably exist. If true this would mean that most potential synapses could never become stable actual synapses because the majority of potentially connected neuron pairs have less than 4 potential synapses (e.g., see Fares and Stepanyants, 2009, **Figure 1**). As a consequence, actual P_{pot} would be significantly lower than assumed in our work, perhaps only slightly larger than P, strongly limiting a possible increase of effectual connectivity P_{eff} by structural plasticity. On the other hand, the data of Fares and Stepanyants (2009) are based only on neuron pairs having very low distances ($< 50\,\mu m$), whereas our model rather applies to cortical macrocolumns where most neuron pairs have much larger distances. Thus, unlike Fauth et al. (2015a), our theory of structural plasticity increasing effectual connectivity and synaptic storage efficiency predicts that neuron pairs within a macrocolumn should typically be connected by a much smaller synapse number (e.g., 1 or perhaps 2).

AUTHOR CONTRIBUTIONS

Conceived, designed, and performed experiments: AK. Analyzed the data: AK, FS. Contributed simulation/analysis tools: AK. Wrote the paper: AK, FS.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnana.2016.00063

REFERENCES

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge: Cambridge University Press.

Arshavsky, Y. (2006). "The seven sins" of the hebbian synapse: can the hypothesis of synaptic plasticity explain long-term memory consolidation? *Progress Neurobiol.* 80, 99–113. doi: 10.1016/j.pneurobio.2006.09.004

Attwell, D., and Laughlin, S. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metabol.* 21, 1133–1145. doi: 10.1097/00004647-200110000-00001

Aviel, Y., Horn, D., and Abeles, M. (2005). Memory capacity of balanced networks. *Neural Comput.* 17, 691–713. doi: 10.1162/0899766053019962

Bliss, T., and Collingridge, G. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361, 31–39.

Bosch, H., and Kurfess, F. (1998). Information storage capacity of incompletely connected associative memories. *Neural Netw.* 11, 869–876.

Braitenberg, V., and Schüz, A. (1991). *Anatomy of the Cortex. Statistics and Geometry.* Berlin: Springer-Verlag.

Buckingham, J., and Willshaw, D. (1992). Performance characteristics of the associative net. *Network* 3, 407–414.

Buckingham, J., and Willshaw, D. (1993). On setting unit thresholds in an incompletely connected associative net. *Network* 4, 441–459.

Butz, M., Wörgötter, F., and van Ooyen, A. (2009). Activity-dependent structural plasticity. *Brain Res. Rev.* 60, 287–305. doi: 10.1016/j.brainresrev.2008.12.023

Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a hebbian learning rule. *Ann. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.125639

Cepeda, N., Vul, E., Rohrer, D., Wixted, J., and Pashler, H. (2008). Spacing effects in learning: a temporal ridgeline of optimal retention. *Psychol. Sci.* 19, 1095–1102. doi: 10.1111/j.1467-9280.2008.02209.x

Chklovskii, D., Mel, B., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature* 431, 782–788. doi: 10.1038/nature03012

Clopath, C., Büsing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat. Neurosci.* 13, 344–352. doi: 10.1038/nn.2479

Cover, T., and Thomas, J. (1991). *Elements of Information Theory.* New York, NY: Wiley.

Crowder, R. (1976). *Principles of Learning and Memory.* Oxford: Lawrence Erlbaum.

Deger, M., Helias, M., Rotter, S., and Diesmann, M. (2012). Spike-timing dependence of structural plasticity explains cooperative synapse formation in the neocortex. *PLoS Comput. Biol.* 8:e1002689. doi: 10.1371/journal.pcbi.1002689

DePaola, V., Holtmaat, A., Knott, G., Song, S., Wilbrecht, L., Caroni, P., and Svoboda, K. (2006). Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. *Neuron* 49, 861–875. doi: 10.1016/j.neuron.2006.02.017

Dobrushin, R. (1959). General formulation of shannon's main theorem in information theory. *Ushepi Mat. Nauk.* 14, 3–104.

Ebbinghaus, H. (1885). *Über das GedÄchtnis: Untersuchungen zur Experimentellen Psychologie.* Leipzig: Duncker & Humblot.

Engert, F., and Bonhoeffer, T. (1999). Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* 399, 66–70.

Fares, T., and Stepanyants, A. (2009). Cooperative synapse formation in the neocortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16463–16468. doi: 10.1073/pnas.0813265106

Fauth, M., Wörgötter, F., and Tetzlaff, C. (2015a). Formation and maintenance of robust long-term information storage in the presence of synaptic turnover. *PLoS Comput. Biol.* 11:e1004684. doi: 10.1371/journal.pcbi.1004684

Fauth, M., Wörgötter, F., and Tetzlaff, C. (2015b). The formation of multi-synaptic connections by the interaction of synaptic and structural plasticity and their functional consequences. *PLoS Comput. Biol.* 11:e1004031. doi: 10.1371/journal.pcbi.1004031

Frey, U., and Morris, R. (1997). Synaptic tagging and long-term potentiation. *Nature* 385, 533–536.

Friston, K. (1994). Functional and effective connectivity. *Hum. Brain Mapp.* 2, 56–78.

Fu, M., and Zuo, Y. (2011). Experience-dependent structural plasticity in the cortex. *Trends Neurosci.* 34, 177–187. doi: 10.1016/j.tins.2011.02.001

Greene, R. (1989). Spacing effects in memory: evidence for a two-process account. *J. Exp. Psychol.* 15, 371–377.

Hebb, D. (1949). *The Organization of Behavior. A Neuropsychological Theory.* New York, NY: Wiley.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol. Cybernet.* 82, 111–121. doi: 10.1007/PL00007964

Holtmaat, A., and Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* 10, 647–658. doi: 10.1038/nrn2699

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558.

Knoblauch, A. (2008). Neural associative memory and the Willshaw-Palm probability distribution. *SIAM J. Appl. Mathemat.* 69, 169–196. doi: 10.1137/070700012

Knoblauch, A. (2009a). "The role of structural plasticity and synaptic consolidation for memory and amnesia in a model of cortico-hippocampal interplay," in *Connectionist Models of Behavior and Cognition II: Proceedings of the 11th Neural Computation and Psychology Workshop*, eds J. Mayor, N. Ruh and K. Plunkett (Singapore: World Scientific Publishing), 79–90.

Knoblauch, A. (2009b). "Zip nets: Neural associative networks with non-linear learning," in *HRI-EU Report 09-03, Honda Research Institute Europe GmbH, D-63073* (Offenbach/Main).

Knoblauch, A. (2010a). Bimodal structural plasticity can explain the spacing effect in long-term memory tasks. *Front. Neurosci. Conference*

Abstract: Computational and Systems Neuroscience. 2010, doi: 10.3389/conf.fnins.2010.03.00227

Knoblauch, A. (2010b). "Efficient associative computation with binary or low precision synapses and structural plasticity," in *Proceedings of the 14th International Conference on Cognitive and Neural Systems (ICCNS)* (Boston, MA: Center of Excellence for Learning in Education, Science, and Technology (CELEST)), 66.

Knoblauch, A. (2010c). "Structural plasticity and the spacing effect in willshaw-type neural associative networks," in *HRI-EU Report 10-10, Honda Research Institute Europe GmbH, D-63073* (Offenbach/Main).

Knoblauch, A. (2011). Neural associative memory with optimal bayesian learning. *Neural Comput.* 23, 1393–1451. doi: 10.1162/NECO_a_00127

Knoblauch, A. (2016). Efficient associative computation with discrete synapses. *Neural Comput.* 28, 118–186. doi: 10.1162/NECO_a_00795

Knoblauch, A., Hauser, F., Gewaltig, M.-O., Körner, E., and Palm, G. (2012). Does spike-timing-dependent synaptic plasticity couple or decouple neurons firing in synchrony? *Front. Comput. Neurosci.* 6:55. doi: 10.3389/fncom.2012.00055

Knoblauch, A., Körner, E., Körner, U., and Sommer, F. (2014). Structural plasticity has high memory capacity and can explain graded amnesia, catastrophic forgetting, and the spacing effect. *PLoS ONE* 9:e96485. doi: 10.1371/journal.pone.0096485

Knoblauch, A., and Palm, G. (2001). Pattern separation and synchronization in spiking associative memories and visual areas. *Neural Netw.* 14, 763–780. doi: 10.1016/S0893-6080(01)00084-3

Knoblauch, A., Palm, G., and Sommer, F. (2010). Memory capacities for synaptic and structural plasticity. *Neural Comput.* 22, 289–341. doi: 10.1162/neco.2009.08-07-588

Kononenko, I. (1989). Bayesian neural networks. *Biol. Cybernet.* 61, 361–370.

Krone, G., Mallot, H., Palm, G., and Schüz, A. (1986). Spatiotemporal receptive fields: a dynamical model derived from cortical architectonics. *Proc. R. Soc. Lond. B* 226, 421–444.

Lansner, A., and Ekeberg, O. (1989). A one-layer feedback artificial neural network with a Bayesian learning rule. *Intern. J. Neural Syst.* 1, 77–87.

Latham, P., and Nirenberg, S. (2004). Computing and stability in cortical networks. *Neural Comput.* 16, 1385–1412. doi: 10.1162/089976604323057434

Laughlin, S., and Sejnowski, T. (2003). Communication in neuronal networks. *Science* 301, 1870–1874. doi: 10.1126/science.1089662

Lennie, P. (2003). The cost of cortical computation. *Curr. Biol.* 13, 493–497. doi: 10.1016/S0960-9822(03)00135-0

Marr, D. (1971). Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. Ser. B* 262, 24–81.

Martin, S., Grimwood, P., and Morris, R. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis. *Ann. Rev. Neurosci.* 23, 649–711. doi: 10.1146/annurev.neuro.23.1.649

O'Connor, D., Wittenberg, G., and Wang, S.-H. (2005). Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9679–9684. doi: 10.1073/pnas.0502332102

Palm, G. (1980). On associative memories. *Biol. Cybernet.* 36, 19–31.

Palm, G., Knoblauch, A., Hauser, F., and Schüz, A. (2014). Cell assemblies in the cerebral cortex. *Biol. Cybernet.* 108, 559–572. doi: 10.1007/s00422-014-0596-4

Palm, G., and Sommer, F. (1992). Information capacity in recurrent McCulloch-Pitts networks with sparsely coded memory states. *Network* 3, 177–186.

Paulsen, O., and Sejnowski, T. (2000). Natural patterns of activity and long-term synaptic plasticity. *Curr. Opin. Neurobiol.* 10, 172–179. doi: 10.1016/S0959-4388(00)00076-3

Petersen, C., Malenka, R., Nicoll, R., and Hopfield, J. (1998). All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl. Acad. Sci. U.S.A.* 95, 4732–4737.

Poirazi, P., and Mel, B. (2001). Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron* 29, 779–796. doi: 10.1016/S0896-6273(01)00252-5

Raisman, G. (1969). Neuronal plasticity in the septal nuclei of the adult rat. *Brain Res.* 14, 25–48.

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/PhysRevLett.85.461

Schwenker, F., Sommer, F., and Palm, G. (1996). Iterative retrieval of sparsely coded associative memory patterns. *Neural Netw.* 9, 445–455.

Shannon, C., and Weaver, W. (1949). *The Mathematical Theory of Communication.* Urbana/Chicago: University of Illinois Press.

Sommer, F. (2000). On cell assemblies in a cortical column. *Neurocomputing* 32–33, 517–522. doi: 10.1016/S0925-2312(00)00207-1

Sommer, F., and Palm, G. (1999). Improved bidirectional retrieval of sparse patterns stored by Hebbian learning. *Neural Netw.* 12, 281–297.

Song, S., Miller, K., and Abbott, L. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neurosci.* 3, 919–926. doi: 10.1038/78829

Sporns, O. (2007). Brain connectivity. *Scholarpedia* 2:4695. doi: 10.4249/scholarpedia.4695

Steinbuch, K. (1961). Die Lernmatrix. *Kybernetik* 1, 36–45.

Stepanyants, A., Hof, P., and Chklovskii, D. (2002). Geometry and structural plasticity of synaptic connectivity. *Neuron* 34, 275–288. doi: 10.1016/S0896-6273(02)00652-9

Waydo, S., Kraskov, A., Quiroga, R., Fried, I., and Koch, C. (2006). Sparse representation in the human medial temporal lobe. *J. Neurosci.* 26, 10232–10234. doi: 10.1523/JNEUROSCI.2101-06.2006

Willshaw, D., Buneman, O., and Longuet-Higgins, H. (1969). Non-holographic associative memory. *Nature* 222, 960–962.

Witte, S., Stier, H., and Cline, H. (1996). *In vivo* observations of timecourse and distribution of morphological dynamics in Xenopus retinotectal axon arbors. *J. Neurobiol.* 31, 219–234.

Wyner, A. (1978). A definition of conditional mutual information for arbitrary ensembles. *Inform. Control* 38, 51–59.

Xu, T., Yu, X., Perlik, A., Tobin, W., Zweig, J., Tennant, K., Jones, T., and Zuo, Y. (2009). Rapid formation and selective stabilization of synapses for enduring motor memories. *Nature* 462, 915–919. doi: 10.1038/nature08389

Yang, G., Pan, F., and Gan, W.-B. (2009). Stably maintained dendritic spines are associated with lifelong memories. *Nature* 462, 920–924. doi: 10.1038/nature08577

Yu, X., and Zuo, Y. (2011). Spine plasticity in the motor cortex. *Curr. Opin. Neurobiol.* 21, 169–174. doi: 10.1016/j.conb.2010.07.010