Computational principles of memory

Rishidev Chaudhuri & Ila Fiete

The ability to store and later use information is essential for a variety of adaptive behaviors, including integration, learning, generalization, prediction and inference. In this Review, we survey theoretical principles that can allow the brain to construct persistent states for memory. We identify requirements that a memory system must satisfy and analyze existing models and hypothesized biological substrates in light of these requirements. We also highlight open questions, theoretical puzzles and problems shared with computer science and information theory.

'Memory' refers to any of the panoply of changes in the activity or connectivity of neural systems that are triggered by stimuli or brain states and then persist over a duration longer than the triggering events. Memory is adaptive: an agent with memory can learn from experience[1–3], generalize faster[4,5], recall prior information to better infer or predict with incomplete data[6,7], and perform other computations that depend on accumulating information over time, including integration[8–10], search[11] and decision making[12]. Many of these functions require elements beyond memory, but memory—defined and discussed narrowly in this Review as the persistence of states over time—is an essential component. Here we provide a glimpse, from the computational and theoretical perspective, of some of the principles, mechanisms and biological substrates that are believed to form the building blocks of neural memory. We focus on the question of memory maintenance, while also discussing questions that perplex us.

It is traditional to partition memory into short-term and long-term varieties, with varying definitions for the boundary[13]. One distinction is based on persistence time, with short-term memory (STM) being defined as lasting for seconds to tens of seconds and long-term memory (LTM) being defined as lasting for hours to decades. Another is that STM refers to sustained changes in activity while LTM refers to changes in the presence of connections and the strengths of the corresponding synapses[14–16] between neurons. From a dynamical perspective, it is unclear how sharp the boundary between STM and LTM is, and some of the central computational challenges associated with maintaining states over time are similar, whether 'time' refers to seconds or decades and whether 'state' refers to activity or structure.

What features might be desirable in any memory storage system? First, by definition, the system should possess states that can persist over time. Second, it should contain a sufficient 'capacity' or number of states, with the capacity scaling up in an appropriate (efficient) way with resources consumed. This capacity constraint might be very different for STM and LTM. Third, different inputs to be remembered should trigger the persistence of different memory states. Fourth, the states should be robust to noise, and concurrently stored memories

should not seriously interfere. Finally, the stored memories should be correctly retrievable, given appropriate cues.

State variables in the brain are too numerous to list; examples include neural voltage, synaptic activation, synaptic strengths, synaptic connections, states of presynaptic vesicles, phosphorylation levels, mRNA concentrations, transcriptional regulation, neuromodulatory signals, and the entire complement of states in glia and other constituent cells. In principle, any or all of these may be pressed into service as memory states; many have indeed been shown to be, and we will discuss how some of these may be made to persist over time.

Why is setting up memory states a hard problem? Biophysical quantities have intrinsic timescales over which they decay to some baseline level if not reactivated. Membrane time constants are milliseconds to tens of milliseconds[17], postsynaptic potentials last for tens to hundreds of milliseconds[18] and certain facilitation processes can last for several hundred milliseconds[19]. However, behavioral timescales for STM are on the order of tens of seconds to minutes. With respect to structural and other forms of LTM, molecules turnover with typical protein lifetimes on the order of days[20], whereas memories can persist for years. Thus, in both LTM and STM, the brain must construct states that persist using, as we understand so far, comparatively memory-less substrates (**Figs. 1–3**).

Next, noise is ubiquitous in the brain. Synapses transmit signals with some probability of failure[21]. Neurons receive fluctuating inputs and spike stochastically[22,23]. Copy numbers of important proteins and ionic species can be small in individual spines and boutons[24], permitting large fluctuations. Even with persistent states, noise can drive the system into a non-persistent regime or into the wrong persistent states (**Fig. 4**).

Finally, the desirables of memory are in opposition; trying to simultaneously satisfy them involves tradeoffs. For robust recovery, distinct memory states should be well-separated so that states with noise can be mapped to the correct memory. But robustness imposes a capacity constraint: fewer well-separated states will fit in a fixed representational space (**Fig. 5**). Alternatively, memory may be made both robust and high-capacity using advanced error-correction as in communications theory[25–27], but this requires complex encoding and decoding[26] (**Fig. 6b**) that might be neurally implausible. Slow biophysical processes can generate long persistence times with less circuit fine-tuning, but they render the system relatively unresponsive to inputs[28]. Plasticity that is strong enough for rapid learning of new information

Center for Learning and Memory and Department of Neuroscience, The University of Texas at Austin, Austin, Texas, USA. Correspondence should be addressed to I.F. (ilafiete@mail.clm.utexas.edu).

Figure 1 Stable states from positive feedback. (**a**) A system with sigmoid-shaped positive feedback (black) can exhibit bistability: when positive feedback exceeds intrinsic decay (gray), total activation is pushed up(right)ward ('+' or green areas and arrows), and vice versa when decay dominates ('−' or red areas and arrows). White circles mark stable states or attractors. (**b**) An exponentially decaying unit (decay proportional to activation) with self-excitation (strength W) and a saturating response can produce the dynamics shown in **a**. (**c**) Bistability with an alternative nonlinearity. For weak input, the response is small. Above a threshold, activity increases linearly and then saturates. In general, nonlinear feedback with initial acceleration followed by saturation can produce bistability. (**d**) Presented as in **b** with the unit replaced by a homogeneous population. Shaded cube, the state space or activation space of all units. r_1 represents activation level of unit 1, etc. The vector of activation of all units is a point in state space. The high and low activity states are shown as white circles. (**e**,**f**) Feedback tuned to cancel decay across a range of activations, producing a continuum (a line attractor) of stable states, shown in **f**. (**g**,**h**) Variant of **e** in which the continuum of states forms a ring (shown in **h**). Here, neurons physically arranged on a ring excite immediate neighbors (top, green) and inhibit all other neurons (orange). Synapses are shown from one neuron. The stable activity profile is a bump and all shifts of the bump around the ring (bottom, three stable states plotted over each other).
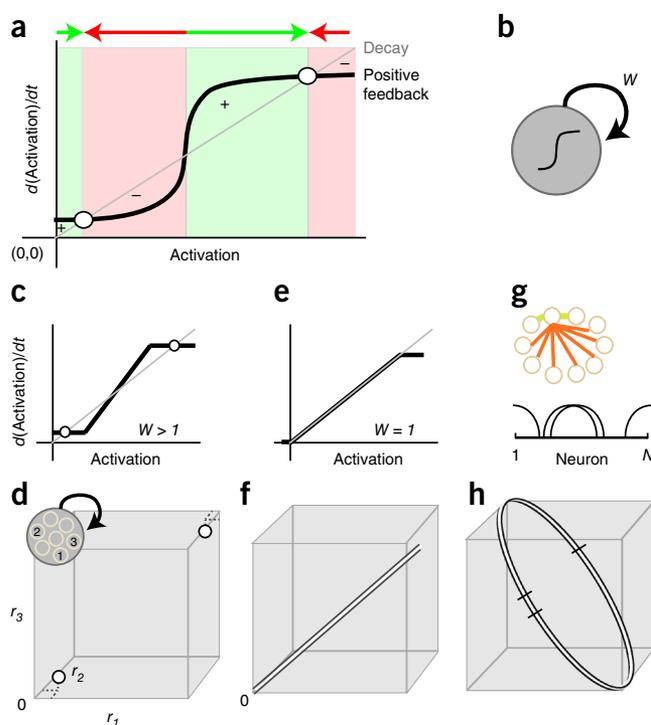
causes rapid overwriting of existing memories[29–31]. To learn where the brain is situated on these tradeoffs is to better understand the imperatives of neural memory. Here we discuss mechanisms for the generation of persistent states, their robustness and capacity, the biological evidence for these mechanisms, and the major differences in neural and von Neumann computer architectures for memory.

Creating persistence from memory-less components

Most models of STM rely on some form of persistent activity. Experimentally, persistent activity has been observed in various cortical areas during the delay periods of a range of memory tasks and it correlates with task load and memory performance in a way that allows the remembered item to be decoded[32–34]. The substrate of LTM is persistent or lasting synaptic change[35], according to the central dogma of neuroscience, which dates back to Ramón y Cajal[36] and others[37,38]. Indeed in experiments, LTM is associated with stable changes in synaptic weights and structure, and memory acquisition is impaired by blocking long-term plasticity (LTP)[39,40].

How is it possible to create long-lasting states for neural memory? To persist beyond the time constants of the constituent elements, states must be stabilized by circuit interactions. As we discuss next, circuit mechanisms can involve across-neuron signaling (STM) or molecular interactions in a synapse (LTM), and can produce diverse persistent responses, including fixed states, slowly decaying states and dynamic trajectories.

Positive feedback. Positive feedback is a general principle for generating persistent states from inherently non-persistent variables (**Fig. 1**). Units in a circuit that excite one another allow excitation to persist beyond the duration of the exciting stimulus, as in the 'reverberating activity loops' proposed by Lorente de Nó and Hebb[38,41].

Discrete memory states. Non-persistent, nonlinear units that strongly excite themselves or each other can, once activated, maintain a high-activity or 'up' state (**Figs. 1** and **2d**), whereas the 'down' state remains stable if initial activity is low; input pulses can drive transitions between these stable states.

In engineering, this is how flip-flops or switches are constructed from leaky capacitive elements[42] and is the basis for static random-access memory (static RAM)[43,44]. Autocatalytic molecular reactions are a form of positive feedback that have been proposed to support bistable switch dynamics[45,46], as are strong synaptic interactions between neurons[47].

More generally, strong feedback through excitatory connections can support multiple discrete stable states (called 'attractors'), each a different distributed pattern of activation across the neurons (**Fig. 2c**), as shown by Hopfield and others[48,49]. Surprisingly, and despite the pervasive influence of the discrete attractor Hopfield network in how we think about memory in neural systems, it is hard to identify uncontroversial examples of discrete attractors in the brain that persist in the absence of stimuli.

A continuum of memory states. Positive feedback, when carefully adjusted, can alternatively produce a continuum of stable states[9,10,50,51]. If the excitatory drive from positive feedback balances the intrinsic decay of the state, and if this balance can be achieved for a range of states, the system can maintain a continuum of states (**Figs. 1f,h** and **2e**) and use them to store an analog variable[8–10,51].

In general, these 'continuous attractors' emerge only when circuit interactions have some symmetry or other fine-tuning. Consider a ring of neurons, each with the same strong excitatory drive to immediate neighbors and inhibition to the rest (**Fig. 2e**)[10,50]. The stable state on the ring is an activity bump and all of its translations (rotations), defining a one-dimensional continuous attractor. Neural plasticity could form and calibrate such structured interactions[52,53], and intrinsic bistability could help to stabilize such networks[54,55], but usually some fine-tuning requirements remain. How the brain might solve this problem is an open theoretical and experimental question.

Experimental evidence suggests that the brain may use continuous attractors in multiple brain systems, both cortical and subcortical, including the oculomotor, head direction, grid cell and prefrontal working memory circuits[8,9,56–60].

Persistent states from inhibitory interactions. Positive feedback can arise from mutual inhibition rather than excitation: units that inhibit one another effectively disinhibit themselves. Thus, combining an excitatory feedforward drive with mutual inhibition can produce switching dynamics[61] (**Fig. 2d**). Similarly, circuits with all-inhibitory recurrent interactions can form continuous attractors[51,62,63].

Figure 2 Circuit mechanisms for persistent states. Top row, schematics of various architectures (green, excitatory interaction; orange, inhibitory; line weight proportional to strength). Shaded cubes show state space of network population activity. (**a**) Non-interacting units. Any state decays to the sole fixed point at 0 with the short intrinsic time constant of each unit. Traces at bottom show activity of three units over time in response to a transient input (input duration shown by thin red bar below). (**b**) Moderate-strength positive- or negative-derivative feedback can produce long transients. (**c**) Strong structured recurrent feedback (as in a Hopfield network) can generate multiple discrete attractors (stable states; white circles). Each corresponds to one pattern of population activity (row) at bottom, with the all-black pattern representing 0. (**d**) Bistable switches can be implemented by two mutually exciting pools or by mutually inhibiting ones (combined with a uniform excitatory drive), and can be viewed as a special case of a Hopfield network. Fixed points of the mutual-excitation switch are on the main diagonal; mutual-inhibition fixed points are on the cross-diagonal. Bottom, activity of two units from different pools of the mutual inhibition network (gray, black at bottom) and inputs to these pools (short gray, black pulses, top). (**e**) Highly structured networks with symmetries or fine tuning in weights can generate continuous attractors via positive feedback or negative derivative feedback. (**f**) Complex trajectories and responses can be generated in networks with feedforward and non-normal architectures such as synaptic chains (depicted here). Traces at bottom show population activity at different points in time.
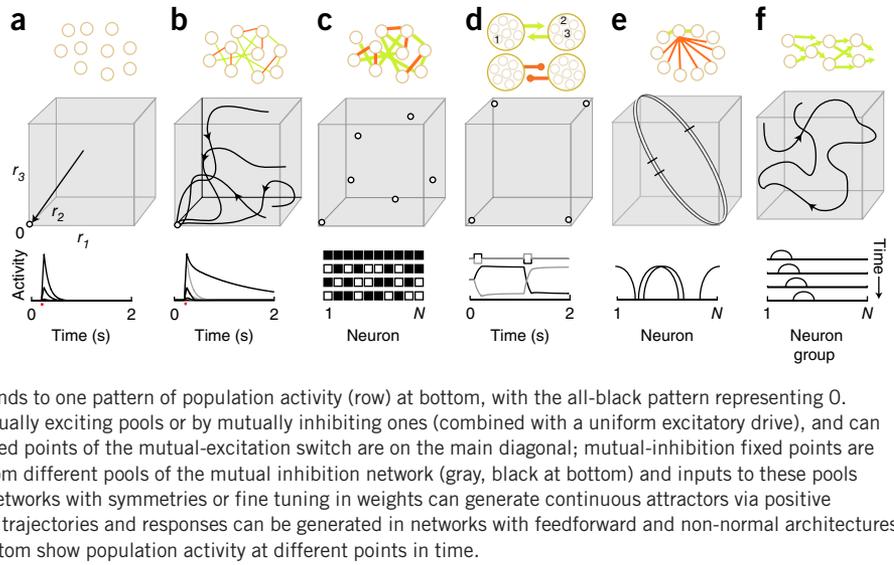
LTM maintenance. How is a long-term memory, once acquired, maintained over months and years, given that protein lifetimes are typically on the order of days? A widely hypothesized possibility is that LTM stability arises from synaptic stability (see below for alternatives). Positive feedback is often discussed for STM, but the concepts are general and relevant to LTM. Autocatalytic reactions or mutually exciting or inhibiting molecular interactions can stabilize intrinsically non-persistent molecular states[45,64,65], allowing them to serve as the synaptic maintenance signal (**Fig. 3**).

If synaptic strength were specified solely by the state of a single bistable molecular switch, a synapse could have only two sizes, unless different copies of the molecules are compartmentalized and independently switchable or if there exist multiple switches involving different molecular species with staggered switch thresholds. In the latter two cases, the synapse could have multiple discrete levels. Experimentally, leading candidates for maintaining synaptic state through positive feedback include calcium/calmodulin-dependent protein kinase II (CaMKII) and protein kinase Mζ (PKMζ).

CaMKII autophosphorylation acts as positive feedback, potentially allowing it to serve as a bistable switch[66]. Inhibiting the interaction of CaMKII with the NMDA receptor after LTP induction disrupts potentiation in a manner that persists after the inhibitor is removed[67,68], and disrupting CaMKII-NMDAR complexes has long-term effects on spatial learning[69]. Moreover, CaMKII may form a bistable switch with the translation factor CPEB[70]. However, not all CaMKII inhibitors, in the procedures tested, affect LTP maintenance[71].

The constitutively active protein kinase PKMζ inhibits a protein that suppresses its translation (positive feedback through mutual inhibition)[72–74]. The PKMζ inhibitor ZIP disrupts both LTP and memory maintenance[75], and models suggest that PKMζ can maintain its state once it is highly expressed[76,77]. However, ZIP is not specific to PKMζ, and intact LTP maintenance and memory in PKMζ knockout mice complicate PKMζ's role in memory maintenance[78,79], although the latter results might not apply to wild-type mice[80,81].

Clearly, the case remains open for both of these candidate mechanisms.

As in STM, feedback could generate a molecular maintenance signal of analog strength. If a synapse produces a number of molecules of a certain type in proportion to its current size (synapse size is proportional to strength; **Fig. 3b**)[82,83], and if these molecules capture

resources for synaptic maintenance, the process can correspond to tuned positive feedback. As discussed below, creating analog persistent states with positive feedback involves fine-tuning. Even if well-tuned, and more so if not, the states can drift over time—problematic over the long timescales of LTM—suggesting that maintenance is more likely to be based on discrete attractors. Synapse strengths may therefore be discretized. Electron microscopy results showing that different synapses between the same axon and dendrite have essentially the same volume seem to be consistent with a discrete set of sizes[84], but much more needs to be done to settle this question.

Negative derivative feedback. A system with strong excitatory coupling, similarly strong inhibitory feedback and slower excitatory than inhibitory interactions can also generate memory states[85,86]. The general principle is negative feedback control[87,88], commonly used by biological systems[89,90]. Fast-acting inhibitory couplings counteract changes of state, allowing the network to maintain the states it is placed into. Similar to positive feedback networks, a (neural or molecular) network with negative derivative feedback can exhibit stable states or very long transients, is responsive to inputs despite its long persistence times, and is in some ways more robust (see below). Experimentally, it is unclear whether observed continuous attractors are based on positive feedback, negative derivative feedback or a combination of both.

Long transients, feedforward structures and chaotic states. The above mechanisms, when well-tuned and in the absence of noise or other inputs, in principle allow memory to persist for infinite time. There are also more transient memory mechanisms.

Sufficiently strong positive feedback with particular architectures creates attractors, but more generally positive feedback creates 'slow modes', which exhibit activity decay at timescales slower than the biophysical time constants of the individual elements (**Fig. 2b**). These slowly decaying traces can be used to store inputs and decode them from the instantaneous output of the system[91–93].

Another way to prolong decay is through delay lines, which were used to construct memory in the earliest days of computing[43,94]. In the brain, a delay line can consist of neurons coupled in a feedforward architecture, each feeding activity into the next[95–98], or a cascade of subcellular molecular reactions[24,40]. The signal decays rapidly at

Figure 3 Long-term maintenance of synapse size. In both panels, green hexagons represent molecules synthesized at the synapse, whose role is to capture centrally trafficked resources (purple triangles) for synapse maintenance. (**a**) Synapse strength is specified by the state of a bistable molecular switch (orange); the state of the switch determines how many capturing molecules (green) are produced, and thus the synapse size. (**b**) Here, the synapse manufactures capturing molecules (green) in direct proportion to its size. These molecules capture synapse maintenance resources in proportion to their number; thus, the synapse can obtain resources in proportion to its size, for any given size. In both **a** and **b**, the soma can simply sum feedforward signals from the synapses to determine how many resource molecules to transcribe. Note that the persistent specification of synapse size, or the maintenance signal, must be localized in the synapse, as somatic changes alone cannot plausibly account for synaptic specificity. To be directly involved in generating a synapse-specific persistent maintenance signal, the soma would require at least 2^K persistent states (assuming the K synapses can independently take on one of two strengths), a biologically implausible scenario.
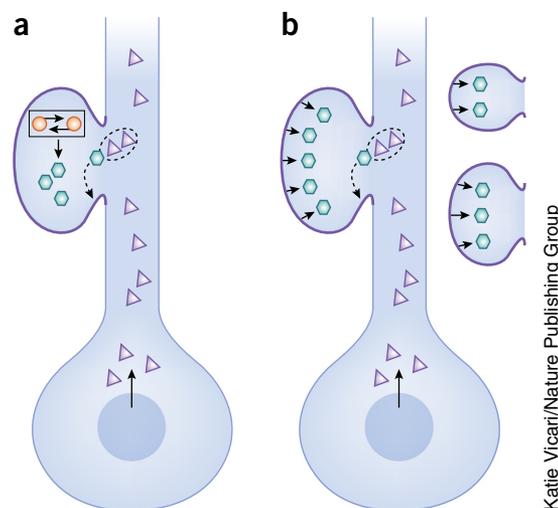
Katie Vicari/Nature Publishing Group

each node, but is passed along the chain (**Fig. 2f**), extending signal lifetime in proportion to the number of nodes. Delay lines are a special case of 'non-normal' networks[97,98], which include negative derivative feedback networks[85] and other so-called 'balanced' networks[99–101]. General balanced networks tend to generate fast transient responses, possibly in opposition to the desiderata for memory networks.

When strongly coupled, large non-normal and normal networks can generate complex sustained dynamical responses, including oscillations and chaos[100,102,103]. Whether these rich dynamics can reliably serve as STM remains an interesting theoretical question.

Experimentally, long stimulus-dependent trajectories are observed in multiple neural systems[104,105], but it is unclear whether these trajectories actually subserve memory or simply move the system from one state to another for other purposes.

Stable or slow biophysical processes. Let us briefly return to the assumption that memory is created out of relatively memory-less components. What are the computational implications of constructing STM and LTM out of fundamentally persistent or slowly decaying elements?

Slow synaptic dynamics for STM maintenance. Many biophysical timescales in the brain are short compared with persistence times for STM, but longer time constants at the cellular and synaptic levels, including synaptic facilitation (τ of tens to hundreds of milliseconds), post-tetanic potentiation (τ of one to tens of seconds) and calcium-sensitive cation currents (tens of seconds to minutes) might help to stabilize persistent neural activity states[19,106–108].

However, a drawback of STM models based on slow biophysical timescales τ is that, unlike those based on positive feedback with faster elements, they are slow to respond to input changes. A unit-strength input applied for duration Δt produces a change in the persistent state proportional to $\Delta t / \tau$; to elicit a unit change in state, the input would have to persist for the long duration τ. By contrast, positive feedback–stabilized states built with fast elements permit relatively fast changes of state, even if the effective network time constant for memory is very long or divergent. On the other hand, the stability of the oculomotor integrator to perturbation[109] and the sensitivity of positive feedback–stabilized circuits of fast elements raises the possibility that slower biophysical time-scales may be important in STM maintenance[108,110].

Intrinsic biochemical multistability for LTM maintenance. LTM maintenance at synapses could also involve slow states, in the form of intrinsic molecular multi-stability. A molecule with two low-energy

states can, if the energy barrier between the states is high enough, persist in either state over very long timescales; these states can serve as markers of a potentiated synapse. However, this molecule would either have to avoid turnover and retain its state for the memory life-time (indeed there are long-lived proteins with lifetimes of a decade or more, but these might contribute to other problems, such as aging from disrepair[111]) or be able to transmit its state to new copies.

Consistent with the latter possibility, the prion-like cytoplasmic polyadenylation element–binding protein (CPEB) has been proposed to be the synaptic maintenance signal[40,112]. CPEB, which is activated by NMDA signaling and regulates protein translation in dendrites, exists in two stable folded conformations, one of which forms self-perpetuating multimers[112,113]. Experimentally, knocking down CPEB after memory consolidation disrupts previously stable hippocampal memory[114] and LTP, and CPEB mutations disrupt cerebellar learning[115].

Robustness to noise
Given the apparent ubiquity of noise in the brain[22,23], the plausibility of hypothesized memory mechanisms depends on their robustness. Here we consider the effects of noise in the state variables themselves (for example, neural activity and protein concentration) and in the parameters of the underlying system (for example, synaptic weights and reaction rates).

Discrete attractors are robust. For ongoing uncorrelated noise, bistable switches built from strong positive feedback have a probability of spontaneous switching that is exponentially small in the number of participating units[116,117]. Thus, the states powerfully resist noise and can recover the attractor from a perturbed network state. Strong positive feedback generally drives responses into the saturated regime of activation; thus, bistable switches are also robust against perturbations in the coupling strength. For sufficiently strong coupling, small changes in strength will not destroy or even substantially alter bistable states[118]. Similarly, other discrete attractor states (such as those in Hopfield networks) are robust to noise that is smaller than half the distance between any pair of attractor states (**Fig. 4a**): this noise cannot move the system to another attractor, and the dynamics regains the original attractor.

Continuous attractors are partially robust. Continuous attractors are a mathematical idealization: slight perturbations in network structure break the continuum into a string of stable points (**Fig. 4b**), much as a narrow stream of water is prone to necking off into closely

Figure 4 Robustness of persistent activity architectures. (**a**) In discrete attractor networks, states in the neighborhoods (light blue) of attractors (dark blue dots) decay to their attractors and are thus automatically corrected. One may view this process as memory retrieval from partial cues. If the image of the brain (far right, bottom) is an attractor, then a noisy input (second image from right) leads the network to recover the intact image. If noise is sufficiently large (second image from left), the network fails to recover the intact image, and converges to another memory state (first image). (**b**) Top, a continuous attractor corrects noise perpendicular to the attractor manifold (red arrows), but not noise along the manifold (green arrows), because this noise puts the system into another permissible state, corresponding to a different value of the represented variable. Bottom, sensitivity to structural noise. With special architectural symmetries or fine tuning, the set of stable states is continuous (flat energy valley, upper panel), but small perturbations to network architecture break the continuum into a set of closely spaced and thus quasi-continuous set of fixed points (lower panel). (**c**) When coding with long transients in a non-persistent system, the initial separation between states decays over time (converging blue traces). Thus, the ability of a given noise strength to confuse two states grows with time. (**d**) Complex trajectories in balanced networks can be surrounded by a small region that pulls states back onto the trajectory. However these regions seem to be vanishingly small with system size[119].
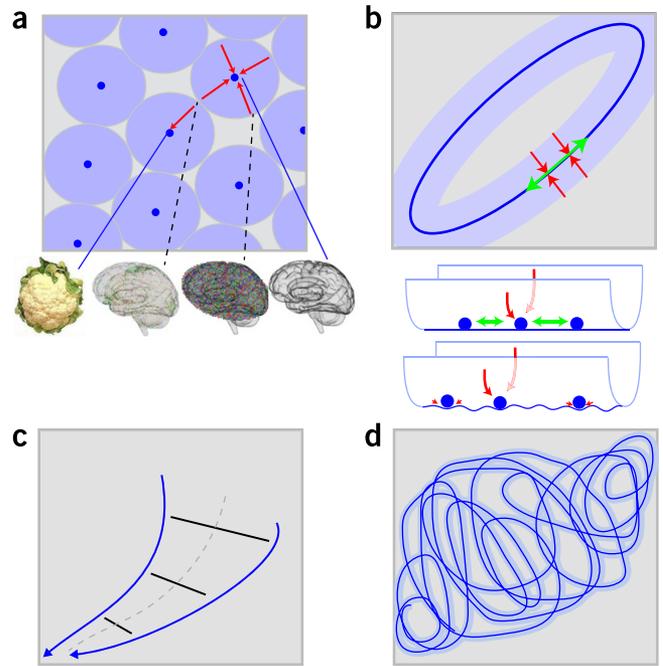
spaced droplets. However, narrowly spaced fixed points approximate a continuum, and henceforth we refer to both 'truly' continuous and quasi-continuous attractors formed from fixed points as continuous attractors.

Continuous attractors are not fully robust to ongoing noise: perturbations off the attractor manifold quickly decay away and are therefore corrected (**Fig. 4b**). However, components of noise along the manifold push the state to another stable point on the manifold (**Fig. 4b**). With ongoing noise, the state diffuses along the attractor[119,120], moving away from the initial state with squared distance growing linearly in time at a rate proportional to the attractor dimension and inversely proportional to the number of units in the network[120,121]. Approximately continuous attractors with some degree of discreteness along the attractor can resist diffusion resulting from low-amplitude noise[54,55], albeit at the cost of a rounding-off or discretization error in the represented variable[122]. However, they do not allow for integration of small, smoothly varying inputs (whose task is to systematically push the state along the attractor) and at the same time are not as robust as well-separated discrete attractors (such as those in Hopfield networks).

Robustness of negative derivative feedback networks. Negative derivative feedback networks are more resistant to small perturbations in interaction strength than positive feedback networks. Starting with a network tuned to have 100-fold longer persistence than the biophysical time constant τ, a 5% reduction in excitatory feedback will cause a fivefold drop in the persistence time of positive feedback networks, but only a 5% change for negative derivative feedback. This happens because τ is divided by a number close to zero in well-tuned positive feedback networks; a small change in feedback shifts the denominator away from zero, quickly collapsing the persistence time to $\sim\tau$. Negative derivative feedback networks, loosely speaking, multiply τ by a large number; small changes in excitatory strength result in similarly small percentage changes in this multiplier and thus the time constant.

Noise tolerance in networks with long transients. In positive feedback networks that encode information with long, decaying transients, network responses converge toward zero over time (**Fig. 4c**); thus, for temporally punctate inputs and ongoing noise, initial states

can begin well-separated, but become more easily confused through noise over time (although noise orthogonal to the slow directions decays away rapidly). Thus, noise contributes to information loss in addition to loss from intrinsic activity decay. In networks with feedforward or hidden feedforward structure, such as synaptic chains, the amount of (Fisher) information grows in proportion to network size and decreases with time, at best as the inverse of elapsed time[121].

Networks that generate complex trajectories[99,100] can exhibit attractive dynamics that correct small perturbations (**Fig. 4d**). However, in these statistically homogeneous random networks, the attracting region is small and shrinks with network size[123].
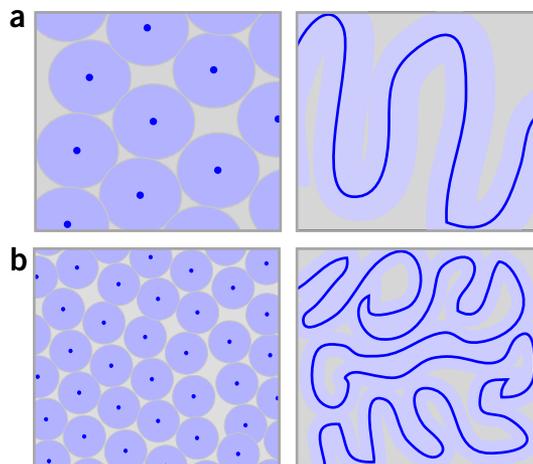
Memory capacity

Given a computational model of memory storage, how much information can be stored as a function of network size? This is the question of memory capacity. In real neural systems, what are the demands on STM and LTM capacity?

As we discuss below, a number of factors limit memory capacity. One is decay, if the memory states are not truly persistent over time. Another is noise, which can cause the stored memory state to jump to another (**Fig. 4**), and increasing noise tolerance via well-separated memory states reduces capacity (**Fig. 5**). The third is interference: even if existing memory states can stably persist over time and the stored states are robust to noise, if the number of inputs written into memory over the organism's lifetime exceeds the space available for memory, accommodating new memories requires overwriting old ones.

How much LTM is enough? Meaningful events and scenes in the world may result from only special combinations of elements, but if the number of elements is large this can quickly lead to a combinatorial explosion. LTM capacity, if it scales linearly with neurons or synapses, can be overwhelmed even if the number of neurons and synapses is very large. Theoretical estimates of the brain's LTM capacity vary wildly, from 10^9 to 10^{20} bits[124]. Empirically, human LTM for complex stimuli is large[125,126].

Figure 5 The tradeoff between capacity and robustness. (**a**) If memory states, whether discrete or continuous (dark blue in left, right, respectively), are well-separated, a decoder can robustly recover the state from a relatively large amount of noise because the neighborhoods of each memory state are large (shaded light blue regions). (**b**) Packing more memory states into the fixed state space volume of a given number of neurons necessarily means that the neighborhoods of each memory state will shrink. A small amount of noise drops the state into the neighborhood of a different memory state. Thus, higher capacity means less noise tolerance.

Capacity in LTM models. *Hopfield networks.* The Hopfield network[127] is a classic model of LTM storage and 'auto-associative' retrieval using relatively memory-less units. It can store the memory of a set of input patterns (a pattern is an arbitrary activity state of all the neurons) in its synaptic weights. For retrieval, the neural dynamics, using the learned weights and starting from partial or noisy inputs, flows to a stable activity state corresponding to the closest stored pattern (as in **Fig. 4a**). The network is called auto-associative because the tag to retrieve a state is the state itself, and it correctly retrieves memory from and denoises noisy cues.

General arguments show that a Hopfield network of N neurons can store ~N arbitrary stable patterns in its weights[128,129]. Specific learning rules, such as the associative Hopfield learning rule, in which a weight is strengthened whenever pre- and postsynaptic neurons express correlated activation, achieve this scaling (albeit with a constant < 1 prefactor[130]).

However, learning new inputs once capacity has been reached abruptly destroys the stability of all existing states, which can no longer be retrieved. Notably, adding bounds to the range of each synapse allows the network to function in a 'palimpsest' mode, where old memories gradually decay as new ones are learned; thus, catastrophic erasure from interference is avoided, but capacity (the number of readable memories at any given time) is lowered[131].

Palimpsest LTM networks. The palimpsest mode induces a tradeoff between retaining memories and writing in new information with high fidelity, which quickly overwrites older synaptic strengths. Several studies consider the joint limit on the lifetime and fidelity of memories in the palimpsest mode. Recent results show that if each bounded synapse consists of a set of persisting internal states (rather than just one scalar strength), with structured state transitions, then information can persist for a duration that scales almost linearly with the number of synapses in the network while allowing new memories to be acquired at high fidelity[31,132]. These results are about information contained directly in the synaptic weights; it is currently unclear how the bounds would change for plausible memory retrieval as in Hopfield networks.

High-capacity LTM systems. Information theory shows how to construct sets of states of length N ('codes') that robustly represent exponentially many states (~$e^{\rho N}$, where $0 < \rho < 1$)[25,26]. The theory of error-correcting codes does not, however, account for encoding and decoding (that is, denoising) costs. Usually, denoising good codes involves high complexity, with large costs either in space (many neurons) or time (slow, many neural time constants). In a neural system, encoding, storage, and denoising must all be done by neurons and would consume the same types of resources. The question is whether the brain can encode, store and denoise close to exponentially many states using linearly many neurons.

Interestingly, Hopfield networks of disjoint cliques or bistable switches can have nearly exponentially many stable states that are denoised by network dynamics[117,133]. However, these specially structured states do not correspond to arbitrary input patterns. It may be possible to store arbitrary patterns by mapping them onto these

structured states—in this view, the inputs are stored by 'hashing' them to the persistent states[134]—but this encoding may be computationally complex (**Fig. 6b**). It involves mapping exponentially many inputs that may not be well-separated to as many distinct well-separated memory states without a structure consistent with the represented inputs. It remains an open question whether the structure of natural inputs permits natural (low-complexity) mappings of this sort, to enable robust exponential or at least superlinear storage.

How much STM is enough? Behavioral evidence indicates that human STM capacity is fewer than ten items[135]. This is surprising, as general intelligence strongly correlates with STM performance[136]. It is unclear whether the stringent capacity limit is a result of constraints, such as limited resources, decay or interference, or whether it is a design feature. If the former, it remains unresolved which constraints are the limiting factors[137,138]. If the latter, STM might be optimized for easy encoding and retrieval, and clearing older inputs keeps the buffer clutter-free for fast access, although this objective might be better served by selective erasure of unnecessary items instead of nonspecific refreshing.

Capacity in STM models. The Hopfield network also functions as an STM for its learned patterns, as these patterns are stable states of the dynamics. The same is true for continuous attractor networks. In both cases, STM is encoded by activating a stable state and consists of the persistence of that state over time. In this setting, the capacity of the Hopfield network is again ~N, linear in the number of neurons in the network.

In continuous attractor networks, one cannot enumerate states to estimate capacity. Nevertheless, the range of the represented variable divided by the decoded error some time T after encoding gives the capacity of such networks. Defined this way, the capacity scales linearly with N and inversely with T[120,121]. The scaling with network size is consistent with Hopfield networks, but the degradation with time is an additional penalty incurred by (even well-tuned) continuous attractor networks.

Items in a sequence of input can be recovered from the instantaneous states of networks with long transients rather than fixed-point dynamics. The capacity of such networks, roughly defined as the summed memory of past inputs that can be recovered from the current state, grows linearly with N in linear networks[93,97] (and decays as $1/T^2$, where T is elapsed time). Under the same metric, the best linear networks for sequence memory with ongoing neural noise are organized in a single, maximally long (N neuron long) chain[97]. Networks

Figure 6 Complexity cost of storing a continuous variable in a set of well-separated discrete attractors. (**a**) Consider a continuous circular variable (such as the orientation of one's head relative to some external marker; ring at left); values of the variable can be naturally and continuously mapped onto a (quasi)continuous attractor of the same dimension and topology (shown on the right), preserving metric relationships between different values of the variable. The encoding is relatively simple, with the selection of a different storage state for a different value based on the change in the variable value. (**b**) To encode a continuous variable in a set of well-separated discrete stable states in some other coding dimension involves two steps: a discretization (first set of arrows) followed by the harder step of choosing how to map the discrete values into the attractors. In general, there is no metric-preserving mapping and the complexity of this encoding problem is high.
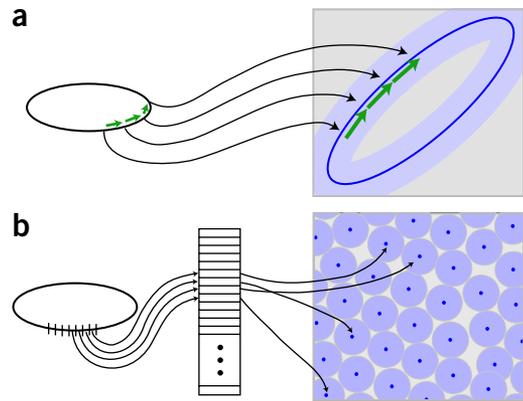
with transient dynamics can alternatively store a single input (rather than a sequence). Here again, performance improves linearly with N (and at short times decays as $1/T$)[121].

Discrete versus continuous attractors for memory?

Why might the brain use (quasi)continuous memory states given the capacity and robustness advantages of well-separated discrete states? Mapping an analog variable onto a continuous manifold of matching dimension can preserve metric information, with nearby values of the variable mapped to neighboring neural states (**Fig. 6**). With a neural representation that preserves the metric of the external analog variable (for example, using a (quasi)continuous set of memory states), it becomes possible for an input representing the time-derivative of the external variable to directly update the neural state to the new value of the external variable (**Fig. 6a**). Indeed, many systems modeled as continuous attractors are hypothesized or established 'integrators'.

Model mechanisms: questions and tests

In view of the computational principles reviewed here, it is clear that much work is required to establish which mechanisms are actually used for STM and LTM in the brain.

Questions and tests for STM. Is persistent activity the basis of STM, and if so, what is the relative role of circuit versus slow subcellular mechanisms? Some subjects of 'absence' (*petit mal*) seizures can process questions and resume conversations after several seconds of abnormal synchronized brain activity[139]. This observation could, in principle, argue against STM mechanisms based on short biophysical time constants and circuit feedback–stabilized activity, but the data are not conclusive, as seizures are often localized and may not have disrupted activity in the relevant areas. Blocking activity for varying lengths of time (for instance, optogenetically) in relevant areas and then measuring the degree to which memory recovers could help to resolve this issue. Another possibility is to measure the gain in the persistent component of neural responses after receiving a transient input: systems relying on long biophysical timescales show small gains (see above).

Neurons believed to be involved in STM can exhibit evolving patterns of activation during the delay period[140]. These data may suggest that the circuit operates with transient dynamics or feedforward structures (**Fig. 2b,f**); alternatively, they could reflect a stable manifold, with the overall state freely flowing along the manifold, whereas a lower-dimensional projection of network activity remains fixed and supports the memory[141]. An experimental probe might use targeted perturbations (such as patterned photostimulation[142]) to find directions that disrupt memory maintenance.

In STM psychophysics, does degradation in recall performance result from temporal decay or interference[137,138]? The former implies

a continuous attractor or a transient memory system in which information is gradually lost over time. The latter implies a palimpsest-like memory in which memories are temporally stable without external drive, but are overwritten by new inputs. Varying delay periods while keeping memory load fixed and vice versa and quantitatively measuring performance can help to resolve the underlying mechanisms.

Finally, despite the tremendous influence of discrete attractor models of neural memory[127], the superior stability and noise tolerance of discrete attractors relative to continuous attractors, and the apparent prevalence of continuous attractors in the brain, there has been no clear identification of neurons involved in discrete attractor dynamics. Perceptual bistability seems to be a compelling example[143], but the neural loci have not been found. Neural states can transition abruptly between two levels[144], but it remains to be shown that the apparent bistability is intrinsic to the network and not a consequence of changing feedforward inputs.

Questions and tests for LTM. The link between synaptic change and LTM is becoming increasingly well-established. Nevertheless, basic questions about the synaptic and circuit-level substrates of LTM remain unanswered.

What molecules constitute the LTM maintenance signal in synapses? Identification of the maintenance signal, among the candidates discussed above and others, will answer whether the mechanism is an intrinsically stable molecular state (with no turnover and thus possible problems with aging) or a self-propagating molecular state (with positive feedback, via either autocatalytic or prion-like properties).

Do individually stable synapses form the substrate of LTM or is LTM a circuit-level property with drifting single-synapse contributions? Synapses in *in vitro* brain slices turnover and change in strength[145,146] over tens of days, raising the specter of *in vivo* turnover. In the latter view, memory would consist of some low-dimensional projection of the synaptic states, so that many different strength and connectivity combinations could result in the same projection, similar in spirit to a proposal discussed above for STM[141]. By moving among these synaptic states, individual synapses would fluctuate, but memory could remain invariant. For this to be possible, there must be a mechanism continually pushing the network to preserve the same lower dimensional memory as synapses vary. Reconsolidation on the basis of recall might be such a process[147], but if recall is spontaneous, then the result is a positive-feedback process that could overwhelmingly favor a few strong memories and delete the rest.

To determine whether memory is sustained by the long-term stability of its substrates (synapses) or must instead be repeatedly re-instantiated as for RAM in computer systems (see below)[44], is important, but extremely difficult. A blue-sky experiment would characterize whether

there is a subspace of synaptic configurations that correspond to a stable memory by tracking the synaptic configurations over time while also measuring the behavioral retention of the memory. However, measuring retention induces recall, which in turn must affect the stored memory[147]. It is probably also impossible to avoid the acquisition of new memories during this time, which would complicate the characterization of invariant subspaces for a given set of memories.

How does new information interact with old in LTM, and does LTM operate in a palimpsest mode? Inducing new learning and then observing how memory degrades (behaviorally) and neural representations change can elucidate how easily neural memories get overwritten. For instance, how do hippocampal representations of familiar environments change after training the animal on several novel environments, as compared with merely waiting as long to retest representations in the familiar environment? More ambitiously, optical stimulation experiments *in vivo*[142,148] could induce multiple new memories using a common set of synapses while measuring how old memories disappear.

Biological versus computer memory

Biological and computer memory differ in multiple ways; we highlight three. First, computer hardware is separate from software, and memory from computation. By contrast, structure and activity are inextricably intertwined in the brain. Patterns of activity that underlie a computation naturally get encoded into LTM via activity-dependent plasticity. Conversely, these memories alter the computations a system can perform. Unlike computer memory, which can be accessed repeatedly without change, retrieval in the brain can reshape the accessed memory[147].

Second, computer memory is accessed by abstract indices generally unrelated to memory content, whereas the brain is believed to work with content-addressable LTM, in which pieces of the stored item can trigger recall of the full memory by completing partial associations.

Third, computers assign segregated locations to different memories, whereas memory storage is parallel, distributed and overlapping in the brain: a set of synapses is thought to participate in multiple memories, and a given memory is distributed across a network. These differences can lead to problems with interference and robustness in the brain, but putting memory and computation in the same place may permit rapid, flexible computation at a lower energy cost and without the wait times to access memory that are a severe constraint on von Neumann computer architectures.

Nevertheless, biological and artificial memory share some key theoretical principles. The STM and LTM distinction may correspond to the primary (RAM) versus secondary (hard drive) memory distinction in computers. Broadly, both STM and RAM require a circuit into which information can be repeatedly and rapidly inserted and retrieved, and the mechanisms for achieving this involve constant expenditure of energy. Computers can continually refresh the decaying states of capacitors, as in dynamic RAM, or use switches stabilized by positive feedback between decaying elements, as in static RAM. Parallels in biological memory are rehearsal strategies in working memory (such as dynamic RAM) and the positive feedback circuits believed to maintain persistent activity (such as static RAM). In both schemes, to counteract decay, the rate of spiking must exceed the inverse biophysical time constant of the neurons or synapses. Interestingly, there exist fast RAM-like mechanisms for computer memory that do not require power for maintenance, such as flash memory. However, the physical device degrades over time.

As with hard drives and RAM, LTM based on synaptic stability may require much less energy to maintain than STM (as it does not

involve neural spiking), is slower to induce, and may exploit more sophisticated encoding and decoding schemes to protect against errors. Although maintaining a persistent synaptic molecular signal might be energetically cheap (or free, if relying on molecular bistability), synapses themselves require energy to maintain, unlike hard drive magnetization.

For excellent reviews on more complex memory phenomena than the basic question of persistence of information over time that we have focused on, including molecular substrates, how to selectively activate neurons on the basis of which items must be put into STM, the human psychophysics of memory, and differences in hippocampal and cortical memory, see refs. 40,149,150.

1. Pavlov, I.P. *Conditioned Reflexes* (Oxford University Press, London, 1927).
2. Carew, T.J., Walters, E.T. & Kandel, E.R. Classical conditioning in a simple withdrawal reflex in *Aplysia californica. J. Neurosci.* **1**, 1426–1437 (1981).
3. Rescorla, R.A. & Wagner, A.R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. in *Classical Conditioning: Current Research and Theory* (Appleton-Century-Crofts, 1972).
4. Mackintosh, N.J. *The Psychology of Animal Learning* (Academic Press, Oxford, 1974).
5. Vidyasagar, M. *A Theory of Learning and Generalization* (Springer-Verlag, New York, 2002).
6. Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
7. Bialek, W., Nemenman, I. & Tishby, N. Predictability, complexity and learning. *Neural Comput.* **13**, 2409–2463 (2001).
8. Robinson, D.A. Integrating with neurons. *Annu. Rev. Neurosci.* **12**, 33–45 (1989).
9. Seung, H.S. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. USA* **93**, 13339–13344 (1996).
10. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
11. Kamil, A.C. & Roitblat, H.L. The ecology of foraging behavior: implications for animal learning and memory. *Annu. Rev. Psychol.* **36**, 141–169 (1985).
12. Gold, J.I. & Shadlen, M.N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
13. James, W. *The Principles of Psychology* Vol. 1 (Holt, New York, 1890).
14. Scoville, W.B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
15. Tetzlaff, C., Kolodziejski, C., Markelic, I. & Wörgötter, F. Time scales of memory, learning, and plasticity. *Biol. Cybern.* **106**, 715–726 (2012).
16. Bailey, C.H., Kandel, E.R. & Harris, K.M. Structural components of synaptic plasticity and memory consolidation. *Cold Spring Harb. Perspect. Biol.* **7**, a021758 (2015).
17. Koch, C., Rapp, M. & Segev, I. A brief history of time (constants). *Cereb. Cortex* **6**, 93–101 (1996).
18. Destexhe, A., Mainen, Z.F. & Sejnowski, T.J. Kinetic models of synaptic transmission. *Methods Neuronal Model.* **2**, 1–25 (1998).
19. Zucker, R.S. & Regehr, W.G. Short-term synaptic plasticity. *Annu. Rev. Physiol.* **64**, 355–405 (2002).
20. Cambridge, S.B. *et al.* Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J. Proteome Res.* **10**, 5275–5284 (2011).
21. Koch, C. *Biophysics of Computation: Information Processing in Single Neurons* (Oxford University Press, New York, 1999).
22. Softky, W.R. & Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334–350 (1993).
23. Shadlen, M.N. & Newsome, W.T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).

24. Hanus, C. & Schuman, E.M. Proteostasis in complex dendrites. *Nat. Rev. Neurosci.* **14**, 638–648 (2013).

25. Shannon, C. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).

26. MacKay, D. *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, 2004).

27. Sreenivasan, S. & Fiete, I. Grid cells generate an analog error-correcting code for singularly precise neural computation. *Nat. Neurosci.* **14**, 1330–1337 (2011).

28. Pereira, J. & Wang, X.-J. A tradeoff between accuracy and flexibility in a working memory circuit endowed with slow feedback mechanisms. *Cereb. Cortex* **25**, 3586–3601 (2014).

29. McCloskey, M. & Cohen, N.J. Catastrophic interference in connectionist networks: the sequential learning problem. *Psychol. Learn. Motiv.* **24**, 109–165 (1989).

30. Fusi, S., Drew, P.J. & Abbott, L.F. Cascade models of synaptically stored memories. *Neuron* **45**, 599–611 (2005).

31. Lahiri, S. & Ganguli, S. A memory frontier for complex synapses. *Adv. Neural Inf. Process. Syst.* **26**, 1034–1042 (2013).

32. Fuster, J.M. & Alexander, G.E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).

33. Goldman-Rakic, P.S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).

34. Ranganath, C. & D'Esposito, M. Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Curr. Opin. Neurobiol.* **15**, 175–182 (2005).

35. Bliss, T.V. & Lomo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol. (Lond.)* **232**, 331–356 (1973).

36. Cajal, S.R.Y. The Croonian Lecture: La fine structure des centres nerveux. *Proc. R. Soc. Lond.* **55**, 444–468 (1894).

37. Konorski, J. *Conditioned Reflexes and Neuron Organization* (Cambridge University Press, Cambridge, 1948).

38. Hebb, D.O. *The Organization of Behavior: a Neuropsychological Theory* (New York: Wiley & Sons, 1949).

39. Takeuchi, T., Duszkiewicz, A.J. & Morris, R.G. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130288 (2014).

40. Kandel, E.R., Dudai, Y. & Mayford, M.R. The molecular and systems biology of memory. *Cell* **157**, 163–186 (2014).

41. Lorente de Nó, R. Vestibulo-ocular reflex arc. *Arch. Neurol. Psychiatry* **30**, 245–291 (1933).

42. Eccles, W.H. & Jordan, F.W. Improvements in ionic relays. British patent number: GB 148582 (1918).

43. von Neumann, J. *First Draft of a Report on the EDVAC* (University of Pennsylvania, Philadelphia, 1945).

44. Jacob, B., Ng, S. & Wang, D. *Memory Systems: Cache, Dram, Disk* (Morgan Kaufmann, 2010).

45. Lisman, J.E. A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proc. Natl. Acad. Sci. USA* **82**, 3055–3057 (1985).

46. Bierman, A. Studies on the effects of structure on the behavior of enzyme systems. *Bull. Math. Biophys.* **16**, 203–257 (1954).

47. Wilson, H.R. & Cowan, J.D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972).

48. Hopfield, J.J. & Tank, D.W. Computing with neural circuits: a model. *Science* **233**, 625–633 (1986).

49. Little, W. The existence of persistent states in the brain. *Math. Biosci.* **19**, 101–120 (1974).

50. Ben-Yishai, R., Bar-Or, R.L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA* **92**, 3844–3848 (1995).

51. Burak, Y. & Fiete, I.R. Accurate path integration in continuous attractor network models of grid cells. *PLoS Comput. Biol.* **5**, e1000291 (2009).

52. Renart, A., Song, P. & Wang, X.-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**, 473–485 (2003).

53. Widloski, J. & Fiete, I.R. A model of grid cell development through spatial exploration and spike time-dependent plasticity. *Neuron* **83**, 481–495 (2014).

54. Goldman, M.S., Levine, J.H., Major, G., Tank, D.W. & Seung, H.S. Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cereb. Cortex* **13**, 1185–1195 (2003).

55. Koulakov, A.A., Raghavachari, S., Kepecs, A. & Lisman, J.E. Model for a robust neural integrator. *Nat. Neurosci.* **5**, 775–782 (2002).

56. Aksay, E., Gamkrelidze, G., Seung, H.S., Baker, R. & Tank, D.W. *In vivo* intracellular recording and perturbation of persistent activity in a neural integrator. *Nat. Neurosci.* **4**, 184–193 (2001).

57. Taube, J.S. & Bassett, J.P. Persistent neural activity in head direction cells. *Cereb. Cortex* **13**, 1162–1172 (2003).

58. Yoon, K. *et al.* Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nat. Neurosci.* **16**, 1077–1084 (2013).

59. Wimmer, K., Nykamp, D.Q., Constantinidis, C. & Compte, A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).

60. Seelig, J.D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).

61. Ferrell, J.E. Jr. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14**, 140–148 (2002).

62. Cannon, S.C., Robinson, D.A. & Shamma, S. A proposed neural network for the integrator of the oculomotor system. *Biol. Cybern.* **49**, 127–136 (1983).

63. Song, P. & Wang, X.J. Angular path integration by moving "hill of activity": a spiking neuron model without recurrent excitation of the head-direction system. *J. Neurosci.* **25**, 1002–1014 (2005).

64. Monod, J. & Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).

65. Crick, F. Memory and molecular turnover. *Nature* **312**, 101 (1984).

66. Lisman, J., Yasuda, R. & Raghavachari, S. Mechanisms of CaMKII action in long-term potentiation. *Nat. Rev. Neurosci.* **13**, 169–182 (2012).

67. Sanhueza, M. *et al.* Role of the CaMKII/NMDA receptor complex in the maintenance of synaptic strength. *J. Neurosci.* **31**, 9170–9178 (2011).

68. Lisman, J. & Raghavachari, S. Biochemical principles underlying the stable maintenance of LTP by the CaMKII/NMDAR complex. *Brain Res.* **1621**, 51–61 (2015).

69. Zhou, Y. *et al.* Interactions between the NR2B receptor and CaMKII modulate synaptic plasticity and spatial learning. *J. Neurosci.* **27**, 13843–13853 (2007).

70. Aslam, N., Kubota, Y., Wells, D. & Shouval, H.Z. Translational switch for long-term maintenance of synaptic plasticity. *Mol. Syst. Biol.* **5**, 284 (2009).

71. Chen, H.X., Otmakhov, N., Strack, S., Colbran, R.J. & Lisman, J.E. Is persistent activity of calcium/calmodulin-dependent kinase required for the maintenance of LTP? *J. Neurophysiol.* **85**, 1368–1376 (2001).

72. Hernandez, A.I. *et al.* Protein kinase M zeta synthesis from a brain mRNA encoding an independent protein kinase C zeta catalytic domain. Implications for the molecular mechanism of memory. *J. Biol. Chem.* **278**, 40305–40316 (2003).

73. Sacktor, T.C. How does PKMζ maintain long-term memory? *Nat. Rev. Neurosci.* **12**, 9–15 (2011).

74. Kwapis, J.L. & Helmstetter, F.J. Does PKM(zeta) maintain memory? *Brain Res. Bull.* **105**, 36–45 (2014).

75. Pastalkova, E. *et al.* Storage of spatial information by the maintenance mechanism of LTP. *Science* **313**, 1141–1144 (2006).

76. Ogasawara, H. & Kawato, M. The protein kinase Mζ network as a bistable switch to store neuronal memory. *BMC Syst. Biol.* **4**, 181 (2010).

77. Jalil, S.J., Sacktor, T.C. & Shouval, H.Z. Atypical PKCs in memory maintenance: the roles of feedback and redundancy. *Learn. Mem.* **22**, 344–353 (2015).

78. Volk, L.J., Bachman, J.L., Johnson, R., Yu, Y. & Huganir, R.L. PKM-ζ is not required for hippocampal synaptic plasticity, learning and memory. *Nature* **493**, 420–423 (2013).

79. Lee, A.M. *et al.* Prkcz null mice show normal learning and memory. *Nature* **493**, 416–419 (2013).

80. Hsieh, K., Tsokas, P. & Sacktor, T. Compensation for PKMζ function in spatial long-term memory in mutant mice. *Soc. Neurosci. Abstr.* 573.01 (2015).

81. Tsokas, P. & Sacktor, T. Compensation for PKMζ function in late-LTP in mutant mice. *Soc. Neurosci. Abstr.* 573.15 (2015).

82. Bourne, J.N. & Harris, K.M. Coordination of size and number of excitatory and inhibitory synapses results in a balanced structural plasticity along mature hippocampal CA1 dendrites during LTP. *Hippocampus* **21**, 354–373 (2011).

83. Meyer, D., Bonhoeffer, T. & Scheuss, V. Balance and stability of synaptic structures during synaptic plasticity. *Neuron* **82**, 430–443 (2014).

84. Bartol, T.M. *et al.* Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife* **4**, e10778 (2015).

85. Lim, S. & Goldman, M.S. Balanced cortical microcircuitry for maintaining information in working memory. *Nat. Neurosci.* **16**, 1306–1314 (2013).

86. Boerlin, M., Machens, C.K. & Denève, S. Predictive coding of dynamical variables in balanced spiking networks. *PLoS Comput. Biol.* **9**, e1003258 (2013).

87. Wiener, N. *Cybernetics or Control and Communication in the Animal and the Machine* (MIT Press, 1961).

88. Sontag, E.D. *Mathematical Control Theory: Deterministic Finite Dimensional Systems* (Springer Science & Business Media, Berlin, 2013).

89. Liddell, E.G.T. & Sherrington, C. Reflexes in response to stretch (myotatic reflexes). *Proc. R. Soc. Lond. Biol.* **96**, 212–242 (1924).

90. Alon, U., Surette, M.G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).

91. Jaeger, H. & Haas, H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80 (2004).

92. Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).

93. White, O.L., Lee, D.D. & Sompolinsky, H. Short-term memory in orthogonal neural networks. *Phys. Rev. Lett.* **92**, 148102 (2004).

94. Eckert, J.P. Jr. A survey of digital computer memory systems. *Proc. IRE* **41**, 1393–1406 (1953).

95. Abeles, M. *et al. Local Cortical Circuits* (Springer, Berlin, 1982).

96. Long, M.A., Jin, D.Z. & Fee, M.S. Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394–399 (2010).

97. Ganguli, S., Huh, D. & Sompolinsky, H. Memory traces in dynamical systems. *Proc. Natl. Acad. Sci. USA* **105**, 18970–18975 (2008).

98. Goldman, M.S. Memory without feedback in a neural network. *Neuron* **61**, 621–634 (2009).

99. van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).

100. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).

101. Murphy, B.K. & Miller, K.D. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* **61**, 635–648 (2009).

102. Sompolinsky, H., Crisanti, A. & Sommers, H.J. Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259–262 (1988).

103. Sprott, J.C. Chaotic dynamics on large networks. *Chaos* **18**, 023135 (2008).

104. Laurent, G. Olfactory network dynamics and the coding of multidimensional signals. *Nat. Rev. Neurosci.* **3**, 884–895 (2002).

105. Hahnloser, R.H.R., Kozhevnikov, A.A. & Fee, M.S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).

106. Shen, L. Neural integration by short term potentiation. *Biol. Cybern.* **61**, 319–325 (1989).

107. Mongillo, G., Barak, O. & Tsodyks, M. Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008).

108. Egorov, A.V., Hamam, B.N., Fransén, E., Hasselmo, M.E. & Alonso, A.A. Graded persistent activity in entorhinal cortex neurons. *Nature* **420**, 173–178 (2002).

109. Aksay, E. *et al.* Functional dissection of circuitry in a neural integrator. *Nat. Neurosci.* **10**, 494–504 (2007).

110. Kalmbach, B.E., Chitwood, R.A., Dembrow, N.C. & Johnston, D. Dendritic generation of mGluR-mediated slow afterdepolarization in layer 5 neurons of prefrontal cortex. *J. Neurosci.* **33**, 13518–13532 (2013).

111. Toyama, B.H. *et al.* Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**, 971–982 (2013).

112. Si, K., Lindquist, S. & Kandel, E.R. A neuronal isoform of the aplysia CPEB has prion-like properties. *Cell* **115**, 879–891 (2003).

113. Stephan, J.S. *et al.* The CPEB3 protein is a functional prion that interacts with the actin cytoskeleton. *Cell Rep.* **11**, 1772–1785 (2015).

114. Fioriti, L. *et al.* The persistence of hippocampal-based memory requires protein synthesis mediated by the prion-like protein CPEB3. *Neuron* **86**, 1433–1448 (2015).

115. McEvoy, M. *et al.* Cytoplasmic polyadenylation element binding protein 1–mediated mRNA translation in Purkinje neurons is required for cerebellar long-term depression and motor coordination. *J. Neurosci.* **27**, 6400–6411 (2007).

116. Miller, P. & Wang, X.-J. Stability of discrete memory states to stochastic fluctuations in neuronal systems. *Chaos* **16**, 026109 (2006).

117. Fiete, I., Schwab, D.J. & Tran, N.M. A binary Hopfield network with 1/log(n) information rate and applications to grid cell decoding. Preprint at http://arxiv.org/abs/1407.6029 (2014).

118. Strogatz, S.H. *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering* (Westview Press, 2014).

119. Compte, A., Brunel, N., Goldman-Rakic, P.S. & Wang, X.-J. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).

120. Burak, Y. & Fiete, I.R. Fundamental limits on persistent activity in networks of noisy neurons. *Proc. Natl. Acad. Sci. USA* **109**, 17645–17650 (2012).

121. Lim, S. & Goldman, M.S. Noise tolerance of attractor and feedforward memory models. *Neural Comput.* **24**, 332–390 (2012).

122. Kilpatrick, Z.P., Ermentrout, B. & Doiron, B. Optimizing working memory with heterogeneity of recurrent cortical excitation. *J. Neurosci.* **33**, 18999–19011 (2013).

123. Monteforte, M. & Wolf, F. Dynamic flux tubes form reservoirs of stability in neuronal circuits. *Phys. Rev. X* **2**, 041007 (2012).

124. Landauer, T.K. How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cogn. Sci.* **10**, 477–493 (1986).

125. Shepard, R.N. Recognition memory for words, sentences, and pictures. *J. Verbal Learn. Verbal Behav.* **6**, 156–163 (1967).

126. Brady, T.F., Konkle, T., Alvarez, G.A. & Oliva, A. Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. USA* **105**, 14325–14329 (2008).

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

128. Abu-Mostafa, Y.S. & St Jacques, J. Information capacity of the Hopfield model. *IEEE Trans. Inf. Theory* **31**, 461–464 (1985).

129. Gardner, E. & Derrida, B. Optimal storage properties of neural network models. *J. Phys. A* **21**, 271 (1988).

130. Amit, D.J., Gutfreund, H. & Sompolinsky, H. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.* **55**, 1530–1533 (1985).

131. Parisi, G. A memory which forgets. *J. Phys. A* **19**, L617 (1986).

132. Benna, M.K. & Fusi, S. Computational principles of biological memory. Preprint at <http://arxiv.org/abs/1507.07580> (2015).

133. Hillar, C., Tran, N. & Koepsell, K. Robust exponential binary pattern storage in Little-Hopfield networks. Preprint at <http://arxiv.org/abs/1206.2081> (2012).

134. Valiant, L.G. The hippocampus as a stable memory allocator for cortex. *Neural Comput.* **24**, 2873–2899 (2012).

135. Miller, G.A. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97 (1956).

136. Conway, A.R., Kane, M.J. & Engle, R.W. Working memory capacity and its relation to general intelligence. *Trends Cogn. Sci.* **7**, 547–552 (2003).

137. Barrouillet, P., De Paepe, A. & Langerock, N. Time causes forgetting from working memory. *Psychon. Bull. Rev.* **19**, 87–92 (2012).

138. Lewandowsky, S., Oberauer, K. & Brown, G.D.A. No temporal decay in verbal short-term memory. *Trends Cogn. Sci.* **13**, 120–126 (2009).

139. Hughes, J.R. Absence seizures: a review of recent reports with new concepts. *Epilepsy Behav.* **15**, 404–412 (2009).

140. Mante, V., Sussillo, D., Shenoy, K.V. & Newsome, W.T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).

141. Druckmann, S. & Chklovskii, D.B. Neuronal circuits underlying persistent representations despite time varying activity. *Curr. Biol.* **22**, 2095–2103 (2012).

142. Emiliani, V., Cohen, A.E., Deisseroth, K. & Häusser, M. All-optical interrogation of neural circuits. *J. Neurosci.* **35**, 13917–13926 (2015).

143. Sterzer, P., Kleinschmidt, A. & Rees, G. The neural bases of multistable perception. *Trends Cogn. Sci.* **13**, 310–318 (2009).

144. Colgin, L.L. *et al.* Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *J. Neurophysiol.* **104**, 35–50 (2010).

145. Loewenstein, Y., Yanover, U. & Rumpel, S. Predicting the dynamics of network connectivity in the neocortex. *J. Neurosci.* **35**, 12535–12544 (2015).

146. Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* **10**, 647–658 (2009).

147. Dudai, Y. The restless engram: consolidations never end. *Annu. Rev. Neurosci.* **35**, 227–247 (2012).

148. Nabavi, S. *et al.* Engineering a memory with LTD and LTP. *Nature* **511**, 348–352 (2014).

149. Lisman, J. The challenge of understanding the brain: where we stand in 2015. *Neuron* **86**, 864–882 (2015).

150. Jonides, J. *et al.* The mind and brain of short-term memory. *Annu. Rev. Psychol.* **59**, 193–224 (2008).