NEUROBIOLOGY OF LEARNING AND MEMORY **70,** 252–267 (1998)
ARTICLE NO. NL983851

The Neurophysiology of Reminiscence

Bruce L. McNaughton

Department of Psychology, University of Arizona, Tucson, Arizona 85749

When this stimulation of the hippocampal gyrus was carried out, the hippocampal formation and amygdaloid nucleus were still intact. But the rest of the anterior half of the temporal lobe had been removed. The fact that stimulation could still produce a flash-back of former experience would support the suggestion that comes from other evidence (Milner and Penfield, 1958) that the hippocampus of the two sides is, in fact, the repository of the ganglionic patterns that preserve the record of the stream of consciousness. If not the repository, then each hippocampus plays an important role in the mechanism of reactivation of that record (Penfield and Roberts, 1959, *Speech and brain mechanisms.* Princeton, NJ: Princeton University Press).

INTRODUCTION

Reminiscence can be defined subjectively as the process of "calling into mind" a previous event or episode. This seems to occur mainly during periods when the brain is not otherwise fully preoccupied with the processing of external inputs. Some basic questions in the neurosciences concern the objective description of what goes on in the brain during reminiscence, what mechanisms support this form of retrieval, and what is its biological function. A reasonable, although unproven, starting assumption is that during reminiscence, in some part or parts of the brain, patterns of neuronal activity resembling those which occurred during the corresponding experience are reactivated (e.g., Farah, 1995). Some form of reminiscence, or memory trace reactivation during "off-line" periods such as sleep, has been thought to play an important role in the process of memory consolidation, through which episodic information is incorporated into the brain's knowledge base (Marr, 1971; McClelland & McNaughton, 1995; Buzsaki, 1989). Among the earliest neurophysiological evidence of such off-line reactivation of previously experienced patterns of neural activity was the observation by Pavlides and Winson (1989) that hippocampal place cells which had been robustly active during a period of waking behavior were selectively more active during a subsequent episode of sleep. Because only single neurons were recorded in these studies, however, it was not possible to conclude that the actual patterns of activity were necessarily being recapitulated, and hence one could not be sure that mnemonic representations of the experience were being reinstated. In 1994, Wilson and McNaughton reported that when groups of hippocampal CA1 pyramidal neurons were simultaneously recorded from, those cells which tended to fire together during behavior, as a consequence of the overlap of their place fields, had an enhanced tendency to

This work was supported by MH46823. The author is grateful to Carol Barnes, Hemant Kudrimoti, Bartlett Mel, Bill Skaggs, Alexei Samsonovich, and Francesco Battaglia for helpful discussions and/or comments on the manuscript.

Address correspondence to B.L. McNaughton, University of Arizona, Room 384, Life Sciences North Building, Tucson, AZ 85724. Fax: (520) 626-2618. E-mail: bruce@nsma.arizona.edu.

1074-7427/98 $25.00

fire together during subsequent slow wave sleep (SWS). By studying the temporal asymmetries of pairwise cross-correlation functions of simultaneously recorded cells, Skaggs and McNaughton (1996) found that some information about the sequences of activation within the recorded hippocampal neural ensemble during behavior was also preserved during sleep. More recently, Qin et al. (1997) have reported that the reinstatement, in sleep, of patterns of neuronal discharge correlation during behavior also occurs in both neocortical–neocortical cell pairs and hippocampal–neocortical pairs. Thus, it appears that some form of coherent reactivation of memory traces and trace sequences occurs over much of the brain during SWS. These results as a whole provide neurophysiological evidence for both of Hebb's (1949) central constructs, the "cell assembly" and the "phase sequence." Thus far, multiple single-neuron recording studies have used as the dependent variable correlations over time in the activities of pairs of cells (Abeles & Goldstein, 1977; Aertsen et al., 1989; Gerstein et al., 1985; Palm et al., 1988). This approach has been greatly facilitated by the ability to record from many (up to 150) cells at once, because the number of pairs available for correlation analysis increases in proportion to the square of the number of cells recorded [$N(N\text{-}1)/2$]; however, the pairwise activity correlation approach only indirectly assesses the variable of interest, which is the patterns themselves, and as discussed below has several severe limitations. Thus, more refined approaches need to be explored.

Beginning with Hebb (1949) and leading up to modern theories of "attractor" dynamics in neural networks (Amit, 1989; Hopfield, 1982), much has been written about the possible synaptic mechanisms and network architectures that could lead to memory reactivation. Less has been said, however, about how we might detect and quantify this process experimentally or of the possible complexities inherent in this effort. This article explores, in the light of the meager body of available data, some of the conceptual issues involved in the study of memory reactivation at the neurophysiological level, in particular the use of spike train correlation techniques and the possible inferences that can be drawn from them, and the population or state vector approach. The goal of this exercise is not to achieve any mathematically rigorous conclusions, but to outline the scope of the problem and to point out some possible solutions.

WHAT IS MEMORY REACTIVATION?

What is implied by the concept of memory reactivation? It is generally assumed that the internal representation of an instantaneous experience is based on a unique distribution (mathematically, a vector) of neural spiking activity within the CNS and that the memory of this experience results from the reinstatement of some facsimile of the original pattern, in the absence of the corresponding input. Similarly, the experience of and memory for a sequence of events are based on the establishment and reactivation of a unique sequence of activity vectors, what Hebb (1949) referred to as a phase sequence. A vector of spike rates estimated over some small interval is not the only possible definition of a perceptual event or even necessarily the best definition. Indeed, there are examples of differences in auditory perception resulting from temporal differences in inputs that are shorter than the duration of a single spike. It is also possible that the smallest unit of mnemonic representation is not an instantaneous pattern of spiking activity, but rather a short phase sequence.

Nevertheless, spike rate vectors are a useful way to start, and to apply them, it is necessary select some "appropriate" time window (Δt) over which to integrate the spike activity. In a noisy system, there is always a tradeoff between temporal resolution and accuracy in the estimate of firing rate, particularly in systems like the hippocampus, in which information is represented by sparse activity patterns (Marr, 1971; McNaughton and Morris, 1987). For example when reconstructing a rat's spatial location from the place cell population code, there tends to be a U-shaped function for error versus Δt with a minimum around 0.5 s (Wilson & McNaughton, 1994; Zhang et al., 1997). One definition of an appropriate time scale is one which minimizes the mean difference between adjacent vectors; however, even this might be problematic. For example, phase sequences in the hippocampus are oscillatory at the time scale of the theta rhythm (O'Keefe & Recce, 1993; Skaggs & McNaughton, 1995). In other words, within each theta cycle, the internal representation of position begins at the rat's current location and then moves, through a short sequence of locations, to a point some distance ahead (typically about 10–20 cm) of the rat. At the end of each theta cycle, the state vector jumps back to some point along the trajectory during the previous theta cycle corresponding to the new current location (Samsonovich & McNaughton, 1997; Tsodyks et al., 1996). Thus, the states change relatively quickly at time scales finer than the theta period and more slowly at longer time scales; moreover, integration even over a single theta cycle seriously smears the location vectors.

Temporal integration time is not the only issue. It should be clear that even with the best of current methods, only a very small proportion of the total population of neurons is sampled. In general, the accuracy in the estimate of the similarity of two patterns will be proportional to the square root of the number of recorded neurons. An x-fold increase in accuracy requires an x^2-fold increase in sample size. This was verified approximately by Wilson and McNaughton (1993) for position reconstruction from place cell populations. The upshot is that the appropriate integration time will depend on the question at hand, on parameters of the system, in particular, the rate at which cells fire and the rate at which the states of activity in the network change with time, and on the numbers of cells that can be simultaneously recorded.

ORGANIZATION OF DATA

Suppose one records simultaneously from N neurons, over T time intervals. One then has an N by T data matrix (**Q** in Fig. 1) in which the row vectors represent the time series of rates for each neuron and the column vectors represent the distribution of "instantaneous" firing rates in the ensemble. The rows can thus be called temporal vectors and the columns can be called state vectors. The state vector defines a location in the N-dimensional space of possible firing rate distributions among the N-neurons and thus corresponds to the notion of an activity pattern. The simple correlation between any two temporal vectors represents the correlation of firing between two cells over the time interval in question and is a function of the trajectory of the system during that interval, whereas the correlation between any two state vectors reflects the similarity of the global state of the system at the corresponding times. Let us call the interrow correlations *temporal correlations* and the intercolumn correlations *state correlations.* It should be clear that, in considering the pro-

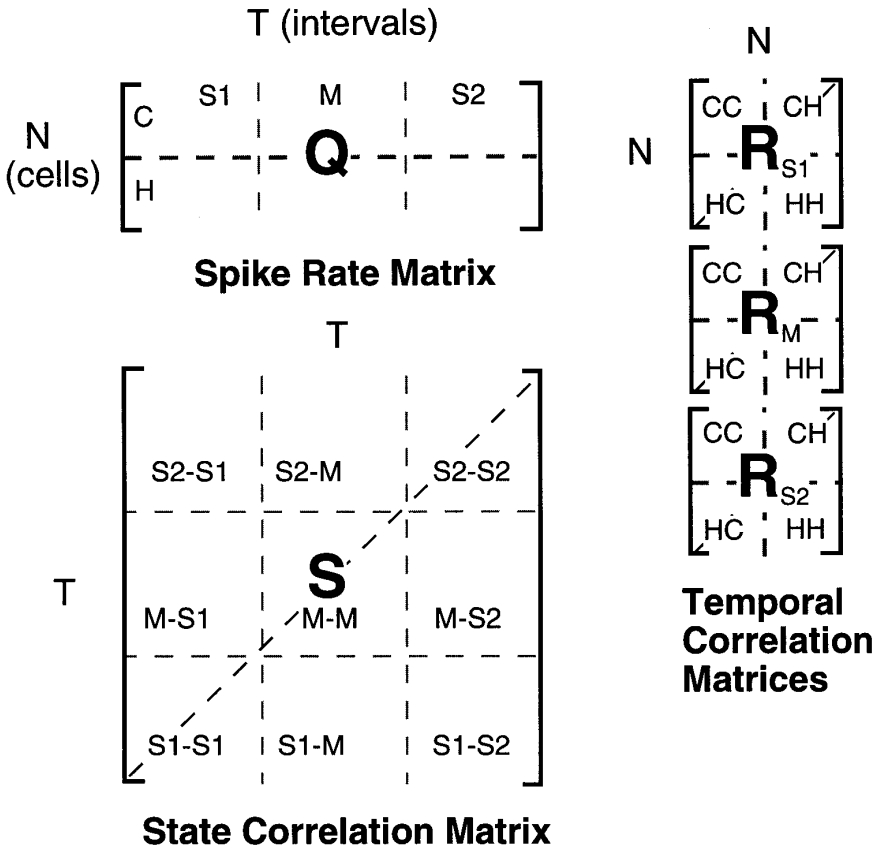


FIG. 1. Data structures for ensemble recording study of mnemonic processes. The raw spike rate data consist of an $N \times T$ matrix [**Q**] containing N rows of spike trains (temporal vectors), corresponding to the N recorded neurons, binned in intervals of Δt, and T columns of state vectors (a.k.a., "population" vectors). The matrix **R** is the matrix of all pairwise correlations among spike rates. The matrix **S** contains all pairwise correlations among state vectors. **R** and **S** are diagonally symmetric. If the rows of **Q** are subdivided into sets (e.g., C, H) because cells of different types or from different anatomical regions (e.g., cortex and hippocampus) are included in the sample, then **R** is subdivided into different submatrices of within- and between-class temporal correlations. Similarly if the columns of **Q** are subdivided into different epochs (e.g., a period of sleep, S1, a period of behavior on a maze, M, and another period of sleep S2), then **S** is subdivided into different submatrices of within- and between-epoch state correlations. The off-diagonal submatrices of **S** are of particular interest in the question of memory reactivation. Also, one may generate different **R** matrices for each temporal epoch in **Q** (e.g., $\mathbf{R}_{S1}$, $\mathbf{R}_{M}$, $\mathbf{R}_{S2}$), and different **S** matrices for each cell class in **Q** (e.g., $\mathbf{S}_{l}$, $\mathbf{S}_{m}$, $\mathbf{S}_{n}$). The similarities of the different R matrices have been used to make inferences about memory reactivation. Finally, one may consider time-lagged temporal cross-correlations (a.k.a., the cross-correllogram) and the mean time-lagged state vector correlation. The former generates an additional R matrix for each Δt [e.g., $R_i(\Delta t)$] and the latter generates a vector containing the average correlation between state vectors as a function of the interval between them.

cess of memory reactivation, we are really interested in state correlations and not temporal correlations; yet, as we shall see, it is sometimes possible to draw useful inferences from the latter about the former. Two other useful constructs can be defined, the $N \times N$ matrix of all temporal correlations, which we shall call the *temporal correlation matrix* over the interval 1-T (**R** in Fig. 1), and

the $T \times T$ matrix of all state correlations, or *state correlation matrix* for the set of N cells over the same interval (**S** in Fig. 1). Both of these matrices are diagonally symmetric, with values of 1.0 by definition along the diagonal. The size of the state correlation matrix is proportional to $1/\Delta t^2$. For a typical recording session it can become very large if Δt is small.

In general, because the assignment of identification numbers to cells in a sample is typically arbitrary, we do not expect to see any particular structure in the temporal correlation matrix (unless we encounter effects of anatomical topography); however, at least under some conditions, we do expect to see structure in the state correlation matrix. For example, suppose that we are considering hippocampal place cells in a rat performing a rather slow walk along a linear track, without changing direction. We expect the values of the of the state correlation matrix elements to be uniformly rather high near the diagonal and typically to decline with distance from the diagonal. If the rat runs more quickly, the correlations between state vectors will also decline more quickly as a function of the interval between them (Fig. 2). It is also possible that intrinsic dynamics may lead to abrupt changes in the state vector (for example, at the transition between two orthogonal attractor states). In general, we can define an instantaneous state vector velocity as the distance (or angle) between two state vectors divided by the time. We can also define the average velocity as the average of the instantaneous values. If the velocity is constant, or at least stationary, and the trajectory is random, the mean cosine of the angle between two vectors will tend toward zero exponentially as a function of interval, and hence one can define a state vector time-constant (McNaughton & Skaggs, 1997). Interestingly, the time constant determined in this manner is substantially smaller during SWS than during the awake theta state (AWΘ) or REM sleep. This presents some complexities in the analysis that will be discussed below. Of course, if the rat's trajectory in space is periodic, then the hippocampal state correlation matrix for this epoch will have periodic stripes of high values, parallel to the diagonal, at periods corresponding to trajectory cycles (Fig. 3); similarly, if motor behavior or sensory inputs are periodic, there may be periodicities in the state correlation matrix of areas which exhibit neural activity that is correlated with these variables. More generally, if the rat's brain occupies any region of its state space more than once, then some off-diagonal elements will have unexpectedly large values. If large state correlations occur between instants of waking behavior and instants of subsequent sleep, and if these states do not occur before the behavior, one may reasonably conclude that memories are being reactivated (Fig. 4).

A theoretical (or perhaps semantic) caveat to the foregoing statement is that some states may result from patterns of synaptic connections that are not learned during the behavioral episode in question, but are either programmed genetically or learned early in development. For example, there is a system of neurons that signal relative head direction (Taube et al., 1990). This system has a limited set of observable state vectors and a limited set of possible transitions among states. This set defines a preconfigured, one-dimensional closed manifold (a ring) in the space of theoretically possible states (Skaggs et al., 1994). Similarly, the relative positions of place fields of subicular cells appear to be independent of which environment the animal is in and hence appear to define a two-dimensional manifold that may be predetermined in the synaptic matrix (Sharp, 1997); Samsonovich and McNaughton (1997) have proposed that the synaptic matrix of CA3

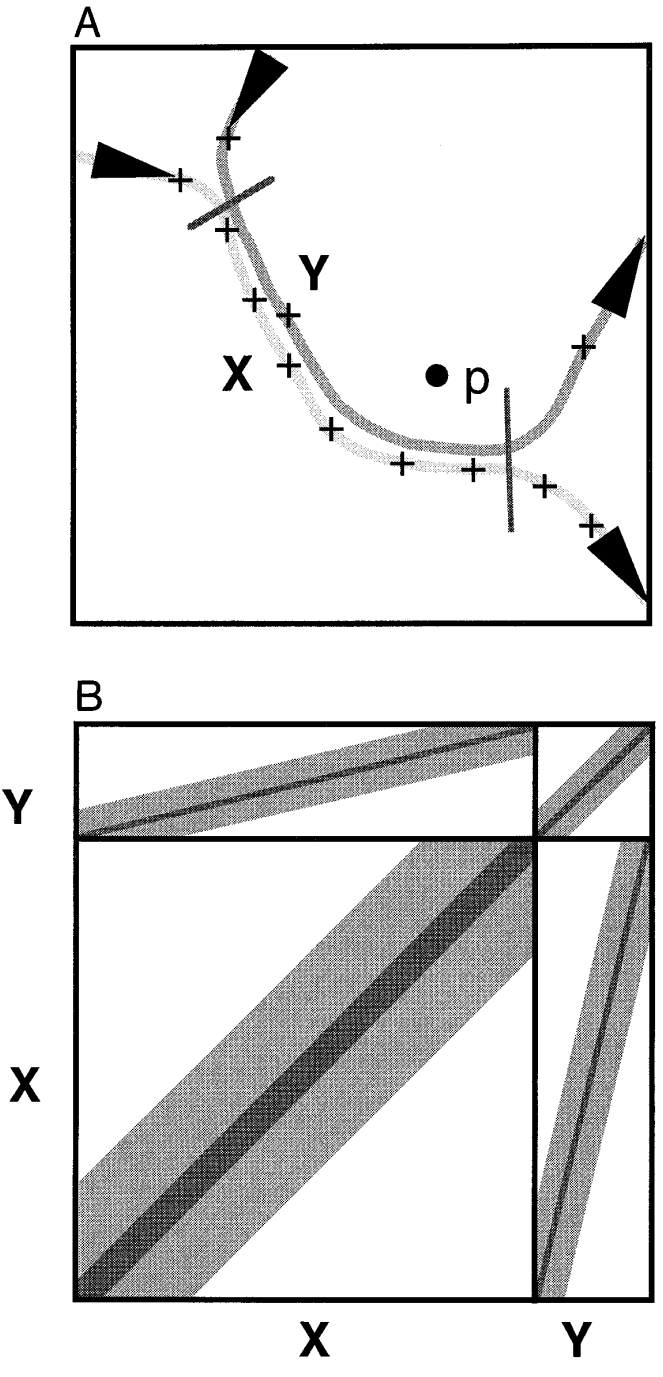


FIG. 2. (A) Two hypothetical trajectories (**X, Y**) through the same path in network state space. The plus signs indicate successive time intervals. Even though the sequences of states are highly similar, the speed of trajectory **Y** is about four times faster than that of **X.** The point *p* represents the sort of error that would arise in estimating trajectory **Y** if the integration time Δt was comparable to the plotted interval markers. (B) Hypothetical state vector correlation matrix for the common region of the two trajectories **X** and **Y** (demarcated by lines in A). The magnitude scale of the state vector correlations is represented by grayscale shading. When the vector speed is slow, the diagonal band of high correlation is broad. When the speed is high, the diagonal band is narrow. The change in speed between **X** and **Y** is represented by the slope of the diagonals of the off-diagonal submatrices.

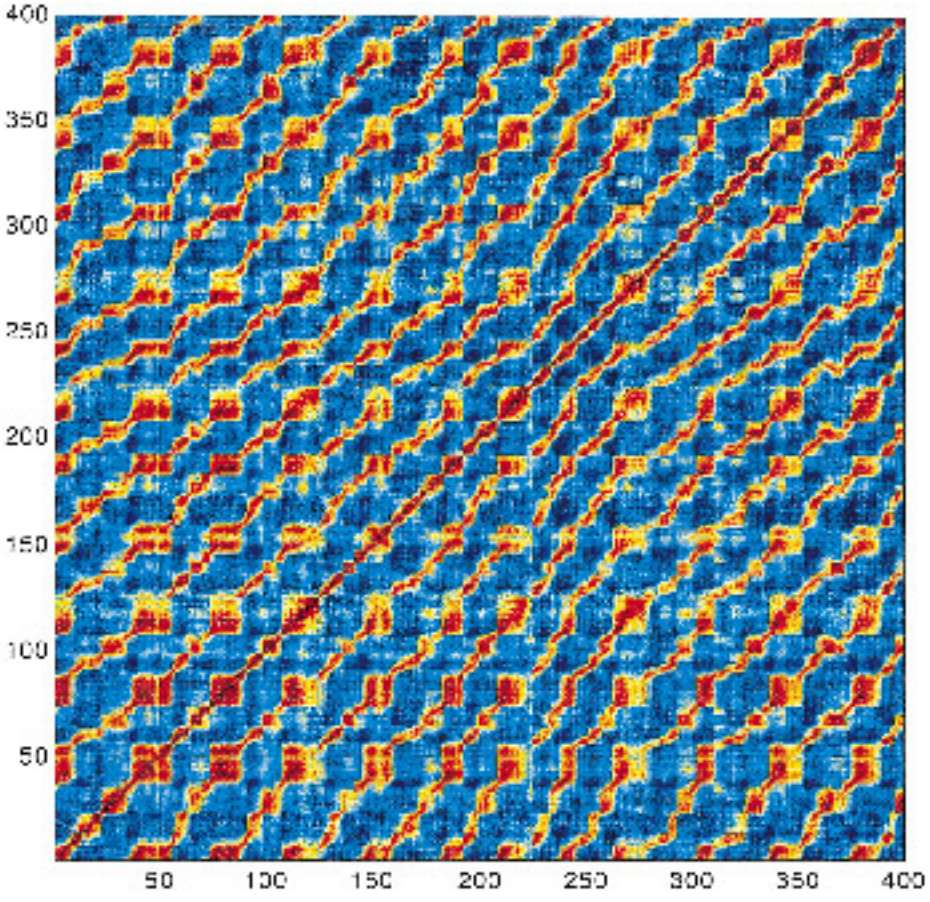
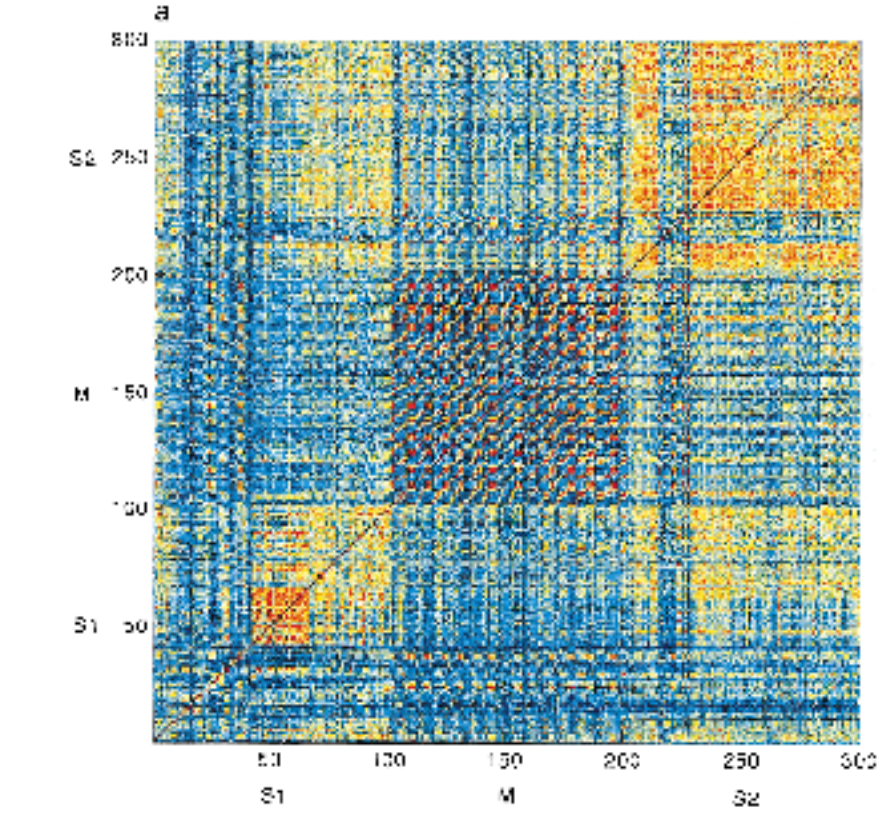


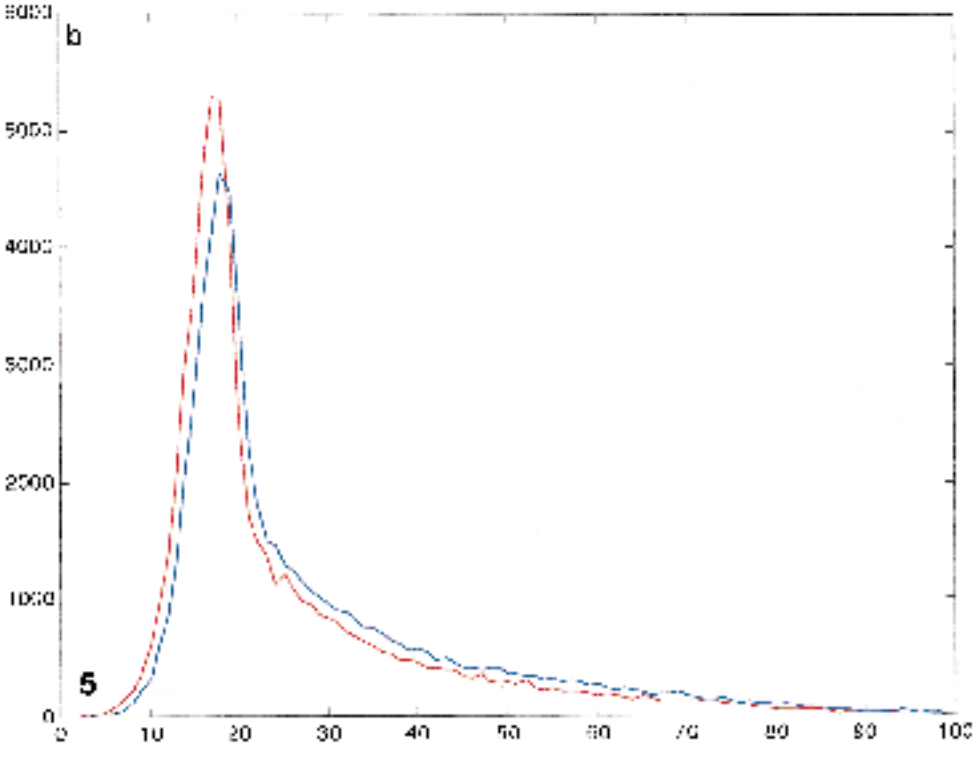
FIG. 3. State vector correlation matrices from a set of hippocampal neurons recorded simultaneously while the rat ran successive laps around a closed rectangular track over a period of 400 s. Note the diagonal stripes of high correlation which reflect the similarities of the state vectors at each location on the track.

may be preconfigured to define a large set of such 2-D manifolds, although this remains to be confirmed. Similarly, it is to be expected that the state space occupancy of motor cortex may be constrained to relatively simple manifolds, such as the sphere that defines the population codes for reaching (Georgopoulos et al., 1988), and it is well established that the state space occupancy of the superior colliculus is inherently two-dimensional (Sparks & Mays 1990). It has been shown that, given the appropriate synaptic matrix, such predefined states can, in princi-

---

FIG. 5. (A) **S** matrix for three equal segments (500 s each) of data from the experiment from which Fig. 3 was constructed, in which the rat first slept (S1), ran on a rectangular track (M), and then slept again. (The *T* axes are in units of 5 s, and only every fifth bin is plotted due to memory limitations). The central diagonal submatrix is essentially the same data as shown enlarged in Fig. 3. Note that there appear to be more high values in the M–S2 submatrix than in the S1–M submatrix. This is confirmed in the histograms in (B) which reflect the distributions of state correlation values for the S1–M (red) and M–S2 (blue) submatrices. There were significantly ($p < .0001$, χ^2) more higher values in the latter, indicating that the states in S2 were more similar to those of M than were the states of S1.


a
S2
M
S1
S1
M
S2


b
5
0
10
20
30
40
50
60
70
80
90
100

ple, be selectively reactivated without recourse to the associative synaptic modification normally thought to underlie memory (Shen & McNaughton, 1994). In some circumstances (i.e., sparse coding), a simple selective bias on the firing probabilities of the cells involved in one of the preconfigured states is sufficient to obtain activation of that state. Thus, the increased probability of occurrence of a state vector following its appearance during a behavioral episode may reflect either the development of new cell assemblies through associative synaptic modification or the selective reactivation of cell assemblies that were already defined in the synaptic matrix before the behavioral episode in question. In either case there is some form of memory at play, but the expression of this memory per se does not necessarily indicate any particular encoding mechanism.

TEMPORAL CORRELATIONS OF SPIKE TRAINS

Let us now consider some properties of the temporal correlation matrix. Suppose as described above, we allow the rat to take a long walk without changing directions (i.e., the states never repeat) and then allow him to sleep. If the rat's dream perfectly reconstructs his trajectory, then obviously the temporal correlation matrices for these epochs will be identical (assuming error free recording); however, if the hippocampus represents each location uniquely (and if all states with the normal sparsity level are equally probable), then, as the length of the trajectory increases, the mean and variance of the correlation distributions will tend to zero. Given some noise in the recording, this result suggests that it will become increasingly difficult to detect reminiscence using temporal correlations as the length of the sequence (i.e., number of states occupied) increases (note that this problem is alleviated if the same set of states is visited repeatedly). It also reveals that the absolute magnitudes of the correlations are theoretically irrelevant. It is the relative magnitudes within the *R* matrix that are of interest.

Now suppose that during the reminiscence period, the temporal order of events is completely scrambled. Because the temporal order of data pairs does not enter into the formula for the correlation coefficient, scrambling the events has no effect at all. The temporal correlation matrix depends only on the state–

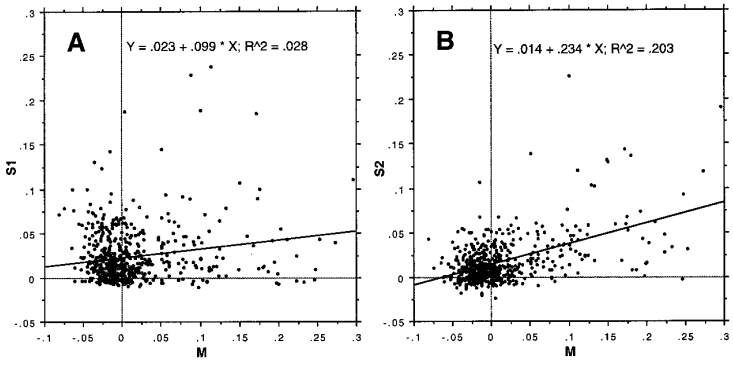


FIG. 4. A typical example of the comparison of the distributions of *R* in sleep before (S1) and after (S2) a period of behavior on a track maze (*M*). Each point represents the zero-lag correlation of spike rates (100-ms bins) for two cells. The effect of *M* accounts for only about 3% of the variance of correlations in S1, but about 20% of the variance in S2.

space occupancy distribution, i.e., the relative amount of time spent by the system in each region of its state space. Thus, the similarity of two temporal correlation matrices is independent of event order. A corollary is that the temporal correlation matrix is independent of the state vector velocity profiles, so long as the occupancy distributions are the same and so long as the smearing effect of the integration time is not large (i.e., an "appropriate" integration time is used). These two properties almost certainly account for why it is possible at all to detect memory reactivation using temporal correlation methods. In contrast, if the state–space occupancy distribution is changed, for example, if some state vectors are reinstated more or less frequently than they occurred during the behavior, then there will be a corresponding change in the temporal correlation matrix. One problem this raises is that the temporal correlation method cannot distinguish between the case in which the reminiscence consists of noiseless recall of a mixture of states, some of which occurred during the target behavioral session and some of which occurred during some other period for which there were no data recorded, and the case in which only target vectors are recalled, but the recall itself is noisy. Thus, for a given experience, temporal correlations cannot distinguish between a noisy memory trace that is reinstated frequently and a robust trace that is reinstated infrequently. In addition, it should be clear that, even if recall for each individual vector was noiseless, if the data segments were of unequal lengths, then the temporal correlation matrices would be dissimilar because the state occupancy distributions would be dissimilar. The latter problem might be overcome, however, if the memory replay occurred at high speed, as appears to be the case in both hippocampus and neocortex during SWS (Qin et al., 1997; Skaggs & McNaughton, 1997) but not during REM sleep (Skaggs & McNaughton, 1997). This might permit the state–space occupancy distributions to be similar for two epochs of different durations.

Another intrinsic difficulty in the use of temporal correlations to compare the state-space occupancy distributions in two epochs is the existence of differences in extrinsic or intrinsic modulatory influences on firing probabilities. Modulation can be defined as activation of inputs that affect all of the principal cells in essentially the same way at any given instant. Such effects lead to nonstationarity or periodicity of firing rates in the population (i.e., in the length of the state vectors) and induce temporal correlations which may have nothing to do with the memory recall per se. In general terms, modulation may change the state–space occupancy distribution by changing the lengths of vectors, the angles of vectors, or both, depending on the nature of the modulator. Whatever the effect, modulation will tend to increase the values of the temporal correlations of all cell pairs and distort the temporal correlation matrix. Nevertheless, in many circumstances, it may be possible to compare two data sets which have been subject to different modulatory effects. What one is comparing is the similarity of the temporal correlation matrices, which can be measured by the coefficient of variation (r^2) of their elements. Ideally, if the form of the modulation function is known exactly, and if no information has been lost, then the inverse function could be applied to the data.

A typical experimental protocol used in the author's laboratory involves a period of sleep (S1) followed by a period of waking behavior, typically maze running of some sort (M), followed by a second period of sleep. The question addressed is how much of the variance in the temporal correlation matrix for

S2 can be accounted for by the variance in M, after removing any effects of S1. In other words, is the temporal correlation matrix for S2 more similar to that of M than the matrix for S1 which preceded the experience (Fig. 4)? If the partial correlation is significant, then it is highly probable that the state-space occupancy distribution of S2 is more similar to that of M than is the distribution for S1 and hence that there has been storage and reactivation of mnemonic traces. A similar approach would be to ask, "Given the correlation distribution of S1, how much more information about the correlation distribution of M is added by knowing the distribution in S2?"

CORRELATION OF STATE VECTORS

Let us now consider what might be accomplished with the same data using state vectors. Suppose we construct the state vector correlation matrix for the entire S1, M, S2 sequence just described. This partition of the time series leads to six unique submatrices (Fig. 1) of which we are principally interested in the S1–M submatrix and the M–S2 submatrix (Fig. 5A). Without considering temporal order, we can ask the simple question: Is the average of the elements of the latter significantly larger than that for the former? If so, then the state-space occupancies are more similar and memory is being expressed during S2. The problem is in detecting the difference statistically. In general, the broader the distribution in state-space of the vectors in M, the harder will be the detection problem. If the K state vectors in M are all substantially different from one another, then we expect at best K "hits" and $K(K\text{-}1)$ "misses." If K is large, the means will not differ significantly even if there is perfect reminiscence. The statistics will likely need to be based on the shapes of the two distributions (Fig. 5B). The odds of detection improve considerably if only a relatively few states are sampled repeatedly. If the states are well correlated with externally observable events which occur repeatedly, such as the animal's location or motor patterns, then one may construct average state vectors for the events and use these as templates to match with the off-line vectors. This would reduce the number of target vectors and thus simplify the analysis.

REPLAY OF SEQUENCES

As suggested above, memory recall is likely to involve not just instantaneous vectors, but trajectories of the system through its state space (i.e., temporally extended events). How might the retraversal of a route through state space be detected and quantified? In general, we are interested in the question of state vector sequence homology in two epochs, T1 and T2.

One approach is to make use of the correlations between temporally shifted ("lagged") spike trains. Shifting one spike train relative to the other by $\pm\Delta t$ gives a new correlation, and the sequence of these correlations for successive lags of $\pm\Delta t$ is called the cross-correlation function. For each lag there is another $N \times N$ temporal correlation matrix. In other words, there is an $N \times N \times K$ array of correlations, where K is the number of positive, negative, and zero lags. The array has a symmetry about the zero lag because the correlation between cells i and j at lag n is equal to that of cells j and i at lag $-n$. Thus, one need consider only those elements in which $i \geq j$. For each i,j, the cross-correlation function is dependent on the exact state trajectory. If the trajectory

itself is quasi-periodic, then there will be peaks in the cross-correlation occurring at multiples of the basic period and most cross-correlograms will be asymmetric about the origin. Illustrative cross-corellograms for two parietal cortical cells during sleep and waking are shown in Fig. 4. In general, unless the trajectory is perfectly periodic, the peaks will decline as a function of lag. Skaggs and McNaughton (1996) defined "bias" as a simple measure of asymmetry in two cross-correlations. Bias is defined as the sum of all bins from zero to $-t$, minus the sum of all bins from zero to $-t$. For N cells, this leads to an $N \times N$ bias matrix from the cross-correlation array. The similarities of the bias matrices for two epochs can be compared, and Skaggs and McNaughton (1995) found that there was a significant correspondence in the bias matrix for a period of SWS with that of the preceding behavior. In contrast, the bias matrix for the preceding sleep was not related to that of the behavioral epoch. Thus, they concluded that there must be significant replay of sequences during SWS. The bias analysis, although relatively crude, involves few *a priori* assumptions and is reasonably robust. It depends, however, on selecting an appropriate time window for analysis. Another approach investigated by Skaggs and McNaughton (1997) is the analysis of the lag latency of the first peak in the (smoothed) cross-correlation. It was found that peak latencies in behavior and SWS were significantly correlated, but the distribution during SWS was highly compressed (Fig. 6) and slope of the regression line for peak latencies between the two states was about 40, suggesting a large increase in state vector velocities in SWS compared to behavior.

It is possible that a more comprehensive approach using the cross-correlations could involve the comparison of the rate correlation matrices at different lags. Suppose we have two sequences M and S2 as in Fig. 1. If we compute the similarity of the temporal correlation matrices of M and S2 at each temporal lag in M with every lag in S2, we obtain what could be called the cross-correlation similarity matrix. If the two sequences are similar, there will be high values along the diagonal. If the two sequences are similar but the time-scale has been compressed in one of them, then there will still be a stripe of high values, but the slope of the stripe will differ from 45°. It should be possible to estimate the degree of compression from the change in this slope.

What alternative strategies can be devised from the state vectors? In considering the replay of sequences, one is comparing two different subsets of **Q,** for example, S2 and M in Fig. 1. The question is whether segments of S2 correspond to segments of M. One approach would be to use an analogy to the methods used in molecular biology to study DNA sequence homology. In this case we can consider every subsequence of S2 of length L and compute its fit to the every size L subsequence in M. The fit measure in this case is simply the mean correlation of the individual vectors in the two sequences. The maximum fit is then taken as the measure of sequence homology, and the average over all subsequences gives a measure of the overall homology at that characteristic length.

A possibly more powerful, but related, approach would be to take each L by L diagonal submatrix of **S** and convolve it with the corresponding L by T submatrix of **S** (i.e., slide the $L \times L$ submatrix laterally along **S,** taking the dot product at each point). This will result in $(T - L)^2$ values whose distribution can be compared to the expected distribution based on random shuffles of the columns of Q.

Such procedures might be successful for examining sequence replay during REM sleep, for which there is evidence that the assumption of equal state

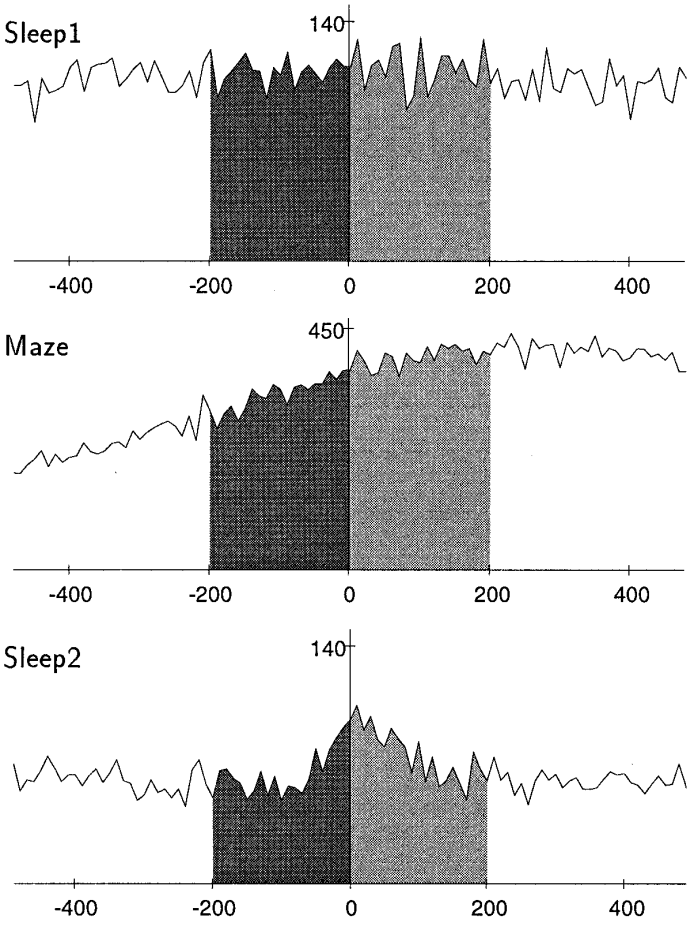


FIG. 6. Representative cross-correllograms for a pair of parietal cortical neurons in S1, M, and S2. Note that the cross-correllogram is flat in S1, has a broad asymmetric peak in M, and has a narrow peak with proportional asymmetry in S2. (Reproduced from Y.-L. Qin, B. L. McNaughton, W. Skaggs, & C. E. Barnes, 1998, Memory reprocessing in corticortical and hippocampocortical neuronal ensembles, *Philosophical Transactions of the Royal Society of London B,* **352,** 1525–1533, Fig. 2. Reprinted with the kind permission of the Royal Society.)

vector velocities is approximately valid. It may or may not work for comparisons between SWS and waking behavior where, as just discussed, the playback velocities may be as much as 40-fold faster. In this case a more refined approach would be necessary.

The first problem in comparing the homology of state-space trajectories in two epochs when the velocities are not equal has to do with the choice of integration times. Take the case of comparing SWS to waking behavior as an illustration. It is known that most of the mnemonic reactivation that occurs during SWS occurs during sharp waves, whose duration is on the order of 100 ms. Suppose one uses 100 ms as the integration time. If the velocity of replay is actually 40 times faster in SWS, this means that the SWS vectors represent the average of 40 AWΘ vectors (i.e., 4 s). Available data suggest that during typical maze running behavior, two hippocampal vectors separated by 4 s are virtually orthogonal. Thus, even if an exact trajectory was replayed in SWS, the apparent trajectory would be seriously distorted.

Assuming that Δt has been selected according to the appropriate definition given above, another strategy for sequence detection in the face of possibly varying playback speeds is to consider only the closest proximity of each vector in the reminiscence phase to any vector in the encoding phase and quantify the relative temporal order of these values. Suppose $X = [x_i]$ $(i = 1, t_1)$ and $Y = [y_j]$ $(j = 1, t_2)$ are two spike rate matrices from the same set of N cells recorded during the encoding and reminiscence periods, respectively (the lowercase letters refer to the constituent state vectors). Let $r_{ij} = \text{corr}\ (x_i, y_j)$ be the correlation between two state vectors in X and Y and let $d_j = \max r_{ij}$ be the maximum correlation of state vector j in Y with any state vector i in X. Let $\alpha_j = i^{dj}$ be the index of X corresponding to the maximum correlation d_j and $\tau_j = \alpha_j - \alpha_{j-1}$ be the time lag in X corresponding best to the trajectory element $(y_{j\text{-}1}, y_j)$ in Y. If μ_τ, the mean value of τ, is significantly greater than zero, then one can conclude that there is significant replay in Y of trajectory segments of X, and the value of μ_τ reflects the mean relative playback speed. The shape and spread of the distribution of τ will provide information about the consistency of playback speed.

Suppose we plot d_j vs j. If the data contain some repeated sequences separated by regions of nonrepeats (for example if Y contains sequences that do not appear in X but are reflections of some other memories), then in general d will not be distributed randomly about its mean value. There will be runs of high values interspersed with runs of low values and there will be a significant correlation between d_{j-1} and d_j.

INTER-REGION INTERACTIONS AND THE PROCESS OF MEMORY CONSOLIDATION

Let us finally consider the possible dynamics of regional interactions during memory reactivation. One model for memory consolidation is that, during behavior, patterns that occur in the neocortex become associated with patterns that occur in the hippocampus, in such a way that subsequent reactivation of the hippocampal pattern may cause reactivation of the corresponding neocortical pattern (Marr, 1971; McClelland, McNaughton, & O'Reilly, 1995; Squire, 1982; Squire, Cohen, & Nadel, 1984). This might occur, for example, through modifiable bidirectional connections at each successive stage of information processing in the cortical hierarchy from primary sensory areas such as V1 up to the hippocampus (see Felleman & van Essen, 1991). Let us consider for example, the possible relationship between some rather low-level cortical area (C) and the hippocampal area CA3 (H), which has traditionally been assumed to be where autoassociation takes place, by virtue of attractors formed by modification of the abundant intrinsic connections there (Marr, 1971; McNaughton & Morris, 1987; Treves & Rolls, 1991). Suppose that these areas are separated by several levels in the synaptic hierarchy, such that there is a substantial delay before a pattern in one area affects the pattern in the other. Depending on connectivity, the delay in the forward (C → H) direction may be different from the delay in the backward (H → C) direction. If, during behavior, pattern C imposes patterns on H, whereas during SWS, memories are spontaneously recalled in H and reimposed on C, then the relationship between temporal patterns in the two structures will be different in the two states. Suppose one records simultaneously cells in both C and H during behavior

and during subsequent sleep. Then one again has vertical and horizontal partitions of the spike rate matrix as illustrated in Fig. 1 (e.g., M^C, M^H, and $S2^C$, $S2^H$). Consider first the temporal correlation matrices for the two epochs S2 and S2. Assuming accurate recall, they will be highly similar, both within the diagonal partitions (within-area correlations) and within the off-diagonal partitions (between-area correlations), because, as stated above, the simple correlations depend only on the state space occupancy distributions and not on temporal order. Such similarity between sleep and waking of the off-diagonal partition of the temporal correlation matrix was observed by Qin et al. (1997), in a study of hippocampal–parietal interactions. Interestingly, although the bias matrices for the cross-correlations were also similar within the diagonal (within-area) submatrices, no significant similarity was observed for the off-diagonal matrices. This lack of similarity is at least consistent with the hypothetical reversal in the direction of information flow.

One possible approach to verifying the information flow reversal hypothesis would involve time-shifting the C and H submatrices of S2. If the hypothesis is correct, then there should be an optimum time shift that maximizes both the similarities of the bias matrices and the integral of the M–S2 submatrix of S.

CONCLUSIONS

Clearly, the study of the neurophysiology of reminiscence is in its infancy. This is largely a reflection of the fact that, although progress seems to be accelerating, methods for simultaneously monitoring multiple single neurons are in their infancy. The refinement of these methods is crucial for this field, and the lack of suitable data sets, i.e., sets containing large numbers of cells, for analysis has led to a corresponding lack of effort to develop suitable analysis methods to address the problem and a general lack of appreciation for some of the possible complexities. It is hopefully clear from the foregoing, however, that the temporal correlation method is limited in its scope and that unraveling the details of the process of memory retrieval at the neurophysiological level will require the elaboration of the state vector approach.

REFERENCES

Abeles, M., & Goldstein, M. H. (1977). Multiple spike train analysis. *Proceedings of the IEEE,* **65,** 762–773.

Aertsen, A. M. H. J., Gerstein, G. L., Habib, M. K., & Palm, G. (1989). Dynamics of neuronal firing correlation: Modulation of "effective connectivity." *Journal of Neurophysiology,* **61,** 900–917.

Amit, D. J. (1989). *Modelling brain function: The world of attractor networks.* New York: Cambridge Univ. Press.

Buzsàki, G. (1989). Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience,* **31,** 551–570.

Farah, M. J. (1995). The basis of mental imagery. *In* M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 963–975). Cambridge, MA: MIT Press.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex,* **1,** 1–47.

Georgopoulos, A. P., Kettner, R. E., & Schwartz, A. B. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *Journal of Neuroscience,* **8,** 2928–2937.

Gerstein, G. L., Perkel, D. H., & Dayhoff, J. E. (1985). Cooperative firing activity in simultaneously

recorded populations of neurons: Detection and measurement. *Journal of Neuroscience,* **5,** 881–889.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the USA,* **79,** 2554–2558.

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society of London B,* **262,** 23–81.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review,* **102,** 419–457.

McNaughton, B. L., & Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences,* **10,** 408–415.

O'Keefe, J., & Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus,* **3,** 317–330.

Palm, G., Aertsen, A. M. H. J., & Gerstein, G. L. (1988). On the significance of correlations among neuronal spike trains. *Biological Cybernetics,* **59,** 1–11.

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of Neuroscience,* **9,** 2907–2918.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms.* Princeton NJ: Princeton Univ. Press.

Qin, Y.-L., McNaughton, B. L., Skaggs, W. E., & Barnes, C. A. (1998). Memory reprocessing in corticortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London B,* in press.

Shen, B., & McNaughton, B. L. (1994). Modeling the spontaneous reactivation of experience specific hippocampal cell assemblies during sleep. *Hippocampus* **6,** 685–692.

Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. J. (1994). A model of the neural basis of the rat's sense of direction. *Advances in Neural Information Processing Systems,* **7,** 173–180.

Skaggs, W.E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following experience. *Science,* **271,** 1870–1873.

Skaggs, W. E., & McNaughton, B. L. (1998). Neuronal ensemble dynamics in hippocampus and neocortex during sleep and waking. *In* H. Eichenbaum and J. L. Davis (Eds.), *Neuronal ensembles* (pp. 235–246). New York: Wiley.

Sparks, D. L., & Mays, L. E. (1990). Signal transformations necessary for the generation of saccadic eye movements. *Annual Review of Neuroscience,* **13,** 309–336.

Taube, J. S., Muller, R. U., & Ranck, J. B., Jr. (1990). Head direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *Journal of Neuroscience,* **10,** 420–435.

Treves, A., & Rolls, E. T. (1991). What determines the capacity of autoassociative memories in the brain? *Network,* **2,** 371–397.

Tsodyks, M., Skaggs, W. E., Sejinowski, T. J., & McNaughton, B. L. (1996). Population dynamics and theta rhythm phase precession of hippocampal place cell firing: A spiking neuron model. *Hippocampus,* **6,** 271–280.

Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science,* **261,** 1055–1058.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science,* **265,** 676–679.

Zhang K., Ginzburg, I., McNaughton, B. L., & Sejnowski, T. J. (1998). Interpreting neuronal population activity by reconstruction: A unified framework with application to hippocampal place cells. *Journal of Neurophysiology,* **79,** 1017–1044.