2. Specific Aims: We will probe the circuitry of spatial working memory using multi single-unit recording in the dorsolateral prefrontal cortex (DLPFC) to study how neuronal activity changes over long delays, and use this information to better understand the circuitry underlying working memory and spatial computations.

Spatial working memory decays over time. The neuronal correlates of this decay, measured at the single neuron level, will help elucidate how spatial information is encoded and how spatial working memory functions at the singe neuron and small circuit level. Single cells in DLPFC have circumscribed mnemonic fields, analogous to receptive fields (Goldman-Rakick 1987; Funahashi, Bruce, and Goldman-Rakick, 1989). We will study whether and how these fields change as memory decays. We consider this in the computational context of an attractor network, the premier model for spatial working memory (Compte et al. 2000). These models posit a "bump" of activity, formed by the cooperative action of many cells, each with their own mnemonic field. In this model, the attractor circuitry maintains the shape and amplitude of the bump over a memory period, but noise may cause the bump to drift in random directions. In the model, it is this drift that underlies memory decay. We will test this strong hypothesis, and consider several (non-exclusive) alternatives – that with the passage of time, the bump amplitude decays, width broadens, or that fewer neurons participate in the bump. Correlating neuronal changes over time with behavioral measures of memory decay will reveal how spatial information is encoded and read out from these neurons, and will provide strong constraints for the architecture of working memory attractor networks.

Aim 1: Test how the mnemonic fields of individual neurons in DLPFC change as memory decays, and relate these changes at the single cell level to a behavioral assay of memory decay over time. We will obtain tuning curves, in time and space, for the mnemonic fields of individual cells. By recording from multiple cells simultaneously, we will have both absolute and relative information about how these curves change over time, and how these changes relate to trial-by-trial changes in spatial working memory content. From these data, we will determine whether the activity bump formed by the population drifts, drops in amplitude, broadens or thins out over a long delay period. We will also determine which of these changes are correlated with changes in memory decay, e.g., the bump may drift and drop in amplitude, but only the drift may be correlated with memory decay reflected in behavioral measures. The results will reveal how spatial information is read out from DLPFC neurons (e.g., is activity normalized or does only the location of the bump matter), and will allow us to refine existing computational models of working memory so that they fit our data.

Aim 2: Test the generality of the spatial working memory mechanisms revealed in Aim 1 by comparing the neuronal correlates of storing a single spatial location with multiple locations. We will construct tuning curves when a macaque is holding two spatial locations in working memory and examine how these tuning curves compare to tuning curves generated when each of the same two locations is maintained individually. In an animal trained to hold either one or two targets in memory, DLPFC may contain two independent memory networks. When a single target is held, one or both networks may be recruited. Alternatively, a single attractor network may be used to store two separate locations. Finally, a completely different strategy may be used to store two targets, e.g., a line may be stored rather than two discrete points. We will test each of these hypotheses. The results are critical to understanding how general the attractor network framework might be, and to what extent spatial working memory networks are general purpose versus purpose-specific.

Working memory underlies a wide array of cognitive functions, and many common psychiatric disorders include deficits of spatial working memory (e.g., schizophrenia, Alzheimer's Disease) (Park and Holzman, 1992). Understanding how working memory is supported and structured within the brain will advance our understanding of cognition and the variety of functions that depend on working memory, as well as provide clues about the pathophysiology of psychiatric disorders.

3a. Significance: This project will make use of multi single-unit recording to test several hypotheses regarding the neuronal architecture of spatial working memory. Working memory plays a crucial role in many cognitive functions and as a result has become an important topic of psychophysical, physiological, and computational study. Impairments in working memory – and in particular spatial working memory – are major factors of many disorders such as Alzheimer's Disease and schizophrenia (Park and Holzman 1992, Green et al. 2000). Understanding the neural mechanisms and circuitry involved in working memory is vital for understanding cognitive functions that depend on working memory as well as for understanding the pathophysiology of related disorders. Additionally, attractor network models are well-suited to describing a wide array of qualitatively different cortical functions using a common base architecture, with only small differences in synaptic and cellular parameters (X.J. Wang, 2008). Probing the architecture of cortical attractor networks can provide insight about cortical functions that share basic attractor network architecture.

Because decay is a fundamental property of working memory, studying how neuronal activity changes as a function of working memory decay can reveal fundamental information about how memory circuits are implemented. Although there has been some study of working memory neuronal activity in the context of error trials (e.g. Funahashi et al. 1989), surprisingly, to our knowledge there have been no studies that look at the effects of decay at the neural level. Furthermore, the effects of decay cannot be adequately studied in the short delay intervals (1 to 3 seconds) typically found in the literature, since the amount of decay over these intervals is minimal (preliminary findings). To address these issues we will use multi single-unit neuronal recording to systematically look at correlations between drops in behavioral memory performance and changes in the neuronal activity of working memory circuitry over long delay periods of 15 to 20 seconds. These data will be interpreted in the framework of attractor networks. Some of the predictions made by these types of networks have seen support from behavioral studies (e.g Chumbley et al. 2008, Macoveanu et al. 2007). However, many of the specific predictions regarding properties of the spatial working memory neuronal architecture are better suited for neurophysiological study.

We will also test whether the spatial working memory mechanisms we find can generalize to multiple targets, or if instead the structure of working memory changes when more locations must be remembered. The specific ways in which neuronal activity differs between single and multiple target conditions will reveal if spatial working memory networks are general purpose or purpose-specific. We will test several plausible multi-target architectures to determine exactly how the structure of the working memory network changes as more items are added.

3b. Innovation: Although behavioral studies (e.g. Chumbley et al. 2008, Macoveanu et al. 2007) and computational studies (Compte et al. 2000) have investigated spatial working memory decay, no studies have investigated decay at the neural level. In addition, most working memory studies have only looked at very short delay intervals of one to three seconds. However, these short delay intervals are not sufficient to study the effects of working memory degradation. In the proposed study we will close these gaps in the working memory literature by comparing how neuronal activity changes correlate with working memory behavioral performance degradation using delay periods of 15 to 20 seconds. The data we collect will reveal the encoding and read-out mechanisms of spatial working memory circuitry and will test key hypotheses and predictions about working memory attractor networks. Additionally, neurophysiological studies have not addressed how the spatial working memory network behaves when more spatial locations must be remembered. We will test a variety of plausible hypotheses and determine whether or not the single-target structure of spatial working memory generalizes to multiple targets, or if an alternative strategy is used instead. Our results will provide strong constraints for the network architecture and will put us in a position to suggest possible revisions to attractor network models of working memory.

Furthermore, although working memory has seen much study at many different levels of analysis – from human brain imaging, to single-unit recordings in non-human primates, to computational modeling – these approaches remain largely independent. While these studies have had much success in determining some of the neural mechanisms and architecture involved in the maintenance of working memory, it is important to combine the different levels of analysis in order to unify the field of working memory. The work in this proposal is innovative in this respect, serving as a critical part of a larger collaborative project aiming to bridge the gap between previous independent approaches. The overall project, of which this proposal is a significant part, will make use of a matched experimental design to compare memory activity across species (humans and non-human primates) and levels of analysis (human imaging and behavior, and monkey imaging, recording, and behavior). The results of these experiments will be analyzed and interpreted in the context of spatial working memory computational models such as those by Compte and his colleagues (2000).

3c. Approach

Overall Goal: Lesion studies in PFC produce significant working memory deficits, implicating the PFC as an important structure for working memory functions (e.g. Jacobsen 1936, Milner 1963, Goldman-Rakic 1987). Neurophysiological studies in non-human primates later demonstrated that PFC cells sustain their activity after a presented stimulus has been removed (Fuster and Alexander 1971, Kubota and Niki 1971). This sustained activity is a likely candidate for the neural correlate for working memory. Within PFC, different neurons are tuned to different spatial locations (Funahashi et al. 1989).

Computational studies have also contributed significantly to our understanding of working memory mechanisms. For a review of the variety of computational models characterizing working memory functions see Durstewitz and colleagues (2000). The prevailing models that simulate sustained activity and the ones that have seen by far the most use and success are attractor network models that use recurrent excitation and inhibition to maintain the sustained activity. Compte and colleagues (2000) have specifically applied this framework to model spatial working memory. In their model, nodes in a fully connected network are each tuned to a specific angular location, with the population of nodes ranging from 0 to 360 degrees. Neurons that are tuned to similar locations have excitatory connections with each other. The strength of excitatory connections decreases as the distance between the spatial locations to which they are tuned increases. The network is also structured with broad inhibition, so that excited neurons activate inhibitory interneurons with feedback connections to all neurons in the network. Therefore, neurons that are tuned to far-apart locations only inhibit each other, whereas neurons that are tuned to similar locations have a net excitatory effect on each other. The result of this model and others like it is that a spatial memory is represented as a continuous "bump" of activity on the axis of neurons tuned to the full range of spatial locations. Typically the neurons also receive random signals to simulate noise in the external inputs.

As a result of their architecture, these models make key predictions about spatial working memory that can be tested to determine if the model architecture is present in the neural circuitry. In these models, the bump of population activity maintains its structure during delay, and memory degradation is the result of random noise inputs that cause the "bump" of the population activity to randomly drift. Trial-to-trial this is equivalent to a synchronized drift of the spatial tuning functions of the individual neurons. That is, in a given trial the drift of the population is in a random direction, and this manifests as a synchronized drift of the individual tuning curves of single neurons in that direction in which the population activity is moving. Because the drift is random, its range increases as a function of time and the effect therefore becomes more apparent as delay increases. Behaviorally, the effect of delay manifests as a decrease in the precision (the variance around the target stimulus location) of the memory over time, as shown in our preliminary data. Figure 1 shows behavioral performance from a monkey in the oculomotor delay response (ODR) task during 3 different delay periods. The results show that precision decreases as a function of delay, a finding that is consistent with the hypothesis that the peak of activity drifts randomly around the target location as a function of delay, causing memory degradation.

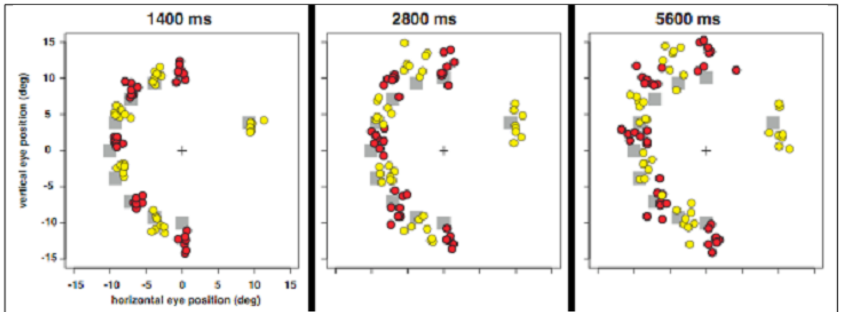

Figure 1 Memory-guided saccade endpoints as a function of time in the monkey. Animals fixated a central cross while a memory target (grey square) was briefly presented. After a memory interval of 1.4, 2.8 or 5.6 s, the cross disappeared and the animal saccaded to the remembered endpoint for a liquid reward. Nine targets were to the left of fixation and one to the right. Endpoints for adjacent targets are color-coded yellow or red for ease of distinction. The mean endpoint for each target shifts upwards with time, especially for upper targets, and the dispersion of the endpoints about the mean increases with time.

Other possible mechanisms not typically explored by the models that could account for the degradation of memory performance over time include a change in amplitude (flattening) or width (broadening) of the "bump" of population activity, which on the single-unit level would manifest as a systematic change in the height or width of individual neuron tuning functions. The activity "bump" may also dissipate or thin out. Although the general simplified model (Compte et al. 2000) does not predict a change in the structure of the "bump" of activity, our preliminary single-cell recording data (see Figure 2) suggests that such an effect may exist.

In our preliminary efforts, we have recorded from 58 single neurons in frontal eye fields of two macaques during a short (6s) ODR task and fit spatial tuning curves to the single-unit activity at different points in time. We found no relationship in tuning curve width across delay or across behavioral error (see Figure 2b & 2d), which is in agreement with the models. However, we found that there was an effect in tuning curve height of individual neurons across delay time and across behavioral variable error (see Figure 2a & 2c) which is equivalent to a change in either height or width (or both) of the "bump" of population activity. This result not predicted by the models such as Compte et al. (2000) which instead predict that behavioral is due to random drift of the population activity "bump" induced by external noise inputs.

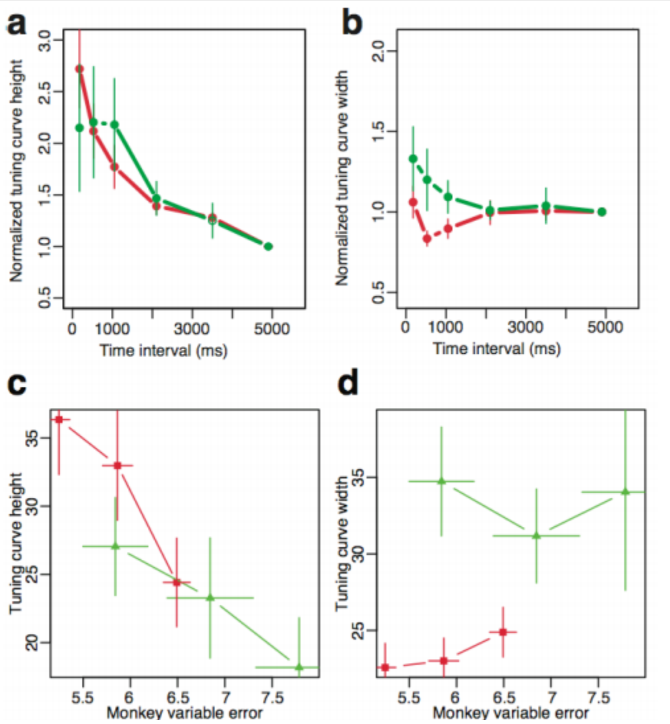

Figure 2. Neuronal-behavioral correlations. **A)** the tuning curve height fell exponentially with time in two animals (red and green) but was flat in a third. The third animal had intensive training in a specialized spatial cognition task. **B)** Tuning curve width was not systematically related to time, matching the predictions of computational work from Compte et al. (2000). **C)** In two animals, tuning curve height was linearly related to behavioral (saccade endpoint) variability. This was not true in the third animal. **D)** Tuning width as a function of variable saccade endpoint error. There was no systematic relationship between tuning width and variable error in any animal.

Furthermore computational studies have examined how the spatial working memory system may behave when multiple locations need to be remembered. Typically, most computational studies assume a single network which responds to all of the presented stimulus locations. Under this assumption when two spatial locations that are close together are maintained in the network the overlapping excitatory connections in the neurons intermediate to two remembered locations should cause the "bump" of activity to shift to those intermediate locations. In contrast, for locations that are far apart there are no overlapping excitatory connections and the activity is not predicted to show this shift. These predictions have seen some support in behavioral experiments (e.g. White et al. 1994, Macoveanu et al. 2007, Chumbly et al. 2008) but no neurophysiological studies exist to systematically study these effects over long delay periods. In addition, alternatives to the single network may be considered. The DLPFC may contain two or more independent networks or may employ different strategies when multiple target locations need to be remembered. Our experimental paradigm will allow us to test specific hypotheses about how the structure of working memory may change as the number of remembered items increases.

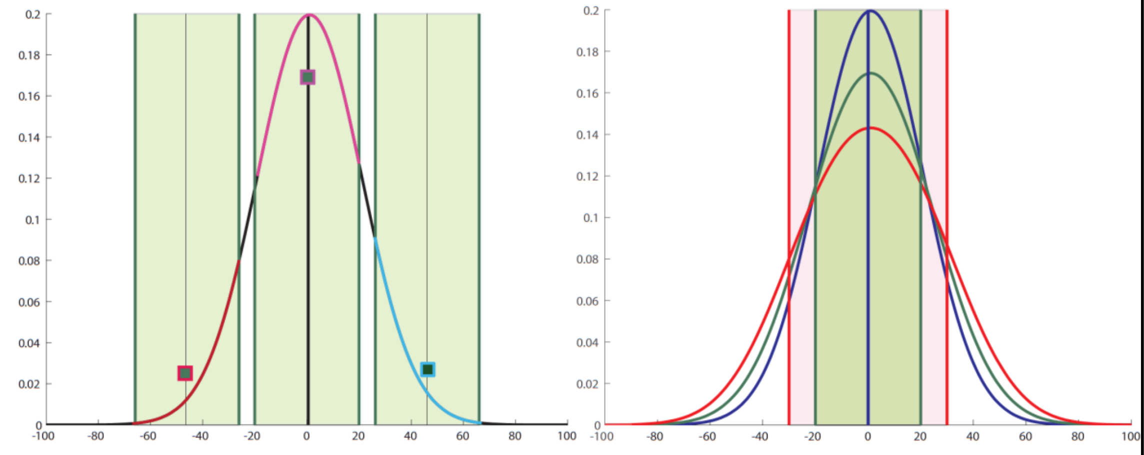

Figure 3. How population activity drift appears to single-unit methods. Left – Individual neuron tuning curve at zero delay, and 3 points of how its tuning curve appears after a given delay in which the population activity can drift in the range depicted by the green rectangles. Presenting target stimuli at a given location over multiple trials and recording after a delay during which the population activity has drifted will produce activity that falls somewhere in the range of the drift. Averaging the values of multiple trials produces points of the single neuron tuning curve after such a delay. Right – Calculating many such points continuously (by presenting targets across the entire range from -180 to 180). Notice that as delay increases (red > green > blue, as depicted by the corresponding rectangles) tuning curves appear to widen. As a result, we cannot distinguish between widening or random drift in the population activity using single-unit methods.

When using single-unit recordings some of the mentioned mechanisms for memory degradation cannot be readily distinguished from one another. Specifically, single-unit tuning functions constructed from multiple trials at a certain point in a delay period will look identical for the cases of random drift and broadening of the population activity “bump”, and because of this single-unit recording techniques cannot determine which of these hypotheses is correct. For an illustration of this, see Figure 3. We can get around this limitation in the preliminary data by recording multiple single units all at the same time, which allows us to study correlated activity changes between neurons with overlapping response fields. Examining these correlation patterns will allow us to determine how the “bump” of activity evolves over a delay period (see Data Analysis). We will also make use of longer delay periods (~15 to 20 seconds) in the proposed study as the effects we are looking for are predicted to become larger as delay increases. In addition, we will record under these conditions with multiple target stimuli to determine how additional stimuli affect the evolution of the “bump” of activity during a delay and specifically whether additional stimuli accelerate mechanisms of working memory sustained activity degradation. Our results will enable us to better test the computational models and suggest model revisions.

Subjects: Three macaques will be used as subjects. The macaques will be fitted with a prosthetic to keep their heads stabilized, as well as a scleral search coil to track eye movement. They will also be fitted with a recording chamber over dorsolateral PFC. Fitting will be done under inhalation anesthesia and post-operative analgesics will be provided. All procedures discussed conform to the National Institutes of Health guidelines and are approved by the Washington University Institutional Animal Care and Use Committee.

During the experimental tasks the macaques will be seated in primate restraint chairs and placed in a completely dark room facing a white screen upon which task-relevant stimuli will be projected. A CRT projector (not an LCD projector) will be used to keep the environment dark. Eye position will be tracked using earth-mounted 4’ rectangular field coils and logged every 2ms. Visual stimulus presentation times are tracked to within 1ms.

Behavioral Task: Macaques will be trained to perform the oculomotor delay response (ODR) task (Funahashi et al. 1989). The ODR task is a widely used paradigm across many successful studies of spatial working memory. Many attractor network models have also based their simulations on this task (e.g. Compte et al. 2000). Using this task in our study will allow us to compare our results with previous experimental and computational studies. In our version of the task, a trial begins with the subject fixating at a central point on a blank screen. After 500ms of fixation a target will flash briefly in a random angular location of constant radial distance from the fixation point while the subject continues to fixate. After a delay period the fixation point will disappear, and the subject will be required to make a saccade to the target location. For our experiments we will be using long delay periods of 15 to 20 seconds. We will also train our subjects to perform the task with two sequential target presentations, after which both target locations must be reported. Reward (water or juice) will be given as each correct target is identified. To encourage reporting targets in the order presented reward for the second target will be available only once the first target has been reported. Reporting the second target first will be counted as correct but will not provide any reward to the subject, although reward may still be earned for a subsequent report of the first target. The subject will be required to reacquire the fixation point after reporting the first target and prior to reporting the second target. Finally, early breaking of fixation during the delay period will result in the current trial being aborted with no reward.

Multi Single-unit Recording: Multi single-unit recordings in macaque dorsolateral PFC and related areas – identified in parallel fMRI experiments in monkeys and humans – will be performed while subjects participate in our behavioral task. An alpha-omega system capable of recording from 16 simultaneous channels will be used. We will use multiple electrodes to simultaneously isolate two or more neurons with overlapping spatial working memory response fields and record their activity while a macaque participates in our behavioral task. The feasibility of isolating two neurons with overlapping response fields is indicated by previous spatial working memory studies which have suggested that memory fields in PFC have a topographic columnar organization in which neighboring columns respond to similar locations in the visual field. In addition, the tuning in these neurons tends to be fairly broad (e.g. Funahashi et al. 1989, 1990; Goldman-Rakic 1996; Rainer et al. 1998; Sawaguchi 1996). As a result of these properties we believe isolating neurons with overlapping response fields will be feasible.

Data Analysis: We will use recordings collected from simultaneous isolations of two or more neurons during our behavioral task in order to construct individual spatial tuning curves of neuronal activity. Both correct trials and incorrect trials in which the subject did not break fixation during the delay period will be used in the analysis. Trials in which fixation was broken prior to the end of the delay period will be discarded. We will divide the delay period into several equal-sized time bins and fit a Gausian tuning function to the single-unit activity of each neuron. We will compare the peak locations, the widths, and the amplitudes of these curves as a function of delay and behavioral error to test specific hypotheses about the spatial working memory cortical network.

Aim 1: Test how the mnemonic fields of individual neurons in DLPFC change as memory decays, and relate these changes at the single cell level to a behavioral assay of memory decay over time. In our effort to probe and better understand cortical attractor networks, specifically those involved in spatial working memory, we will test specific hypotheses about working memory degradation mechanisms. We will test these hypotheses by comparing the error in behavioral report of the target location to correlations in the simultaneously recorded activity of neurons with overlapping response fields. The decay mechanism most explored by computational studies is random drift. Random drift predicts that the population activity "bump" retains its structure but randomly drifts along spatial locations due to external noisy inputs to the circuitry, causing decay of the initial location information. In the case of random drift, error in behavioral reporting of a target location intermediate to the preferred directions of two such neurons (e.g., the red spot in Figure 4) will manifest as anti-correlated changes in their activity as memory decays. When a target is presented at non-intermediate location (e.g., the green spot in Figure 4), then correlated changes in activity are expected. Alternatively, the population "bump" may not retain its structure, and may instead decrease in amplitude, become broader, or both, leading to memory decay. In the case where the structure of the bump changes uniformly (e.g. amplitude, width) neural data will show correlated changes in the activity of the two isolated neurons regardless of the location at which a target is presented. The specific pattern of relative activity changes in each pair or group of simultaneously isolated neurons will allow us to determine exactly how the structure of the population activity changes as a function of delay, and comparing this pattern to behavioral error, as measured by the deviation of a response saccade from the actual target location in the ODR task, will reveal the network mechanisms responsible for working memory decay. The results will give us a better understanding of cortical attractor networks and allow us to propose revisions and constraints to working memory network models.

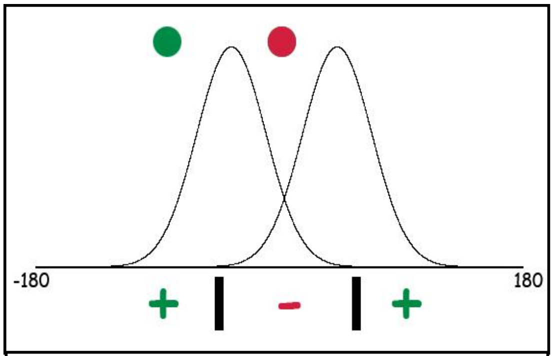

Figure 4. Tuning curves of two neurons with overlapping response fields. Targets presented on the outside of their response fields (green) lead to correlated activity changes as the population activity drifts along the spatial location axis. Targets presented between their response fields lead to anti-correlated changes due to opposite signs in the slopes of their curves in this range. This result will not be observed if the "bump" of population activity changes in structure rather than randomly drifts, allowing us to distinguish these two possibilities.

Aim 2: Test the generality of the spatial working memory mechanisms revealed in Aim 1 by comparing the neuronal correlates of storing a single spatial location with multiple locations. We will construct tuning curves when a macaque is holding two spatial locations in working memory and examine how these tuning curves compare to tuning curves generated when each of the same two locations is maintained individually. Computational models of spatial working memory have generally worked under the assumption of a single attractor network able to maintain one or several activity "bumps" at each presented target location. Alternatively two (or more) independent networks may exist, each encoding one of two targets. To distinguish between these possibilities we will present two targets in far apart locations. One target will be located in the preferred angular location of an isolated neuron (as determined in single-target trials) while the second target will span a range of locations (between 135 and 180 degrees) away from the first target. If there is a single network, the isolated neuron will always encode the target in its preferred direction and will show a unimodal distribution in its response. On the other hand, if there are two networks the isolated neuron will encode the preferred target in some trials (high firing rate), while on other trials it will encode the non-preferred target (low firing rate). The result will be a bimodal distribution in its response. It is important to note that in the case of multiple networks target assignment to a network may not be random, and instead may be based on a rule such as temporal order (first target is always encoded by the same network), proximity to the fovea, or some other relative relationship between the two targets. Care will be taken to present targets so that there are no consistent relationships between them.

If our findings are consistent with a single network, then future study can address a variety of interesting questions. For example, how does the network maintain two target locations that are close together? The single network attractor framework predicts that overlapping excitatory connections of neurons with similar preferred directions should pull the population activity “bumps” of two such target closer together, a result that has seen support from behavioral studies (Chumbley et al. 2008, Macoveanu et al. 2007). With even closer target locations the activity elicited may collapse to a single-peaked “bump”. Alternatively, multi-peaked attractor states may be possible. The effects of adding more locations to the single network can also be explored, opening up questions regarding how a single network can increase the number of target locations it can maintain over the course of training.

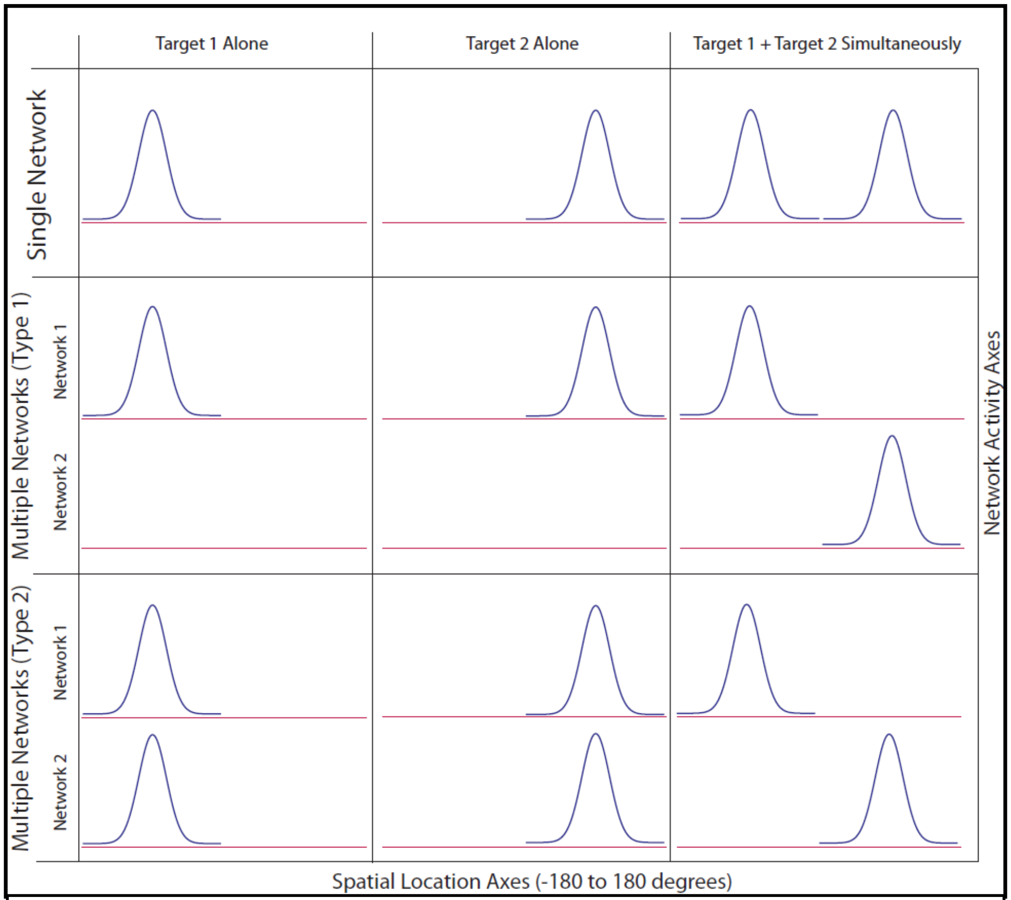
Target 1 Alone
Target 2 Alone
Target 1 + Target 2 Simultaneously
Single Network
Multiple Networks (Type 1)
Network 1
Network 2
Multiple Networks (Type 2)
Network 1
Network 2
Network Activity Axes
Spatial Location Axes (-180 to 180 degrees)

Figure 5 Possible network architectures for multiple targets. Each panel shows network activity in response to two targets located far apart. From left to right, the columns show target 1, target 2, and target 1 + target 2 conditions. In the single network each cell always responds to the same target. However, in the case of multiple networks a given network and therefore a given cell may encode a different target from trial to trial. Note also that in the Type 2 multiple network the number of cells active in the target 1 + target 2 condition is the same as the number of cells active in each of the two single target conditions. This is not true for the Type 1 multiple network architecture or the single network architecture.

If our findings are consistent with multiple independent networks we will further probe the network properties to determine how spatial information is encoded. In one possible encoding scheme each network responds to only a single target such that as the number of targets increases the number of active networks also increases. In another plausible encoding strategy all of the networks may respond when a single location is presented, but when a second location is presented the networks may be distributed such that some networks maintain the first locations and some maintain the second location. To distinguish between these two possibilities we will compare neuronal activation in one target and two target conditions. In the first encoding scheme the number of networks recruited increase as the number of targets increase. Therefore, there will be twice as many active cells in the two target condition than in the one target condition. This finding is not consistent with the second encoding scheme because a single target activates all of the networks.

Finally, networks may use an alternative strategy, such as maintaining relative relationships between two or more targets. One example to illustrate this point is the maintenance of a line between two locations instead of maintaining the locations. This strategy might be implemented by remembering a single location and a vector from the first location to the second. The neural activity patterns we record will allow us to determine if a more complex encoding scheme is used than that predicted by the attractor network framework and will allow us to begin to examine the structure working memory takes when multiple items must be maintained. The results of this aim are critical for determining the structure spatial working memory information takes as more items are added.