

Physiological and Statistical Approaches
to
Modeling of Synaptic Responses

P. G. Patil¹, M. West⁴, H.V. Wheal⁵, and D.A. Turner^{1,2,3}

¹Neurosurgery and ²Neurobiology, Duke University Medical Center
and ³Durham VAMC, Durham, NC 27710 USA

⁴Institute of Statistics and Decision Sciences
Duke University, Durham NC 27708-0251 USA

⁵Neurosciences, University of Southampton, Southampton, UK

Address for Correspondence:

Dennis A. Turner, M.A., M.D.

Box 3807, Neurosurgery

Duke University Medical Center

Durham, NC 27710 USA

(919)-684-6706, FAX (919)-681-8068

Email for D.A. Turner: Dennis.Turner@duke.edu

Email for P.G. Patil: Patil003@duke.edu

Email for M. West: MW@stat.duke.edu

Email for H.V. Wheal: H.V.wheal@soton.ac.uk

RATIONALE OF CHAPTER

This chapter is a review of current physiological approaches to understanding synaptic function, including a detailed account of methods to acquire optimal data for further quantitative analysis and the limitations of these methods. The statistical modeling approaches to synaptic data analysis are then summarized, including particularly binomial, Poisson and more complex (less restrictive) methods. These approaches use several algorithms or statistical methods to gain insight into the data using statistical models, including maximum likelihood and Bayesian approaches. Physiological examples are then given as to how the Bayesian site approach is used for actual data analysis. The degree of heterogeneity between synaptic release sites and how these various statistical approaches compare are then summarized.

ABSTRACT

Synaptic function centers on the physiological behavior of individual and combined synaptic sites, defined at the anatomical level as a unit consisting of a single postsynaptic density and presynaptic release region. However, multiple synaptic sites are usually activated together in a physiological response, even between two individual neurons. These multiple sites may be either grouped together, as at the neuromuscular junction, or distributed spatially, as in a complex dendritic neuron. The physiological efficacy of responses at each synaptic site can be defined by a few critical parameters, including probability of release, postsynaptic site response amplitude ('quantal' amplitude), the degree of fluctuation around that characteristic amplitude ('quantal' variance) distributed both over time and between sites, and the efficacy of electrotonic conduction between the synaptic site of origin and the output or summation site. These parameters all vary as a highly dynamic process, fluctuating with the history of use at each synapse over short and long time frames, and between adjacent synaptic sites on dendritic trees. Such heterogeneity between synaptic responses can lead to summated synaptic responses which are difficult to interpret in terms of the behavior of each individual synaptic site. However, the influence of these various factors on synaptic plasticity at multiple time spans is critical to further understand mechanisms of development, memory and neuronal circuitry function. Statistical models for analysis of synaptic function have become increasingly sophisticated to account quantitatively for these diverse sources of variability in the explanation of physiological processes. We review the use of various statistical models, which are used to infer descriptions of synaptic release characteristics under appropriate physiological conditions, including a novel Bayesian analysis format, to investigate formally many of these sources of variability. Together, constrained physiological preparations and sophisticated statistical models have defined both the presence of considerable heterogeneity between synaptic sites and mechanisms, which may contribute to synaptic plasticity in the CNS.

CONTENTS

1. Introduction

- 1.1 Modeling Synaptic Function in the CNS
- 1.2 Complexity Introduced by Synaptic Heterogeneity and Plasticity
- 1.3 Complexity Associated with Physiological Recordings
- 1.4 Classical Statistical Models

2. A Bayesian Model of Synaptic Transmission

- 2.1 Introduction to the Model and Comparison to Classical Models
- 2.2 The Bayesian Site Analysis
- 2.3 Application of the Bayesian Site Model to Simulated Data
- 2.4 Application of the Bayesian Model to Recorded Data

3. Discussion

- 3.1 Comparison of Simulations and Physiological Data Sets
- 3.2 Analysis of Components in Contrast to Sites
- 3.3 Analysis of Physiological Data Sets
- 3.4 Conclusions and Future Perspectives

Mathematical Appendix

References

1 INTRODUCTION

1.1 Modeling Synaptic Function in the CNS

Transmission of signals across synapses is the central component of nervous system function (Bennett and Kearns, 2000; Manwani and Koch, 2000). Primarily, such synaptic transmission is mediated by chemical neurotransmitter substances, which are released into the synapse by the presynaptic neuron and detected by receptors, located upon the postsynaptic neuron. In many areas of the nervous system (outside of the hippocampus and neocortex) there also may be dendritic neurotransmitter release, which appears to behave in a very different fashion from more traditional axonal neurotransmitter release (Ludwig and Pittman, 2003). However, because synaptic structures are microscopic and inaccessible, direct measurement of synaptic properties is not often feasible experimentally. As a result, our understanding of the mechanisms underlying synaptic transmission and modulation of function derives from statistical inferences, which are made from observed effects on populations of synapses. Synapses in the central nervous system (CNS) have the distinct property of being unreliable and probabilistic, hence a statistical framework is critical to define function. A schematic of a typical synapse is shown in Figure 1, indicating that the information transfer is between a relatively secure pre-synaptic action potential (on the left) and a highly insecure neurotransmitter release and subthreshold post-synaptic response. For most CNS neurons, spatial and temporal integration within a neuron is an additional complication, often requiring hundreds of synaptic responses summing together to reach action potential threshold.

Typically, studies of synaptic function have made productive use of statistical methodologies and conceptual frameworks described in classic work at a prototypical (but very unusual) model synapse, the neuromuscular junction (NMJ) (Del Castillo and Katz, 1954). At the NMJ, an action potential in the presynaptic axon arrives at the synaptic terminal and, with a finite probability of occurrence π , triggers the release of neurotransmitter. Following this probabilistic event, neurotransmitter diffuses across the synaptic cleft to bind postsynaptic neurotransmitter receptors. The binding of postsynaptic neurotransmitter receptors produces electrophysiological responses in the postsynaptic neuron, a change of magnitude μ , in the transmembrane current or potential. Because synapses at the NMJ contain multiple sites, which release neurotransmitter in concert following a single stimulus to produce a single end-plate potential, analysis of the activity and parameters of individual release sites requires statistical inference. The NMJ is unusual in comparison to most CNS synapses in that the overall response (a post-synaptic muscle twitch) is highly reliable, so the inherent insecurity of synaptic release is overcome by multiple synaptic sites, located in very close proximity. In contrast, synapses are separated on the surface of CNS neurons, and are much less potent, creating a large capability for synaptic interactions and summation (Craig and Boudin, 2001).

Figure 1 Here

At synapses such as the NMJ, which are designed for security rather than capacity for computational functions, the amount of released neurotransmitter may saturate post-synaptic receptors. Such saturation ensures that presynaptic action potentials are reliably transduced into postsynaptic depolarizations. As the necessary consequence of such reliability, presynaptic neurotransmitter release at such synapses is relatively information poor, since variation in the magnitude of release only minimally affects the magnitude of postsynaptic depolarization. By contrast, at central synapses (Figure 1), the concurrent activation of hundreds of synapses may be required to produce

a postsynaptic response. The requirement of postsynaptic signal integration enables substantial information content. Together with variation in the numbers of simultaneously activated presynaptic terminals, the modulation of presynaptic release at individual terminals and the modulation of postsynaptic responses convey information between CNS neurons. Synaptic modulation may thereby encode highly complex cognitive processes, including sensation, cognition, learning, and memory (Manwani and Koch, 2000). Since multiple CNS synapses may be concurrently activated between individual neurons, the examination of synaptic release and neuronal response characteristics often requires a complex statistical model.

The goal of this chapter is to review available statistical models, their applicability to CNS synapses typically studied, as well as the criteria under which the models may be evaluated by physiological assessment. Complexities associated with synaptic heterogeneity and synaptic plasticity, which statistical models must address, are discussed first. We then explore the qualities of electrophysiological measurements, which enable meaningful statistical inferences to be drawn from experimental data. After a discussion of classical statistical models of central synapses, we present and evaluate the Bayesian approach used in our laboratory to study mechanisms of synaptic plasticity in CNS neurons, in comparison to other types of models available (Bennett and Kearns, 2000).

1.2 Complexity Introduced by Synaptic Heterogeneity and Plasticity

If central synapses were static and homogeneous, simple mathematical models would adequately describe synaptic transmission. For example, end-plate potentials at the NMJ might be assumed to be produced by an arbitrary but fixed number of independent release sites, n , which release their quanta with average probability, p , in response to presynaptic stimulation. Such release characteristics would produce a binomial distribution of end-plate potential magnitudes. Given such release characteristics, the parameters, n and p , could be estimated from a histogram summarizing the magnitude of responses to presynaptic stimulation (Katz, 1969). The computational requirements of CNS neurons necessitate much greater complexity, and an important objective of the statistical modeling of CNS synapses is to reveal the physiological manifestations of this complexity through effects on synaptic release parameters.

Central synapses exhibit both heterogeneity and plasticity—their properties differ both between synaptic sites (intersite variability) and over time (intrasite variability). Such complexity arises from multiple factors including variable packaging of neurotransmitter, neurotransmitter dynamics in the synaptic cleft, agonist-receptor interactions at the synapse, channel opening properties, and postsynaptic modification of receptor proteins (Bennett and Kearns, 1995; Buhl et al, 1994; Craig and Boudin, 2001; Faber et al, 1992; Harris and Sultan, 1995; Korn and Faber, 1991; Redman, 1990). In addition, the dendritic location of postsynaptic synapses, voltage-dependent dendritic properties, and inhibition of dendritic signaling may modulate the effects of synaptic events upon the summation sites of postsynaptic neurons (Magee and Johnston, 1997; Spruston et al, 1993; Turner, 1984). Such location-dependent heterogeneity is supported by experimental data (Bolshakov and Siegelbaum, 1995; Buhl et al, 1994; Harris and Sultan, 1995; Turner et al, 1997).

There are many sources of variability inherent to synaptic activation. Release is dependent on calcium entry into the presynaptic terminal, often resulting in failures of release (Glavinovic and Rabie, 2001). Most neurotransmitters require some uptake mechanism for recycling, as well as a synthetic cycle to redistribute neurotransmitter, resulting in variable neurotransmitter packaging in synaptic vesicles, which changes with development (Mozhayeva et al, 2002). Once a vesicle is

primed and docked for potential release, the duration of docking and amount of neurotransmitter released per evoked action potential is unknown, particularly whether multiple cycles of release can occur from a single vesicle (Bennett and Kearns, 2000). Once neurotransmitter is released into the synaptic cleft, a receptor-ligand mismatch may result in a variable post-synaptic response from trial to trial by stochastic diffusion and binding (Faber et al, 1992). Depending on the postsynaptic density of ligand-gated receptors and the type of receptor a variable postsynaptic response at the synaptic site may be elicited. The typical post-synaptic response may be further altered by variable dendritic conduction pathways to the soma, and by inhibition or enhancing mechanisms on the conduction path (Magee and Johnston, 1997; Turner, 1984). The large chain of events from presynaptic action potential to post-synaptic response at a soma or summing output site includes many sources of variability. Further, all finite samples of data are subject to sampling variability as well as processes of measurement error, which complicate interpretations.

Synapses are not static; they evolve over time. For example, systematic changes in the glutamate receptor phenotype (from NMDA to non-NMDA), probability of release and degree of potentiation have been suggested to occur in the hippocampus during development (Bolshakov and Siegelbaum, 1995; Craig and Boudin, 2001; Durand et al, 1996; Hanse and Gustafsson, 2001). This change in receptor phenotype has also been observed at ‘new’ synapses induced with long-term potentiation (LTP), by observing normally ‘silent’ NMDA synapses (at resting potential). These ‘silent’ synapses can be transformed into active non-NMDA synapses by a combination of post-synaptic depolarization and synaptic activation, presumably with receptor modification induced by the calcium influx associated with the NMDA receptor activation (Isaac et al, 1995). The overall implication of this induction of ‘new’ synapses is an increased efficacy of synaptic function, but the specific parameters underlying this enhancement remain unresolved. Thus, synaptic release parameters at individual sites vary over time and as a function of development, as well as between individual sites; this variability results in a very complex set of responses, even when reduced to a small number of connections, as noted with unitary (single pre-synaptic to single post-synaptic) responses (Debanne et al, 1999; Kraushaar and Jonas, 2000; Pavlidis and Madison, 1999).

Other critical instances in which statistical models may help to clarify the physiology underlying synaptic function include the phenomena of short-term and long-term plasticity (Debanne et al, 1999; Isaac et al, 1996; Stricker et al, 1996b). Short-term facilitation is considered to involve primarily changes in presynaptic factors, particularly probability of release, since these changes occur over a few milliseconds (Redman, 1990; Turner et al, 1997). However, long-term potentiation may exhibit many different mechanisms of plasticity, particularly structural changes if the duration is longer than 10-30 minutes. Due to the popularity of the study of long-term potentiation and the wide variety of animal ages and preparations in which this phenomenon is studied, multiple and controversial mechanisms have been suggested (Bolshakov and Siegelbaum, 1995; Debanne et al, 1999; Durand et al, 1996; Sokolov et al, 2002; Stricker et al, 1996b). Since the evoked response histograms change radically both before and after the induction of long-term potentiation, various interpretations have included both enhanced probability of release and also increased amplitude of synaptic site responses. Few of these physiological studies have shown detailed statistical evaluation of the physiological signals which allow for extensive variation between synaptic release sites, other than a compound binomial approach (Stricker et al, 1996b). Thus, current statistical approaches in the evaluation of long-term potentiation may be premature in that variation between synaptic release sites (and the possible addition of release sites with potentiation) have been insufficiently studied.

There are many physiological situations where inferring the underlying statistical nature of synaptic release may lead to understanding the mechanism of critical processes, for example, memory formation and retention. Thus, statistical interpretation of synaptic responses has a clear rationale in physiological interpretation. However, the ambiguity of statistical interpretation, at least to a physiologist's perceived need for an unambiguous result, has lent an esoteric slant to the field of statistical analysis of synaptic signals, and various models.

1.3 Complexity Associated with Physiological Recordings

Statistical interpretation of data arising from studies of synaptic transmission may be obscured by ambient noise associated with either the neuron or the recording conditions. Korn and Faber (1991) and Redman (1990) have previously defined the appropriate physiological circumstances for a proper (and reliable) analysis of underlying synaptic response parameters. These requirements focus on two-cell recordings (either using sharp or patch electrodes) where clear, unitary (but usually multi-site) responses can be identified and include:

- (1) the stimulation of a single presynaptic spike and control of the presynaptic element;
- (2) optimization of noise level to isolate miniature events to compare to evoked events, with miniature events occurring from same synaptic sites as those from which evoked events emanate;
- (3) resolution of single site amplitude, with minimum cable or electrotonic filtering (usually requiring a perisomatic or proximal dendritic synaptic location);
- (4) direct data on variations in quantal amplitude at a single site;
- (5) morphological identification of number of active sites involved and their dendritic location;
- (6) independent evidence on uniformity of amplitude and probability between sites.

These fairly rigid criteria are primarily designed for two-cell recording and in circumstances where both elements can be stained and then anatomically analyzed. Korn and Faber have extensively performed this type of analysis on goldfish Mauthner cells, which have a glycinergic inhibitory synapse which fulfills these criteria (Korn and Faber, 1991). However, in most other CNS cells, particularly in mammalian systems, unitary recordings are much more difficult to obtain, particularly with evaluation of both physiological responses and the anatomical details of the number and location of synaptic sites between two neurons. Recently, however, the hippocampal slice culture has shown considerable promise, with Debanne et al showing dual recordings between CA3 and CA1 neurons with a small number of connections and interesting potentiation properties (Debanne et al, 1999; Pavlidis and Madison, 1999). This same connection in acute *in vitro* slices has shown much less promise due to the small percentage of CA1 and CA3 cells showing excitatory synaptic connections (Redman, 1990; Sayer et al, 1990; Stricker et al, 1996a). Other situations which have been explored include Ia afferent synapses onto motoneurons in the spinal cord (Redman, 1990) and isolation of axonal afferents onto CA1 pyramidal cells using microstimulation techniques (Bolshakov and Siegelbaum, 1995; Isaac et al, 1996; Turner et al, 1997). All of these excitatory synaptic types, however, show primarily dendritic synapses, which are at some electrotonic distance from the soma and the recording electrode and generally higher noise levels are prevalent, obscuring individual responses. Since there are thousands (25,000) of synapses onto these cells there usually cannot be a rigorous comparison of evoked and spontaneous synaptic currents since the relationship between the origin of the two separate events is unknown.

Another approach is to identify inhibitory connections, which are usually peri-somatic, have a much larger signal to noise ratio, but also which may have multiple synaptic release sites per individual synapse (Kraushaar and Jonas, 2001; Staley, 1999). However, these connections may be identified using two-cell recordings, since cells are close to each other and have a high frequency of synaptic interconnection.

Yet another critical characteristic of electrophysiological data is a low level of signal noise. This is achievable using whole-cell patch electrode techniques, where the baseline noise standard deviation may be less than 1 pA, producing a signal-to-noise ration of 4:1 or better for many excitatory synaptic ensembles. Having full physiological control of a presynaptic cell or axon is critical, so responses in the postsynaptic cell may be clearly linked to the activity of the presynaptic response, which occurs either spontaneously or may be triggered (by intracellular current or axonal stimulation). Voltage-clamp simulations demonstrate that most dendritic synaptic responses are characterized by poor voltage clamp, and that there may be decreased measurement error by using current-clamp modalities with whole-cell patch electrodes (Spruston et al, 1993).

To minimize analytic complexity, experimental preparations should be developed to limit stimulation to as few synapses as possible (generally 8-10). Such stimulation control allows for better differentiation of individual synapses within the patch-clamp recordings. We have detailed the conditions important for capture and analysis of EPSC data in Turner et al (1997). Briefly, these conditions included limitation of the synaptic signals to a single glutamate receptor type (non-NMDA), careful activation of as few synapses as possible through microstimulation and restriction of the analysis to carefully defined data sets in terms of noise and stationarity. In hippocampal slices cultures the number of synaptic connections between CA3 and CA1 pyramidal neurons is enhanced by re-innervation, so that a few synapses may be obtained, even in cells at a significant distance from each other (Debanne et al, 1999).

Statistical analysis of synaptic transmission requires that the experimental preparation be stable over time. All of the statistical techniques require a certain number of evoked response samples to be obtained to be able to perform the analysis satisfactorily. For example, in the initial least mean squares optimization technique employed by Redman (1990) to analyze synaptic plasticity, the minimal critical sample size was approximately 750, assuming a signal to noise ratio of 3 to 1 or better. These samples need to be obtained at a rate which is stable for the preparation and which does not in itself lead to plasticity (either depression or facilitation), which varies from 2 Hz for motoneurons to 0.5 - 0.1 Hz for many hippocampal preparations. The duration of acquiring such a sample, which is presumably not changing during the acquisition and fulfills the requirements for stationarity, is extensive (up to 40 min at a rate of 0.2 Hz). This property of the ensemble of stability of the fluctuations during the period recorded is termed stationarity, and it implies both the absence of a trend and that the fluctuations during the period are not changing in any essential manner. Stationarity can be partly assessed by a trend plot and also partly by analyzing sequential subgroups of approximately 50 for net significant changes (Turner et al, 1997). A Bayesian time series analysis and formal trend analysis has also been developed to statistically assess stationarity (West and Harrison, 1997).

Finally, skepticism about the value of statistical inference of synaptic responses is widespread among physiologists because of the difficulty in assessing whether a solution is either unique or 'good' in an absolute sense and the indirect nature of the analysis, with numerous assumptions often required and the lack of validation of the statistical analysis schemes.

1.4 Classical Statistical Models

Chemical synapses involve common release mechanisms, though the specific parameters of release, type of neurotransmitter elaborated and pre-and post-synaptic receptors involved differ considerably. Such synapses are highly probabilistic with responses from a few synapses showing considerable fluctuation from trial to trial (Bennett and Stearns, 2000; Del Castillo and Katz, 1954; Faber et al, 1992; Harris and Sultan, 1995; Korn and Faber, 1991; Redman, 1990). Statistical models of synaptic release mechanisms have been developed to explain the probabilistic release of neurotransmitter and to infer critical changes at individual synaptic release sites when recordings are usually performed from multiple sites at once (Del Castillo and Katz, 1954). These models have been adapted to synapses in the hippocampus for understanding of development, neural plasticity, and potentiation. The appropriateness of a statistical model used to describe what is occurring at a single synapse depends very much upon the validity of the simplifying assumptions of the model and whether there are independent physiological methods to confirm these assumptions.

Historically, statistical models at the neuromuscular junction centered on binomial models, since it was assumed that vesicles emptied near-identical quantities of neurotransmitter, multiple vesicles were released with each impulse and each vesicle possessed the same likelihood of release (Bennett and Stearns, 2000; Del Castillo and Katz, 1954; Katz, 1969). With pharmacological manipulation the number of vesicles released could be significantly reduced and the limit of the binomial distribution, a Poisson distribution, could be reached. With the Poisson distribution the binomial parameters n (number) and p (probability) are not separately defined but are merged into a single value, m , applicable only when p is very small compared to n . The adaptation of this framework to central neurons proved to be difficult because of the much higher noise level associated with recording using sharp electrodes, requiring techniques to remove the effects of the noise (Redman, 1990).

The maximum likelihood estimation (MLE) approaches were developed to optimize searches for multiple parameters in light of this contaminating noise, using a binomial statistical model extended to a compound binomial allowing differences in probability of release (Bennett and Kearns, 2000). In this model, it is assumed that a certain number of synaptic sites ‘mix’ with ambient noise, summate with each other according to the respective probabilities of the sites, and form a set of responses which represent the sum of these different levels over time. The statistical inference problem thereby becomes to analyze the contributions of individual sites in this mixture, particularly determining the number of sites and the individual parameters of each synaptic site. The MLE approach optimizes a likelihood function over a set of possible synaptic site parameters. The optimization is performed by fixing one group of the parameters and allowing one or only a few critical parameters to vary, such as the number of sites, n . The resulting number of sites (or components) are then convolved with the recorded noise and the result is compared to the original histogram. If the result is not optimal, according to a chosen measure, further iterations are performed. MLE analysis therefore does not necessarily provide a unique solution, as many ‘optimal’ solutions may exist. Recent MLE approaches have also added estimation of quantal variance and comparison to non-binomial models, though are based on a smoothed function approximation to the original data histogram (Stricker et al, 1996a).

Other approaches have included histogram-based smoothing and ‘peakiness’ functions, using the histogram to identify summation of site responses (termed components) rather than analyzing the primary data. This approach is severely limited by how the histogram is constructed (bin width,

etc.) and very similar data may show highly different histogram appearance randomly (Walmsley, 1995). Thus, approaches based on analyzing the exact, original data are preferred over superficial histogram appearance.

Guidelines for suitability of physiological data sets have also been established, based on simulation approaches (Redman, 1990). Minimal samples sizes are in the range of 500-750 data points, each from an evoked synaptic response, and a signal to noise ratio of at least 3 to 1 is required for optimal extraction of component locations. Stationarity of the degree of fluctuation of the signal is also required since if the basic synaptic parameters vary a reliable analysis cannot be performed (Redman, 1990). Stationarity can usually be assessed by trends of the entire data set or analysis of successive subgroups (Turner, 1997). The noise itself can be approximated by either a single or dual Gaussian normal function (σ_{noise}), rather than using the raw data histogram. The results of the MLE analysis are limited in that this particular algorithm often cannot identify a local minimum for parameters and may reach different conclusions based on the starting point used for the initial parameters. Thus, multiple initial parameters must be tried and whether the final result is ‘optimal’ or ‘good’ in some absolute sense cannot usually be ascertained. Identifying the underlying distribution is a much more difficult task, particularly with a single finite sample, since random samples from any distribution of a finite size may be very unrepresentative in terms of their histogram appearance (Walmsley, 1995). An additional limitation of the current MLE methods is that variability between site amplitudes cannot be detected, even though variability in site probability is allowed with a compound binomial distribution. Together, these limitations severely curtail the number of data sets which can be analyzed to a small fraction of those which are derived from experiments.

However, many of the basic assumptions of such quantal models have not been tested in CNS neurons. These assumptions include near uniformity of synaptic release amplitudes between sites (low intersite variability), and limited variation in release amplitudes over time (low intrasite variability). Together, such variability in synaptic strength is termed ‘quantal variance.’ Recently however, considerable interest has arisen regarding both forms of quantal variance (Bolshakov and Siegelbaum, 1995; Turner et al, 1997). Such heterogeneity requires more complex statistical models to interpret parameter changes at individual synapses than either a simple or compound binomial model allows.

One approach has been to identify or presume a single or unitary connection between cells whenever possible, and using either the rate of failures or the average success (potency) to identify critical changes in release parameter values (Bolshakov and Siegelbaum, 1995; Debanne et al, 1999; Isaac et al, 1995; Liu and Tsien, 1995; Stricker et al, 1996a,b). This approach simplifies analysis, at the risk of mis-identification of the physiological processes underlying the change. A more detailed alternative approach would be to define an open statistical framework without such assumptions, and then to allow the experimental data to define the parameter changes underlying synaptic plasticity on long- and short-term time scales.

As a detailed example, quantitative parameters describing synaptic release sites include the number of activated sites on any one trial (k), probability of release (π), and amplitude (μ) and electrotonic conduction from the site of origin to the summation site. The goal of ‘quantal’ analysis is to infer these parameters at single synapses but usually in the physiological context where a number of release sites may be activated concurrently (Korn and Faber, 1991; Redman, 1990). However, a high level of background synaptic and recording noise often occurs during recordings, obscuring much

detail and so requiring statistical inference to deduce the content of the original signal. Though a simple quantal-binomial model may hold for inhibitory synapses onto the Mauthner cell (Korn and Faber, 1991) a compound [or complex] binomial distribution has been suggested to be more applicable in the mammalian hippocampus, with varying probability of release [and likely varying amplitude] between synaptic sites required (Sayer et al, 1990; Turner et al, 1997). Recent studies of small EPSCs (Bolshakov and Siegelbaum, 1995; Turner et al, 1997) have suggested considerable heterogeneity over time at synaptic sites and between synapses. Thus, techniques to analyze parameters at individual synapses have been limited by adapted statistical frameworks and newer models, including maximum likelihood estimation models (MLE; Korn and Faber, 1991; Stricker et al, 1996a,b), Bayesian component models (Cao and West, 1997; Escobar and West, 1995; Turner and West, 1993; West and Turner, 1994), and site-directed models (Turner et al, 1997; West, 1997) may provide improved analytical capabilities. Newer approaches have also indicated that within site and between site variance may be systematically under-estimated, requiring different approaches for estimation (Frerking and Wilson, 1999; Uteshev et al, 2000).

2 NONTRADITIONAL MODELS OF SYNAPTIC TRANSMISSION

2.1 Introduction to the Bayesian Model and Comparison to Classical Models

Understanding the fundamental processes underlying synaptic function rests upon an accurate description of critical parameters including probability of release, the strength of presynaptic neurotransmitter release, the postsynaptic response, and the relationship between the postsynaptic response and experimental measurements. Classical statistical models of synaptic release generally focus upon changes observed in ensemble histograms of synaptic responses. In such models, the statistical distribution underlying synaptic responses is assumed from the model. The uncertainty surrounding this assumption is often difficult to assess. In addition, the numbers and nature of statistical parameters are often incorporated into the model.

We developed Bayesian site analysis in order to directly represent synaptic structure in terms of measurable parameters and unobserved variables, which describe individual synaptic sites and experimental characteristics (Turner et al, 1997; West, 1997). Bayesian analysis is based upon entire data sets rather than binned histograms or smoothed data set distributions. Our approach performs a global optimization procedure to estimate (and provide bounds for uncertainty) both for the number of synaptic sites and for the parameters of each site (Cao and West, 1997; Escobar and West, 1995; Turner and West, 1993; West, 1997; West and Cao, 1993; West and Turner, 1994). Critical assumptions in our approach are that individual synaptic sites retain characteristic amplitudes, since sites are ordered during the analysis according to magnitude, and that the underlying ambient noise can be characterized by a smooth Gaussian function that is independent of synaptic signals. A similar type of unrestrained analysis, with a different approach and optimization technique, has also been suggested by Uteshev et al (2000).

By comparison with MLE and other pure likelihood-based approaches, Bayesian inference aims to completely explore the regions of parameter spaces supported and indicated by the data, and formally summarize relevant parameter ranges and sets using summary probability distributions rather than just point estimates and error bars. Current standard techniques involve repeated simulation of parameter values jointly from such posterior distributions. Notice that this subsumes and formally manages problems of local maxima in likelihood surfaces, and other such technical problems that often invalidate MLE approaches; the Bayesian strategy is to more thoroughly map

out the parameter surfaces via informed navigation of the surface in regions identified as relevant probabilistically. An initial statistical outline appears in (West, 1997), though our work has since developed to address more fully estimation of intrinsic variation over time (Turner et al, 1997).

Bayesian site analysis begins with initial assumptions about the data, and then performs a multi-dimensional optimization using the likelihood function. Because the resulting analytic function is complex and is without analytic solution, Monte Carlo simulation methods must be used to estimate the solution, termed the predictive probability distribution. The strengths of the Bayesian approach are that it has much stronger theoretical justification that a global solution is obtained than with the MLE approach. In addition, error bounds can be provided on all of the parameters estimated. These error bounds reflect the certainty surrounding both the parameters and the overall solution.

2.2 Bayesian Site Analysis

Bayesian site analysis directly represents individual synaptic site characteristics in terms of amplitude, μ , and release probability, π , parameters, together with an unknown degree of intrinsic, intrasite, ‘quantal’, variation over time, τ . Such quantal variation can be divided into two components, that due to intersite heterogeneity, which is directly analyzed, and that due to intrasite variation over time, which is estimated from the analysis. Analysis is via Bayesian inference, providing point and interval estimates of all unknown parameters, together with full probabilistic descriptions of uncertainty (Turner and West, 1993; West and Turner, 1994). The model represents the interaction of a number of sites, with each synaptic site possessing its representative set of parameters. Importantly, the model does not assume equality of either probability or amplitude between sites, and includes offset of the failures component, an estimation of intrinsic standard deviation and a variable degree of adherence to the initial data set noise standard deviation. Hence the Bayesian format uses a highly unconstrained model.

The statistical model requires initial specification to define the unknown parameters. Firstly, an arbitrary value of k is chosen as an upper bound on the number of sites, depending on the raw data histogram. If there are fewer than the specified k sites, then each of the extra sites may simply assume a zero value. A marginal probability is generated which indicates how likely k numbers of sites will fully explain the data set, termed $P(k|D)$ or the probability of k sites underlying the data set, D . This marginal probability allows inference on the true number of active sites, since both too few and too many sites are clearly identified by low values for $P(k|D)$. Secondly, the amplitudes of individual active sites, μ_j , are allowed to take either equal or different values, allowing for test of equality between subsets of the synaptic sites which take on common values. Thirdly, the release probabilities for the active sites, π_j , are allowed to take on any value between zero and unity. As with amplitudes, the site analysis allows for exact common values among the π_j , to assess for equivalence of this parameter within the subsets of sites. Further prior specification is required to describe the initial uncertainty about the noise variance, for with the measured noise variance of the recorded data may be used to provide initial estimates and ranges.

Statistical site analysis uses methods of stochastic simulation, namely Monte Carlo iterations, to successively generate sampled values of all model parameters, including values for each of the μ_j and π_j pairs, together with estimates of quantal variation and signal noise. These sampled parameter vectors from the data-based Bayesian posterior probability distribution collectively summarize the information in the data and corresponding uncertainties, generating the predictive probability

density function and confidence ranges for all parameters.

We define this approach further below with specific details of both simulations and experimental data sets. Statistical inference is inherent to the Bayesian approach and all parameters are given with appropriate error ranges, including the number of synaptic release sites. Detailed discussion of the model and its theoretical foundations may be found in the references (Cao and West, 1997; Turner et al, 1997; West, 1997).

2.3 Application of the Bayesian Site Model to Simulated Data

Simulations were performed to accurately guide the definition of the prior parameters under particular conditions. The signal to noise ratio and location and number of sites were variously selected but compared to the actual data after all of the analysis. The simulated data sets were in general similar to those expected from companion physiological studies, except that the underlying distribution from which the random sample was drawn is well defined for the simulation ensembles. Once these data sets were generated the Bayesian analysis was performed in a manner identical to that used for the physiological data sets. Once optimized, the parameter values (and error ranges) were compared between the results of the analysis and that of the original distribution. If the original distribution parameter values were bracketed by the 95% error ranges of the analyzed results, that was counted as a correct parameter value, and an error if the true value was outside of the predicted range. The errors in number of sites and parameters were accumulated and calculated as a percentage of the total number of ensembles. This definition of error used the full error estimates available from the Bayesian estimation.

Random ensembles from known distributions were analyzed, composed of 1-5 sites, using varying signal to noise and ensemble size values. The raw histograms varied extensively depending on the random sample obtained from the underlying distribution and thus histogram appearance may be very misleading (Walmsley, 1995), so the underlying data were used for the analysis rather than a histogram. Numerous simulations were performed using random ensembles drawn from distributions with a variety of site numbers, relative amplitudes (both equal and unequal), signal to noise values, quantal standard deviations and sample sizes to estimate the errors associated with the Bayesian analysis. Each set of simulations included 100-200 separate simulated ensembles, each with $n=200$ to $n=750$ data points, which underwent analysis and comparison to the underlying distribution in the same way as the experimental datasets. Errors in the analysis were defined in two separate ways:

- (1) the relative error of the standard deviation normalized to the value of the parameter (analyzed as $SD/parameter\ value$), similar to that suggested by Stricker et al (1996a);
- (2) the percentage of errors per site in which the number of sites, underlying site amplitude or probability did not fall within the 95% confidence interval of the analysis of the data set.

This latter definition shows a more usable definition of error, so that one can estimate how often the analysis will be incorrect taking into account both the limited sample size of the experiment (including sampling errors) and the errors of the analysis. The 95% confidence interval around the median was chosen because the distributions were often asymmetric about the median.

The number of sites (k) derived from the Bayesian analysis reflects partly the prior values, particularly the adherence of the noise in the analysis to the estimated noise standard deviation (σ_{noise})

value. For weak adherence with only a few degrees of freedom ($df = 10$) the noise value itself can fluctuate away from the provisional value, leading to fewer sites than in the distribution. Therefore the adherence of the Bayesian analysis to the noise was usually set to a tighter value ($df = n - 1$), reflecting a strong faith in the preliminary noise estimate from the experiment. The number of sites with the noise $df = n - 1$ evaluated to be either the correct number or larger, in effect to partially overestimate the number of sites to ensure that the site parameters showed the least error. Thus, with the noise set at this level in the analysis the number of sites was highly likely to be at least as large as the number in the simulations. The correct number of sites from the underlying (known) distributions was compared to the predicted number with a cumulative probability of at least 0.5 ($P[k|D] > 0.5$). The summary results of some of the extensive simulations are further defined in Turner et al (1997). Results show that with irregular sites a much larger sample size is needed to resolve the differences between sites. The overlap between sites also increases the difficulty in resolving correctly the actual number of sites.

2.4 Application of the Bayesian Model to Recorded Data

Physiological data sets were acquired as described in detail in Turner et al (1997), using minimal stimulation to assure that only a small, stable number of synaptic sites were activated. The postsynaptic responses were recorded as EPSCs using low-noise whole cell patch electrodes, and the ensembles were carefully assessed for stationarity. Figure 2 shows a physiological average for a minimal stimulation EPSC ensemble (U2N1), recorded from a CA1 pyramidal neuron under highly controlled conditions, using minimal stimulation just above threshold for the presence of a response with averaging. The peak average was 8.87 ± 7.91 pA ($n=244$) and $\sigma_{noise} = 1.22$ pA. Figure 2B shows that the noise (dark circles) and signal (light circles) peak data showed neither a trend nor an irregular pattern over the course of the ensemble. The dataset was then analyzed using noise $df=243$, with a resulting probability favoring a single site [$P(1|D) = 0.93$], and less than 0.07 likelihood that an additional site was needed. The parameters for the single site (Fig. 2D) showed $\mu = 10.2$ pA and $\pi = 0.87$ and a median intrinsic variance of 40%, with the variation around these parameters illustrated as error bars in Fig. 2D. The raw histogram and calculated Bayesian predictive probability density function (pdf) are shown in Figure 2C along with error limits on the pdf as dashed lines. Note the excellent fit of the raw histogram and the pdf. However, the raw histogram was not in any way used in the analysis and is shown only for comparison purposes to the predictive pdf, which was calculated directly from the data. For both of these physiological examples 20,000 Monte Carlo draws were used in the estimation of the parameters.

Figure 3 shows another EPSC ensemble (U4N1), again obtained from the CA1 region during a whole cell patch clamp recording from a CA1 pyramidal neuron (Turner et al, 1997). Figure 3A illustrates an averaged EPSC trace ($n=386$) and Figure 3B shows a trend analysis, with no net changes over the ensemble period. Figure 3C shows the comparison of the raw data histogram and the predictive pdf (smooth line). Figure 3D shows the two sites which were extracted using the Bayesian analysis: $P(1|D) = 0.01$, $P(2|D) = 0.99$ and $P(3|D) = 0$, with parameters $\mu_1 = 7.63$ pA ($\pi_1 = 0.11$), $\mu_2 = 19.0$ pA ($\pi_2 = 0.32$) and intrinsic variance of 22%. Note that the two sites do not overlap and both the amplitudes and probabilities are different, suggesting this ensemble to be a complex response, since neither the simple binomial nor the compound binomial would be a suitable fit. Many further simple and more complex responses are shown in Turner et al (1997), along with the effects of paired-pulse plasticity on site parameters.

The conclusions from the analysis of these data focus on the considerable between site heterogeneity as well as the moderate level of within site (intrinsic or quantal) variance observed. Less than half of the previously identified and analyzed data sets could be fit by a traditional statistical model, indicating that most responses recorded in the CA1 pyramidal neuron showed unequal amplitudes and probabilities, requiring the more complex model for an adequate fit. Informal assessment of convergence was performed as discussed in Section 2. These summaries reflect the inherent uncertainties in estimating the number of active sites. Additionally, it was not possible in the analysis to indicate the degree of ambiguity in the physiological ‘solutions’, since there were always alternative explanations, which could also explain the data. This form of data analysis will clearly require an alternative approach for unambiguous interpretation. The optimal correlative approach would be to identify each post-synaptic site unambiguously (such as with a marker visible with a confocal microscope system) at the time of activation. This method of optical identification could be used to verify which sites were active, and on which trial, to clarify the physiological recorded responses.

Figures 2 and 3 Here

3 DISCUSSION

3.1 Comparison of Simulations and Physiological Data Sets

The simulation data sets were purposefully designed to emulate many of the physiological conditions observed, particularly the level of noise, the signal to noise ratio, the level of probability of release and the amplitudes. The variations on signal to noise and sample size also were solidly established to be close to the physiological values. For most of these realistic parameter values the Bayesian Site analysis performed well, though with low signal to noise (less than 2:1), small sample sizes (particularly less than 300) and irregular sites the extraction of parameters was more difficult. The number of parameter errors allows a quantitative estimate of the likelihood of a single random sample leading to the definition of the underlying distribution from which the sample was drawn. This likelihood includes both the effects of a small sample size (particularly when less than 500-750) and the randomness of the subsequent draw and the ability of the analysis technique to resolve the underlying distribution. Since parameter errors have not been estimated in this practical context previously for neurophysiological applications it is difficult to compare this technique to other currently available statistical analysis formats, particularly the bootstrap technique for the MLE algorithm and the compound binomial approach of Stricker et al (1996a). Further simulations, particularly with more sites, may help define the limits of the Bayesian site analysis. Currently the analysis will work with up to 8-10 synaptic release sites, but beyond 4-5 sites practical problems arise with identifying or clustering each individual sites and more site overlap prevents such effective clustering. Thus, analysis of data sets with 1-3 sites may be more robust than those with greater than 4-5 sites and we are particularly cautious regarding a larger of number of sites yet.

Virtually all of the analysis techniques discussed here have limited availability in terms of distribution, and are highly specialized in their application to particular data sets. Korn and Faber’s binomial analysis program is only applicable under strict conditions and has not become widely available (Korn and Faber, 1991). The Bayesian components (Turner and West, 1993; West and Turner, 1994) and site-directed programs (Turner et al, 1997; West, 1997) are likewise not yet sufficiently robust to become generally available, and require the use of several ranges of parameters to estimate the best solution. In particular, the Bayesian site program does not work well when

there are likely multiple sites (greater than 5-6) and there are very few stable data sets currently available which fit the range of limited applicability of this analysis program. Limited usefulness of each type of program has made it difficult to compare efficacy of parameter and distribution detection; it may be better to compare a standard suite of simulation data between labs rather than share such highly specialized programs. However, this comparison is critical and should be performed. This lack of interoperability between various approaches and difficulty in comparison hampers any validation of any particular approach. Likewise, independent methods (such as an optical methods for direct quantal analysis) to evaluate validity will be critical to underscore the reliability of any particular method (Emptage et al, 2003; Oertner, 2002; Oertner et al, 2002).

3.2 Analysis of Components in Contrast to Sites

The analysis of sites is a more robust approach than using a histogram approximation procedure, regardless of the particular model and optimization approach utilized. For example, Korn and Faber (1991) have used binomial models, which are site-directed, thus ensuring that a low probability tail is adequately represented in the analysis. However, Redman and his group have used a component based approach, beginning with a smoothed histogram but fitting peaks in this smoothed representation of the data (Redman, 1990; Stricker et al, 1996a). The components based approach fits histogram peaks rather than the natural sums and products of underlying sites, and thus the peaks may or may not represent sums of site contributions appropriately. The differences between the two approaches represents an ongoing debate as to the usefulness of the two approaches, but clearly the histograms are sums of individual synaptic site contributions, with their own respective amplitudes and probabilities. Whenever there is more than one site, the statistical prediction is that the appropriate sums of the sites (with the probability being a product of the individual values) should be adequately represented, even though on an individual draw the histogram may not be accurate or representative for the distribution (Walmsley, 1995). To study these different approaches we have evolved through both types of analysis.

The Bayesian data analysis has been previously used for analysis of components (peaks in histograms representing sums of synaptic site contributions - West and Cao, 1993; Turner and West, 1993; West and Turner, 1994) and data are presented here showing the enhanced usefulness of a site-directed approach. If all synaptic sites are assumed equal then a binomial (or in an extreme limiting case, a Poisson) distribution may be appropriate, and the components may be decomposed into their underlying sites. For a binomial distribution, this requires estimation of several variables: n , or the number of sites p , or the probability of release at each site; and the amplitude at each site, q . The histogram is normalized by what is felt to represent the unit amplitude (q) and then integers representing the sums of the individual site contributions are used to assess the probability and number of sites. As shown previously, there is often no unique solution, particularly with the most likely situation that there is contaminating noise and where the first peak may be offset by an arbitrary value (Stricker et al, 1996a). Estimation of n may be particularly difficult. The situation is highly complicated by the additional likelihood that probability is different between sites, leading to a more complex compound binomial distribution (but which still requires equal site amplitude). When the ambient noise becomes significant, removal of the noise effect also requires an optimization routine, usually termed deconvolution (Redman, 1990). Because of the multiple inferences with this form of deconvolution procedure, the results are difficult to confirm, particularly the likelihood that the solution is ‘good’ or ‘optimal’ in any way.

We have developed the Bayesian site method of analysis to overcome many of these problems. By starting with the possibility that sites may not be similar in terms of either amplitude or probability and by not assuming a fixed number of sites, a more flexible and global analysis format is possible. Additionally, as shown in Figs. 2 and 3, the analysis gives error ranges for the predictive pdf and in particular the site parameters, as well as the likelihood for how many sites are required. Though the analysis uses the fitted Gaussian for the noise, this is only a starting point, and depending on the degrees of freedom accompanying the noise value, is more or less strictly used in the analysis. The Bayesian also gives a posterior distribution for the noise variance, indicating how strict this adherence was for a particular data set. Likewise, the Bayesian format is general, and development plans include making the prior distribution for the noise much more flexible, particularly by using a sampled noise distribution rather than any form of fitted smooth distribution. However, this format of analysis begins to give some idea as the ‘goodness’ of the analyzed fit to the raw data, though of course does not in any way guarantee that the fit is unique or optimal. However, the Bayesian approach uses all of the data for the analysis, as opposed to the more conventional MLE and compound binomial (Stricker et al, 1996a) procedures, which utilize optimized histogram data. Additionally, the Bayesian approach provides a more rigorous and extensive investigation of the data-supported regions of the parameter space than the more restricted MLE solution algorithm (West and Turner, 1994). The direct Bayesian site format may eventually prove to be very helpful to identify changes directly with plasticity and experience rather than the more ‘blind’ approaches, which have conventionally been used.

3.3 Analysis of Physiological Data Sets

The results of the two data sets shown here and the larger group of data sets from Turner et al (1997) explicitly highlight the considerable heterogeneity of synaptic site parameters in hippocampal neurons during paired-pulse plasticity (Bolshakov and Siegelbaum, 1995; Hanse and Gustafsson, 2001). The heterogeneity noted between sites (62% of the mean site amplitude) of non-NMDA responses may be attributable to differences in presynaptic neurotransmitter release, postsynaptic receptor density and electrotonic filtering, but is clearly larger than the identified range of intra-site or intrinsic variability (36% of the mean site amplitude after correction), defining the relative importance of significant alterations in presynaptic release over time or ligand-receptor mismatch (Faber et al 1992). One strength of the Bayesian site analysis is to resolve these two separate aspects of quantal variance, that due to either differences at a single site over time (here termed intrinsic variance) or that due to differences between site parameters at a single occurrence. The resolution achieved by this analysis suggests that intersite heterogeneity is clearly the rule in CA1 pyramidal neurons, for a variety of expected and important reasons, whereas intrinsic variability at a single site over time is also important and should be explicitly considered. Together these two sources of variability will lead to difficulty with histogram based approaches, since both tend to merge ‘peaky’ histograms into much smoother and less poorly defined histograms (Turner et al, 1997). This effect of smoothing can easily lead to the two-component histograms defined by earlier studies, in which only a ‘failure’ and a ‘success’ component can be identified. This result confirms that histogram-based approaches may also mislead interpretations regarding underlying sites due to this smoothing effect and random histogram variations from sampling effects (Walmsley, 1995).

A small subset of histograms showed inflections suggesting approximately equal spacing of synaptic site contributions, but the rule (as observed previously in most reports) is that histogram appearance usually was not ‘peaky’ and histograms may be unreliable in pointing to the underlying statistical distribution (Walmsley, 1995). Using a paired-pulse stimulation protocol the histogram distributions clearly changed in many ensembles, but there were also data sets, which did not

exhibit a significant paired-pulse potentiation but rather showed depression (Turner et al, 1997). Paired-pulse plasticity appeared to arise primarily from changes in probability of release, as described by many other authors (Redman, 1990). However, the degree of changes in probability appeared to vary between sites, suggesting that past history of activation and/or other factors may be important in the underlying process leading to the short-term changes. The degree and direction of paired-pulse plasticity response did correlate with the initial probability of release (for the critical synaptic site underlying most of the change) suggesting that enhanced plasticity may be present if the response starts from a lower initial probability. However, other factors may also be important in the determination of the direction and degree of short-term plasticity, since different sites appear to often respond in varying ways to the same conditioning impulse. Thus, the presence of a response on the conditioning pulse may lead to enhanced plasticity compared to responses which showed a failure on the first response. In contrast, Debanne et al suggest that an initial response may lead to depression on the test pulse rather than potentiation (Debanne et al, 1999), possibly due to neurotransmitter depletion.

We have also previously evaluated the hypothesis that different release sites over the dendritic surface lead to amplitude and waveform shape differences due to electrotonic filtering of events (Turner, 1984; Williams and Stuart, 2003). EPSCs from various data sets were found to contain events with different rise-times, which clearly reflected differences in electrotonic filtering. This electrotonic filtering is likely another source of postsynaptic heterogeneity. These data also leads us to suggest that minimal stimulation may lead to activation of only a few release sites onto individual CA1 cells, often located at different electrotonic locations, and that each site may possess different release probabilities and synaptic site amplitudes. Such variation has also been suggested by anatomical studies showing large variation in synaptic shape and presynaptic densities (Harris and Sultan, 1995). Thus, there is evidence for variable amplitudes of synaptic contributions (as recorded at the soma) due to dendritic spatial filtering, and possibly also due to intrinsic differences between synaptic sites located at similar spatial locations. Additionally, the data presented above indicates that probability of release may vary extensively between synapses, and there does not appear to be a strong rationale for assuming that probability is uniform among any mammalian CNS synapses (Redman, 1990). These findings argue for significant intersite heterogeneity due to spatial and perhaps inherent site differences as well as likely a certain contribution of intrinsic ‘quantal’ variability due to changes in release over time. The intrinsic variability can arise from such sources as variable neurotransmitter packaging in vesicles and variable ligand-receptor binding in the synaptic cleft (Faber et al, 1992). The heterogeneity noted at both intrasite and intersite levels of analysis suggests strongly that changes with plasticity may also vary considerably between synapses, possibly as a function of history of use of each synapse. The important role of intrinsic variance is similar to that described by Bolshakov and Siegelbaum (1995; approximately 20-35% in their analyzed data sets) but more than that suggested by Stricker et al (1996a,b; less than 10% in the analyzable data sets).

3.4 Conclusions and Future Perspectives

Depending on the suspected number of synaptic release sites there are several applicable statistical models. Because of suspected significant heterogeneity in the hippocampal CA1 neurons, as discussed above, we have applied the Bayesian analysis because it is the most general and unconstrained in terms of site parameter freedom. As described in Turner et al (1997) the Bayesian complex framework was required for interpretation of the ensemble in a little over half of the datasets, and so clearly allows extension of the number of datasets which can be analyzed. How-

ever, limitations include the potential loss of identifiability of data sets with larger numbers of synaptic sites (possibly greater than 5-6 sites), the current lack of realistic noise modeling for complex noise and the lack of any on-line feedback during recordings to show the number of sites activated on a trial by trial basis. Thus, further developments anticipated include simulations to model additional numbers of sites, improved noise characteristics, better characterization of data set validity and stationarity, addition of a dynamic updating to allow some on-line analysis during an experiment and placing the data stream from individual synaptic sites into a time series format of analysis. The latter will allow historical effects to be included, such as potentiation with time, in a dynamic format, and may be more predictive of future trends. Addition of dendritic models to include conduction from dendritic sites to the soma may also enhance the link between synaptic inputs and cell firing capabilities and cell behavior in general (Williams and Stuart, 2003).

In addition to the scientific and statistical issues arising in connection with investigations of intrinsic variability, several other areas are of current interest and are under initial investigation. As presented, our models make no explicit allowance for outliers in signal recordings, nor for systematic features such as arise when synaptic tissue exhibits inhibitory as well as excitatory responses to stimuli. An example of the latter is discussed in West and Turner (1994), for example. Perhaps the most needed generalization, currently, is to relax the assumption of normally distributed experimental noise. This is currently in development, using noise models based on work in Bayesian density estimation following Escobar and West (1995).

One beginning area of research involves issues of dependence in release characteristics over time and across sites (addressed in a different manner by Uteshev et al, 2000). On the first, synaptic responses in general may be viewed as a stream where, at each site, the response may depend on previous occurrences. Current analyses assume independence. The very nature of the paired-pulse potentiation experiments, however, explicitly recognize dependence effects, both in terms of response/no response and in terms of response levels, and of the importance of time elapsed between successive stimuli generating the responses. Statistical developments of time series issues are of interest here to explore and, ultimately, model time variation in patterns of synaptic transmission. Interest in transmission dependencies also extends across sites, i.e., to investigations of ‘connectivities’ between sites that may have physiological interpretation and importance, though serious inroads into study of this issue is some way in the future.

It will be apparent that the technical aspects of simulation based analysis in the novel mixture models here bear further study, from a practical perspective. It is certainly desirable to investigate variations on the basic Gibbs sampling schemes that enhance and improve convergence characteristics. It is possible that hybrid schemes based, in part, on the use of auxiliary variables (e.g., Besag and Green, 1993) will prove useful here; the need for improved computational tools will be more strongly felt in more complex models incorporating non-normal noise components. That said, the current approach described and illustrated here provides a useful basis for exploring and assessing synaptic data in attempts to more directly evaluating the structure and stochastic characteristics of synaptic transmission as it is currently viewed in the neurophysiological community. Our current work on the applications side is focused on refining this basic framework and in exploring ranges of synaptic data sets to isolate and summarize the diversity of observed response mechanisms in mammalian central nervous systems.

Clearly, physiological methods are needed to clarify the ambiguity associated with the various parameters at even a single synapse. Optical quantal analysis allows the investigation of single synaptic sites, and correlation with physiology and statistical measures (Oertner, 2002). As these

newer physiological approaches narrow the degree of possible solutions, then the statistical methods can be more sharply aligned and more realistic and specific.

ACKNOWLEDGMENTS

Supported by the National Science Foundation (DAT, MW), and Department of Veterans Affairs Merit Review (DAT).

APPENDIX: MATHEMATICAL DERIVATION OF THE MODEL

Ensembles of maximum synaptic signal levels (specifically excitatory post-synaptic potentials [EPSPs], excitatory postsynaptic currents [EPSCs], inhibitory postsynaptic currents [IPSCs], etc.) obtained in response to stimulation are assumed to be independently generated according to an underlying stochastic, synaptic site-specific mechanism described as follows. At the identified synaptic connection, the experiment isolates either both the pre- and post-synaptic neurons for complete control of all interactions between the two cells (as shown diagrammatically in Figure 1), or a region on the axon branch of a pre-synaptic neuron leading to activation of only a small number of postsynaptic sites. Stimulation applied to the isolated presynaptic element leads to chemical neurotransmitter release at a random number of available synaptic transmission sites, similar to those shown in Fig. 1. If release occurs, then each individual site contributes a small electrical signal, which then sums across sites to induce the overall resulting electrical signal in the postsynaptic neuron, which is measured physiologically, though usually in concert with interfering noise. Each of the $s > 0$ release sites transmits to a site-specific maximum level independently of all other sites. There is no identification of individual sites (e.g., such as might arise were we able to physically mark sites in the tissue and identify the location of individual transmissions) so we arbitrarily label the sites $1, \dots, s$. Site by site we assume that, on any trial,

- sites ‘release’ independently, with individual, site-specific chances of release occurring on any occasion;
- a transmitting site produces a site-specific, fixed packet or ‘quantum’ of neurotransmitter leading to a characteristic amplitude response at the synaptic site;
- recorded maximum synaptic signal levels represent the sums of signals induced by the various sites randomly transmitting, with additive synaptic and experimental noise.

In any experiment we assume individual transmitter release probabilities and transmission levels to be fixed and with only slight perturbations over time, but no net trend. Changes in these characteristics over time due to forced experimental and environmental changes is of later interest, and evaluating such changes is one of the guiding motivations in developing these models. Symbolically, we represent recorded signal levels as $y = \{y_1, \dots, y_n\}$ where y_i is the level on trial i ; the sample size n is typically a few hundreds. Given s sites assumed, site j transmits on any trial with probability π_j , independently of other sites and independently across trials. The level of transmission on occurrence is the site-specific quantity μ_j . Physically, signal levels are measured via either sharp intracellular probes or low-noise patch electrodes as electrical signals induced on cell bodies, and responses are in units of either current (pA) or millivolts (mV). By convention, we change the reading to predominantly positive values so that the μ_j must be non-negative values. The recordings are subject to background synaptic and experimental noise, including both electronic distortions in recording the cellular potential and errors induced in computing the estimated maximum potential levels. These noise sources combine to produce additive errors, assumed to be approximately normally distributed independently across trials, as discussed in the introduction; write v for the variance of these normal errors. Concomitant noise measurements are available

to assess these experimental errors and to provide relevant prior information relevant to v ; blank signals measured without experimental stimulus provide these measurements. It is stressed that the noise affecting signal recordings has the same distribution, of variance v , as the raw noise recordings obtained without a signal, there being no scientific or experimental reason to assume otherwise. We also model a systematic bias in the signal recordings; although the associated noise measures are typically close to zero mean, the signal recordings are often subject to a small but non-negligible baseline offset induced by the experimental procedures; call this shift m . We then have the following basic signal model.

For $i = 1, \dots, n$, the signal observations are conditionally independent,

$$y_i \sim N(y_i|\theta_i, v) \quad \text{with} \quad \theta_i = m + \sum_{j=1}^s z_{ij}\mu_j$$

and where the z_{ij} are latent, unobserved, site transmission indicators,

$$z_{ij} = \begin{cases} 1, & \text{if site } j \text{ transmits on trial } i; \\ 0, & \text{otherwise.} \end{cases}$$

Under the model assumptions, these indicators are conditionally independent Bernoulli quantities with $P(z_{ij} = 1) = \pi_j$ for each site j and all trials i .

Thus signal data are drawn from a discrete mixture of (at most) $k = 2^s$ normal components of common variance, induced by averaging normals determined by all possible combinations of the z_{ij} and weighted by the corresponding chances of those combinations. On any trial, the normal component selected is determined by the column s -vector $z_i = (z_{i1}, \dots, z_{is})'$ realized on that trial by individual sites releasing or not; the selected component has mean $\theta_i = m + z_i'\mu$ where $\mu' = (\mu_1, \dots, \mu_s)$, and is selected with chance $\prod_{j=1}^s \pi_j^{z_{ij}} (1 - \pi_j)^{1-z_{ij}}$; specifically, and conditional on all model assumptions and parameters, the y_i are conditionally independently drawn from the distribution with density

$$p(y_i|\mu, \pi, m, v) = \sum_{z_i} p(z_i|\pi) p(y_i|\mu, z_i, m, v) = \sum_{z_i} \left\{ \prod_{j=1}^s \pi_j^{z_{ij}} (1 - \pi_j)^{1-z_{ij}} \right\} N(y_i|m + z_i'\mu, v), \quad (1)$$

where the s -vector z_i ranges over all 2^s possible values. If there are common values among the site levels μ_j , then the mixture distribution function will be equivalent to one with fewer than 2^s components. Also, if either $\mu_j = 0$ or $\pi_j = 0$ then site j disappears from the model and the distribution reduces to a mixture of at most 2^{s-1} components. Additional cases of special interest include:

- (a) So-called compound binomial models, in which the site response levels are equal, all at a basic 'quantum' level μ_0 , but release probabilities differ. The aggregate response levels then run between m and $m + s\mu_0$ and the mixture of 2^s normal components effectively reduces to a mixture of just $s + 1$, with associated weights determined by the discrete compound binomial resulting from the distinct chances π_1, \dots, π_s across sites. These kinds of special cases have received considerable attention in the literature (Stricker et al, 1996a).
- (b) Precise quantal-binomial models, in which the site levels are equal and the release probabilities are constant too, $\pi_j = \pi_0$. Now the mixture reduces to one with distinct normal

means $m + j\mu_0$ ($j = 0, \dots, s$), and corresponding binomial weights $\binom{s}{j}\pi_0^j(1-\pi_0)^{s-j}$ (Martin 1966; Redman 1990).

Our previous work, following others, modelled synaptic data in the framework of ‘standard’ mixture models for density estimation, from a particular Bayesian viewpoint; see West and Cao (1993), West and Turner (1994), for example, and Escobar and West (1995) for statistical background. In these models, the component means and weights are essentially unrestricted, rather than being modelled directly as functions of the underlying synaptic parameters μ and π . A significant drawback is that it is then, in general, very difficult to translate posterior inferences about unrestricted normal mixture model parameters to the underlying μ and π parameters of scientific interest, especially in the context of uncertainty about s . One simple example in West and Turner (1994) shows how this can be done; that is a rare example in which the data appear to be consistent with the full quantal hypothesis, so that the convoluted process of backtracking from the mixture model to the underlying site-specific parameters is accessible. However, we have encountered very few data sets in which this is the case, and this inversion process is generally difficult. Hence the direct approach, inferring the neural parameters directly, has been developed.

1 Prior Distributions for Synaptic Parameters

Model completion requires specification of classes of prior distributions for the determining parameters μ , π , m and v for any given s . Assessment of reasonable values of s , as well as values of the μ and π quantities for any given s , is part of the statistical inference problem. From a technical viewpoint, uncertainty about s can be formally included in the prior distribution, so as to provide posterior assessment of plausible values. The models below allow this. From the viewpoint of scientific interpretation, however, inference about the site parameters are best made conditional on posited values of s , and then issues of sensitivity to the number of sites explored.

Begin by conditioning on a supposed number of sites s . Current implementation assumes priors for all model parameters which, though rather intricately structured in certain dimensions to represent key qualitative features of the scientific context, are nevertheless inherently ‘uniform’ in appropriate ways, providing ‘reference’ initial distributions. Note that other classes of priors may be used, and some obvious variations are mentioned below; however, those used here are believed to be appropriately vague or uninformative in order that resulting posterior distributions provide benchmark or reference inferences that may be directly compared with supposedly objective non-Bayesian approaches (Redman, 1990; comparative analyses will be reported elsewhere in a further article).

In addressing uncertainty about s alone, we make the observation that models with fewer than s sites are implicitly nested within a model having s sites. To see this, constrain some $r > 0$ (arbitrarily labelled) site levels μ_j to be zero; then (1) reduces to precisely the same form based on $s - r$ sites, whatever the value of the π_j corresponding to the zeroed μ_j . Zeroing some μ_j simply confounds those sites with the noise component in the mixture—a site transmitting a zero level, with any probability, simply cannot be identified. Hence assessing whether or not one or more of the site levels are zero provides assessment of whether or not the data support fewer sites than assumed. This provides a natural approach to inference on the number of active sites, assuming that the model value s is chosen as an upper bound.

Note that a similar conclusion arises by considering $\pi_j = 0$ for some indices j ; i.e., an inactive

site may have a zero release probability rather than (or as well as) a zero release level. However, the zeroing of a site probability induces a degeneracy in structure of the model and obviates its use as a technical device for inducing a nesting of models with fewer than s sites in the overall model. Hence our models define inactive sites through zeros among the μ_j , restricting the π_j to non-zero values, however small. Thus $\mu_j = 0$ for one or more sites j is the one and only way that the number of active sites may be smaller than the specific s .

1.1 General structure

For a specified s , analyses reported below are based on priors with the following structure. Quantities μ , π , (m, v) are mutually independent. We will describe classes of marginal priors for μ and π separately; each will involve certain hyper-parameters that will themselves be subject to uncertainty described through hyper-priors, in a typical hierarchical modelling framework, and these hyper-parameters will also be assumed mutually independent. To anticipate development below, hyper-parameters denoted q and a are associated with the prior for μ , a single quantity b determines the prior for π , and the joint prior is then of the form

$$p(\mu, \pi, m, v, q, a, b) = p(\mu, \pi, m, v | q, a, b) p(q, a, b) = p(\mu | q, a) p(\pi | b) p(m, v) p(q) p(a) p(b). \quad (2)$$

The component densities here are now described in detail. Two comments on notation: first, conditioning statements in density functions include only those quantities that are required to determine the density, implicitly indicating conditional independence of omitted quantities; second, for any vector of h quantities $x = (x_1, \dots, x_h)$, for any $j \leq h$ the notation x_{-j} represents the vector x with x_j removed, i.e., $x_{-j} = x - \{x_j\} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_h)$.

1.2 Priors for μ and associated hyper-parameters q, a

We develop a general classes of priors for μ , and comment on various special cases. The class has the following features:

- (a) a component baseline uniform distribution for each μ_j over a prespecified range $(0, u)$, with a specified upper bound u ;
- (b) components introducing positive prior probabilities at zero for each of the μ_j in order to permit assessment of hypotheses that fewer than the chosen (upper bound) s are actually non-zero, and hence to infer values of the number of active sites;
- (c) components permitting exact common values among the elements of μ to allow for the various special cases of quantal transmission, and specifically the questions of whether or not pairs or subsets of sites share essentially the same ‘quantal’ transmission level.

These are developed as follows. Let $F(\cdot)$ be distribution on $(0, u)$, having density $f(\cdot)$, for some specified upper bound u . Write $\delta_0(x)$ for the Dirac delta function at $x = 0$, and $U(\cdot | a, b)$ for the continuous uniform density over (a, b) . Then suppose the μ_j are *conditionally independently* drawn from the model

$$(\mu_j | F, q) \sim q \delta_0(\mu_j) + (1 - q) f(\mu_j)$$

for some probability q . Under this prior, the number h of non-zero values among the μ_j —the number of active sites—is binomial, $(h | s, q) \sim Bn(s, 1 - q)$, with mean $s(1 - q)$, independently of F . If F were uniform, say $U(\cdot | 0, u)$, this prior neatly embodies the first two desirable features (a) and (b) described above. Note the two distinct cases:

- Setting $q = 0$ implies that $h = s$ is the assumed number of active sites, so then inference proceeds conditional on $\mu_j > 0$ for each j .
- Otherwise, restricting to $q > 0$ allows for assessment of the number of active sites, subject to the specified upper bound s .

In practice we will assign a hyperprior to q in the case $q > 0$. The class of beta distributions is conditionally conjugate, and the uniform prior suggests itself as a reference, $q \sim U(q|0, 1)$. One immediate, and nice, consequence of a uniform prior is that the resulting prior for the number of non-zero values among the μ_j has (averaging the binomial with respect to q) a discrete uniform prior over $0, 1, \dots, s$. This is a suitably vague and unbiased initial viewpoint with respect to the number of active sites. Other beta priors may be explored, of course. One specific choice we use in current work is $(q|s) \sim Be(s - 1, 1)$; note the explicit recognition of dependence on the specified values of s . The reasoning behind this choice is as follows. First, we are currently focused on experiments designed to isolate rather small numbers of sites, down to just a few, say 1 – 4 from the viewpoint of scientific intent and expectation. So s values up to 7 or 8 may be explored, but lower values are typical. Whatever value of s is chosen, h is expected to be in the low integers, so guiding choice of the prior for q to induce a prior for h favoring smaller values. Consider values of s in the relevant range $3 \leq s \leq 8$, or so. Then, integrating $p(h|s, q)$ with respect to the specific prior $(q|s) \sim Be(s - 1, 1)$ we obtain a distribution $p(h|s)$ that is almost completely insensitive to s , having a diffuse and decreasing form as h increases, and with $E(h|s) = 1$ for any such s . Thus, this specific beta prior for q has the attractive feature of consistency with a scientifically plausible prior $p(h|s) \approx p(h)$, incorporating the scientific view of likely small numbers of active sites, and almost independently of the upper bound s specified.

This structure provides a baseline uniform prior for release levels, together with the option for allowing a smaller number of sites than the s specified. So far, however, there is no explicit recognition of the special status, in the scientific area, of quantal hypotheses as represented through common values among the μ_j . If $F(\cdot)$ is a continuous distribution, then the prior implies that the non-zero μ_j are distinct. Though this might allow arbitrarily close μ_j values, it is desirable to have the opportunity to directly assess questions about common values, and perhaps subgroups of common values, in terms of posterior probabilities. There is also a significant technical reason, discussed below in connection with issues of parameter identification, that calls for a prior that gives positive probability to exact equality of collections of the non-zero μ_j . We therefore extend the prior structure so far discussed to provide this. We do this using a standard Dirichlet model. Specifically, a Dirichlet process prior for F induces a discrete structure which gives positive prior probability to essentially arbitrary groupings of the set of non-zero μ_j into subsets of common values; this is a general framework that permits varying degrees of partial quantal structure, from the one extreme of completely distinct values to the other of one common value. This structure is most easily appreciated through the resulting set of complete conditional posterior distributions for each of the μ_j given μ_{-j} . These are defined by

$$p(\mu_j | \mu_{-j}, q, a) = q\delta_0(\mu_j) + (1 - q)\{r_j U(\mu_j | 0, u) + (1 - r_j)h_j^{-1} \sum_{i \in N_j} \delta_{\mu_i}(\mu_j)\}, \quad (3)$$

where h_j is the number of non-zero elements of μ_{-j} and N_j is the corresponding set of indices, $N_j = \{i | \mu_i > 0, i = 1, \dots, s; i \neq j\}$, and $r_j = a/(a + h_j)$. The hyper-parameter a is subject to uncertainty and is included in the analysis using existing approaches for inference on precision parameters in Dirichlet models, developed in West (1997) and illustrated in Escobar and West

(1995). As shown there, gamma priors for a , or mixtures of gamma priors, are natural choices, and our application currently uses diffuse gammas models.

In summary, (3) shows explicitly how site level μ_j may be zero, implying an inactive site, or take a new non-zero value, or be equal to one of the non-zero values of other sites. The roles of hyperparameters q and a are evident in this equation. With this structure we have defined prior components $p(\mu|q, a)p(q)p(a)$ of the full joint prior in equation (2).

1.3 Priors for π and associated hyperparameter b

The structure of the prior for π parallels, in part, that of μ in allowing common values. The detailed development for μ can be followed through with the same reasoning about a baseline prior and a structure to induce positive probabilities over subsets of common values. We restrict to positive release probabilities, as discussed earlier, so that the prior structure for the π_j is simpler in this respect. Specifically, we assume the π_j independently drawn from a distribution on $(0, 1)$ that is assigned a Dirichlet process prior. We take Dirichlet precision $b > 0$ and base measure to be $bU(\cdot|0, 1)$. Then, as in the development for μ the full joint prior $p(\pi|b)$ is defined by its conditionals

$$(\pi_j|\pi_{-j}, b) \sim wU(\pi_j|0, 1) + (1 - w)(s - 1)^{-1} \sum_{i=1, i \neq j}^s \delta_{\pi_i}(\pi_j) \quad (4)$$

where $w = b/(b + s - 1)$, for each $j = 1, \dots, s$.

As for the site levels, the induced posteriors will now allow inference on which sites may have common release probabilities, as well as on the precise values of such probabilities.

Non-uniform beta distributions might replace the uniform baseline in application, particularly with expected low levels of release probabilities. Our analyses reported below retain the uniform prior, for illustration and so as to avoid question of overtly biasing towards lower or higher values. Note also that we explicitly exclude the possibility of non-zero release probabilities, though some or all may be very small, and so rule out the use of $\pi_j = 0$ as a device for reducing the number of active sites; that is completely determined by zero values among the μ_j , as discussed earlier.

As with a in the model for the site levels, the hyper-parameter b will be assigned a prior, and again a diffuse gamma prior is natural. With this structure, we have defined the additional prior components $p(\pi|b)p(b)$ of the full joint prior in equation (2).

1.4 Priors for noise moments m and v

In most experiments we anticipate a possible systematic bias m in both noise and signal recordings, induced by the direct measurement of cell membrane potential. This is typically very small relative to induced signal levels, and the raw noise recordings provide data on which to assess this. In addition, the noise measurements inform on background variability v , i.e., they are drawn from the $N(\cdot|m, v)$ distribution. In our analyses, our prior for m and v as input to the signal analysis is simply the posterior from a reference analysis of the noise data alone, i.e., a standard conjugate normal, inverse gamma distribution based on the noise sample mean, variance and sample size. With this structure, we have defined the final component $p(v, m)$ of the full joint prior in equation (2).

2 Posterior Distributions and Computation

Calculation of posterior distributions is feasible, as might be expected, via variants of Gibbs sampling (e.g., Gelfand and Smith 1990, Smith and Roberts 1993). The specific collection and sequence

of conditional posterior distributions used to develop simulation algorithms are briefly summarized here. A key step is to augment the data with the latent site transmission indicators z ; thus the sampling model is expanded as

$$\begin{aligned} p(y, z | \mu, \pi, m, v, q, a, b) &= p(y | \mu, z, m, v) p(z | \pi) \\ &= \left\{ \prod_{i=1}^n N(y_i | m + z'_i \mu, v) \right\} \left\{ \prod_{j=1}^s \prod_{i=1}^n \pi_j^{z_{ij}} (1 - \pi_j)^{1 - z_{ij}} \right\}, \end{aligned}$$

providing various conditional likelihood functions that are neatly factorized into simple components. The full joint posterior density, for all model parameters together with the uncertain indicators z , has the product form

$$p(\mu, \pi, z, m, v, q, a, b | y) \propto p(\mu | q, a) p(q) p(a) p(z | \pi) p(\pi | b) p(b) p(m, v) p(y | \mu, z, m, v) \quad (5)$$

where the component conditional prior terms are as described in the previous section.

3 Conditional Posteriors and MCMC

The posterior (5) yields a tractable set of complete conditional distributions characterizing the joint posterior, and hence leads to implementation of Gibbs sampling based on this structure. Each iteration of this MCMC sampling scheme draws a new set of all parameters and latent variables by sequencing through the conditionals now noted.

- (1) Sampling site levels μ proceeds by sequencing through $j = 1, \dots, s$, at each step generating a new value of μ_j given the latest sampled values of μ_{-j} and all other conditioning quantities. For each j , this involves sampling a posterior that is a simple mixture of several point masses with a truncated normal distribution. Sampling efficiency is improved using variations of so-called configuration sampling for the discrete components, as introduced by MacEachern (1994).
- (2) Sampling site release probabilities π similarly proceeds by sequencing through $j = 1, \dots, s$, at each step generating a new value of π_j given the latest sampled values of π_{-j} and all other conditioning quantities. For each j , this involves sampling a mixture of discrete components with a beta component. Again, configuration sampling improves simulation efficiency.
- (3) Sampling indicators transmission z involves a set of n independent draws from conditional multinomial posteriors for the n individual binary 2^s -vectors z_i , $i = 1, \dots, n$. These simulations are easily performed.
- (4) Sampling the systematic bias quantity m involves a simple normal draw.
- (5) Sampling the noise variance v involves sampling an inverse gamma posterior.
- (6) Sampling the hyper-parameter q is also trivial, simply involving a draw from an appropriate beta distribution.
- (7) Sampling the hyper-parameters a and b follows West (1992), and Escobar and West (1995), and involves a minor ‘augmentation’ of the parameter space, with simple beta and gamma variate generations.

Iterating this process produces sequences of simulated values of the full set of quantities $\phi \stackrel{\text{def}}{=} \{\mu, \pi, z, m, v, q, a, b\}$ which represent realizations of a Markov chain in ϕ space whose stationary distribution is the joint posterior (5) characterized by the conditionals. The precise sequence (1)–(7) of conditional posteriors detailed above is the sequence currently used in simulation; at each

step, new values of the quantities in question are sampled based on current values of all quantities required to determine the conditional in question. Convergence is assured by appeal to rather general results, such as Tierney (1994, especially Theorem 1 and Corollary 2) relevant to the models here; this ensures that successive realizations of $\phi = \{\mu, \pi, z, v, q, a, b\}$ generated by this Gibbs sampling set-up eventually resemble samples from the exact posterior in (5). Irreducibility of the resulting chain is a consequence of the fact that all full conditionals defining the chain are everywhere positive. Practical issues of convergence are addressed by experiments with several short runs from various starting values. Important practical issues of convergence is discussed in the next section.

4 Parameter Identification and Relabelling Transmission Sites

The model structure as described is not identifiable, in the traditional sense of parameter identification especially associated with mixture models. This is obvious in view of the lack of any physical identification of the neural transmission sites we have arbitrarily labelled $j = 1, \dots, s$; for example, π_1 is the release probability for the first site in our labelling order, but this could be any of the actual release sites. As a result, the model is invariant under arbitrary permutations of the labels on sites, and likelihood functions are similarly invariant under permutations of the indices of the sets of parameters (μ_j, π_j) . Also, as the priors for the μ_j and π_j separately are exchangeable, corresponding posteriors are also invariant under labelling permutations.

To impose identification we need a physically meaningful restriction on the parameters. Perhaps the simplest, and most obvious, constraint is to order either site levels or site release chances. We can do this with the site release levels. In posterior inferences, relabel so that site j corresponds to release level μ_j , the j^{th} largest value in μ . As the priors, hence posteriors, include positive probabilities on common values, we break ties by imposing a subsidiary ordering on the release probabilities for (ordered) sites with common release levels. Thus, if μ_3, μ_4 and μ_5 happened to be equal, the sites relabelled 3–5 would be so chosen in order of increasing values of their chances π_j ; were these equal too, as permitted under the priors and hence posteriors here, then the labelling is arbitrary and we have (in this example) three indistinguishable sites: common release levels and common release probabilities. The posterior analysis required to compute corresponding posterior inferences is extremely simple in the context of posterior simulations; each iteration of the Gibbs sampling analysis produces a draw of (μ, π) (among other things). Given a draw, we first create a separate vector μ^* containing the *ordered* values of μ , and a second vector π^* containing the elements of π rearranged to correspond to the ordering of μ to μ^* . If subsets of contiguous values in μ^* are common, the corresponding elements in π^* are then rearranged in increasing order themselves. Through iterations, repeat draws of (μ^*, π^*) are saved, so building up samples from the required posterior distribution that incorporates the identification of sites.

Note that we could alternatively identify the model by a primary ordering on the π_j , rather than the μ_j . In data analyses we routinely explore posterior summaries identified each—i.e., ordered by μ_j and ordered by π_j —as each is only a partial summary of the full posterior. This is relevant particularly in cases when it appears that the data are in conformity, at least partially, with quantal structure.

In connection with this ordering issue, the quantal structure of the priors for μ and π is very relevant, as we alluded to earlier. Consider an application in which the data support common, or very close, values among the μ_j . Suppose that the prior is simply uniform, with no chance that consecutive values of the ordered μ_j are equal. Then imposing the ordering for identification results in a distortion of inferences—values we should judge to be close, or equal, are ‘pushed apart’ by

the imposed ordering. This undesirable effect is ameliorated by the priors we use, allowing exact equality of neighbouring levels. The same applies to the π_j .

Notice that we retain the original, unidentified parametrization for model specification and the simulation analysis operates on the unconstrained posteriors. This has a theoretical benefit in allowing easy and direct application of standard convergence results (Tierney 1994). More practically, it has been our experience that imposing an identifying ordering on parameters directly through the prior distribution hinders convergence of the resulting MCMC on the constrained parameter space. This arises naturally as the ordering induces a highly structured dependence between parameters, relative to the unconstrained (albeit unidentified) parameter space. In addition, one rather intriguing practical payoff from operating the MCMC on the unrestricted space is an aid in assessing convergence of the simulation. This is due to the exact symmetries in posterior distributions resulting from permutation invariance, partly evidenced through the fact that marginal posteriors $p(\mu_j, \pi_j | y)$ are the same for all j . Hence summaries of marginal posteriors for, say, the μ_j should be indicative of the same margins for all j ; the same if true of the π_j .

However, in many applications convergence will be slow. One reason for this is that, in some models and with some observed signal data sets, the posterior will heavily concentrate in widely separate regions of the parameter space, possibly being multi-modal with well separated modes. Such posteriors are notorious in the Markov chain Monte Carlo literature, and the Gibbs sampling routine may get ‘stuck’ in iterations in just one (or typically several) region of high posterior probability, or around a subset of modes. There is certainly need for further algorithmic research to provide faster alternatives to the raw Gibbs sampling developed here (possibly based on some of the methods mentioned in Besag and Green [1993], or Smith and Roberts [1993], for example). However, this is not a problem as far as the ordered site levels, and the correspondingly reordered release probabilities, are concerned, due to the symmetries in the unconstrained posterior. This is because posterior samples from one region of the (μ, π) space may be reflected to other regions by permuting sites indices, while the resulting ordered values remain unchanged. Hence the unconstrained sampling algorithm may converge slowly as far as sampling the posterior for (μ, π) , but the derived sequence for (μ^*, π^*) converges much faster.

Due to the lack of identification, actual output streams from MCMC simulations exhibit switching effects, as parameter draws jump between the modes representing the identification issue. For example, in a study supporting two distinct site release levels with values near 1 and 2 mV, respectively, the simulated series of the unidentified μ_1 (μ_2) will tend to vary around 1 mV (2 mV) for a while, then randomly switch to near 2 mV (1 mV) for a while, and so on. Our experience is that, across several analyses of different data sets, convergence is generally ‘clean’ in the sense that, following what is usually a rapid initial, burn-in period, the output streams remain stable and apparently stationary between points of switching between these posterior modes. Informal diagnostics based on a small number of short repeat runs from different starting values verify summarized analyses from one final, longer run. In addition, the theoretical convergence diagnostic based on the symmetry of the unidentified posterior is used in final analysis.

5 Incorporation of Intrinsic Variability into the Model

Recent attention has focussed on questions of additional variability in synaptic signal outcomes due to so-called *intrinsic variability* in release levels of individual synapses (Turner et al, 1997); we describe this concept here, and define an elaborated class of models incorporating it. This concept gives rise to model modifications in which components of the normal mixture representation have

variances that increase with level, and this leads to considerable complications, both substantive and technical. The technical complications have to do with developing appropriate model extensions, and associated MCMC techniques to analyze the resulting models; some brief development is mentioned here, with examples. The substantive complication is essentially that competing models with and without intrinsic variance components cannot be readily distinguished on the basis of the observed data alone; an observed data configuration might arise from just one or two sites with significant intrinsic variance, or it might arise from a greater number of sites with low or zero intrinsic variance. In such cases, and especially when inferences about site characteristics are heavily dependent on the number of sites and levels of intrinsic variance, we are left reliant on the opinions of expert neurophysiologists to judge between the models. Unfortunately, in its current state, the field is represented by widely varying expert opinions, from the one extreme of complete disregard of the notion of intrinsic variability, to the other of belief in high levels of intrinsic variability as the norm. In some examples we have studied, including the data sets shown here, this issue is relatively benign, as inferences about release levels and probabilities are relatively insensitive to intrinsic variances. In others cases, it is evidently highly relevant. Further collaborative research to refine knowledge of intrinsic variability effects is part of the current frontiers of the field. Here we give a short discussion of the notion and our current approach to modelling intrinsic variability.

Intrinsic variability refers to variation in levels of neurotransmitter release at specific sites. As developed above, site j has a fixed release level μ_j , and, while these levels may differ across sites, they are assumed fixed for the duration of the experiment under the controlled conditions. However, the mechanism of elector-chemical transmission suggests that this may be an over-simplification. A site transmits by releasing a packet of (many) molecules of a chemical transmitter, and these molecules move across the synaptic ‘cleft’ to eventually bind to the postsynaptic receptors. The induced signal response is proportional to the number of molecules binding. So, the assumption of fixed μ_j implies that (a) the number of molecules transmitted is constant across occasions, and (b) all ejected molecules are bound to receptors on the post-synaptic cell. Each of these is questionable, and the issue has given rise to the notion of intrinsic variance, i.e., variability in the site-specific level of release across occasions (Faber et al, 1992). Intrinsic variance may also arise when two sites are close to each other and cannot be separated in any statistical sense. Then, it will appear that one site may be fluctuating. Whatever the actual structure of variability, it is apparent in data sets through mixture components having higher variance at higher release levels.

The basic extension of our models to admit the possibility of intrinsic variability involves remodelling the data y_i as coming from the conditional normal model

$$y_i \sim N(y_i | m + \sum_{j=1}^s z_{ij} \gamma_{ij}, v) \quad (6)$$

where the γ_{ij} are new site-, and case-, specific release levels and v is the noise variance. Retaining independence across sites and trials, we assume that the γ_{ij} are distributed about an underlying expected release level μ_j —the same site-specific level as before, but now representing underlying average levels about which releases vary. Then the form of the distribution of the γ_{ij} about μ_j represents the intrinsic variability in induced responses due to variation in amounts of neurotransmitter released by site j and also to variation in the success rate in moving the transmitter across the synaptic cleft. Various parametric forms might be considered; our preliminary work, to date, is based on exploration of models in which

$$p(\gamma_{ij} | \mu_j, \tau_j) \propto N(\gamma_{ij} | \mu_j, \tau_j^2 \mu_j^2) I(0 < \gamma_{ij} < u) \quad (7)$$

for all trials i and each site j ; here u is the earlier specific upper bound on release levels. The new parameters $\tau_j > 0$ measure intrinsic variability of sites $j = 1, \dots, s$; ignoring the truncation in (7), τ_j is an effective constant (though site-specific) coefficient of variation in release levels about the underlying μ_j . This model has been implemented, extending the prior structure detailed in Section 2.2 to incorporate the full set of site-, and trial-, specific release levels $\{\gamma_{ij}\}$ together with the new parameters τ_1, \dots, τ_s . Note that, as our model allows $\mu_j = 0$ for inactive sites, we use (7) only for $\mu_j > 0$; otherwise, $\mu_j = 0$ implies $\gamma_{ij} = 0$ for each $i = 1, \dots, n$.

We can see the effects of quantal variability by integrating the data density (6) with respect to (7). This is complicated due to the truncation to positive values; as an approximation, for the purposes of illustrating structure here, assume this is not binding, i.e., that μ_j and τ_j are such that the mass of the basic normal distribution in (7) lies well within the interval $(0, u)$. Then (6) and (7) combine and marginalize over γ_{ij} to give

$$y_i \sim N(y_i | m + \sum_{j=1}^s z_{ij} \mu_j, v + \sum_{j=1}^s z_{ij} \tau_j^2 \mu_j^2). \quad (8)$$

The mean here is as in the original formulation, equation (1), with active sites contributing the expected release levels μ_j . But now the variance of y_i is not simply the noise variance v ; it is inflated by adding in factors $\tau_j^2 \mu_j^2$ for each of the active sites. Hence the feature relevant to modelling increased spread of data configurations at higher response levels.

It should be clear that this extended model can be managed computationally with direct extensions of the MCMC algorithms discussed so far. We are now interested in the full posterior $p(\mu, \pi, z, \gamma, \tau, m, v, q, a, b | y)$, extending the original posterior in (5) to include the new quantities $\gamma = \{\gamma_{ij}, i = 1, \dots, n; j = 1, \dots, s\}$ and $\tau = \{\tau_1, \dots, \tau_s\}$. There are difficulties in the posterior MCMC analysis due to the complicated form of (7) as a function of μ_j , and also due to the truncation of the basic normal model. These issues destroy part of the nice, conditionally conjugate sampling structure, and are currently handled using some direct analytic approximations and Metropolis-Hasting accept/reject steps in the extended simulation analysis. It is beyond the scope of the current paper to develop the technical aspects of this fully here, but our discussion would be incomplete without raising the issues of intrinsic variability, currently becoming vogue in the field, and without providing some exploratory data analysis. After further technical refinements and practical experience, full modelling and technical details will be reported elsewhere (West, 1997).

REFERENCES

- Bennett MR and Kearns JL (2000) Statistics of transmitter release at nerve terminals. *Prog. Neurobiol.* **60**, 545-606.
- Besag J and Green PJ (1993) Spatial statistics and Bayesian computation. *J. Roy. Statist. Soc., Ser. B*, **55**, 25-37.
- Bolshakov VY and Siegelbaum SA (1995). Regulation of hippocampal transmitter release during development and long-term potentiation. *Science* **269**, 1730-1734.
- Buhl EH, Halasy K and Somogyi P (1994) Diverse source of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites. *Nature* **368**, 823-828.
- Cao G and West M (1997) Bayesian analysis of mixtures of mixtures. *Biometrics* **52**, 221-227.
- Craig AM and Boudin H (2001) Molecular heterogeneity of central synapses: afferent and target regulation. *Nature Neuroscience* **4**, 569-578.
- Debanne D, Gahwiler BH and Thompson SM (1999) Heterogeneity of synaptic plasticity at unitary CA3-CA1 and CA3-CA3 connections in rat hippocampal slice cultures. *J. Neurosci.* **19**, 10664-10671.
- Del Castillo J and Katz B (1954) Quantal components of the end-plate potential. *J. Physiol. (Lond.)* **124**, 560-573.
- Durand GM, Kovalchuk Y and Konnerth A (1996) Long-term potentiation and functional synapse induction in developing hippocampus. *Nature* **381**, 71-75.
- Emptage NJ, Reid CA, Fine A and Bliss TV (2003). Optical quantal analysis reveals of presynaptic component of LTP at hippocampal Schaffer-associational synapses. *Neuron* **38** , 797-804.
- Escobar DM and West M (1995) Bayesian density estimation and inference using mixtures. *J. Am. Stat. Assoc.* **90**, 577-588.
- Faber DS, Young WS, Legendre P, and Korn, H. (1992) Intrinsic quantal variability due to stochastic properties of receptor-transmitter interactions. *Science* **258**, 1494-1498.
- Frerking M and Wilson M (1999) Differences in unquantal amplitude between sites reduce unquantal variance when few release sites are active. *Synapse* **32**, 276-287.
- Gelfand AE, and Smith AFM (1990) Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398-409.
- Glavinovic MI and Rabie HR (2001) Monte Carlo evaluation of quantal analysis in the light of Ca²⁺ dynamics and the geometry of secretion. *Pflugers Arch. - Eur. J. Physiol.* **443**, 132-145.
- Hanse E and Gustafsson B (2001) Quantal variability at glutamatergic synapses in area CA1 of the rat neonatal hippocampus. *J. Physiol. (London)* **531.2**, 467-480.

Harris KM and Sultan P (1995). Variation in the number, location and size of synaptic vesicles provides an anatomical basis for nonuniform probability of release at hippocampal CA1 synapses. *Neuropharmacology* **34**, 1387-1395.

Isaac JT, Nicoll RA and Malenka RC (1995). Evidence for silent synapses: Implications for the expression of LTP. *Neuron* **15**, 427-434.

Katz B (1969) The Release of Neural Transmitter Substances. Liverpool University Press, Liverpool.

Korn H and Faber DS (1991). Quantal analysis and synaptic efficacy in the CNS. *Trends in Neuroscience* **14**, 439-445.

Kraushaar U and Jonas P (2000) Efficacy and stability of quantal GABA release at a hippocampal interneuron-principal neuron synapse. *J. Neurosci.* **20**, 5594-5607.

Liu G and Tsien RW (1995). Properties of synaptic transmission at single hippocampal synaptic boutons. *Nature* **375**, 404-408.

Ludwig M and Pittman QJ (2003) Talking back: dendritic neurotransmitter release. *Trends Neurosciences* **26**, 256-261.

Magee JC and Johnston D (1997) A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science* **275**, 209-212.

Manwani A and Koch C (2000) Detecting and estimating signals over noisy and unreliable synapses: Information-theoretic analysis. *Neural Comp.* **13**, 1-33.

Mozhayeva MG, Yildirim S, Liu X and Kavalali ET (2002) Development of vesicle pools during maturation of hippocampal synapses. *J. Neurosci.* **22**, 654-665.

Oertner TG (2002) Functional imaging of single synapses in brain slices. *Exper. Physiol.* **87**, 733-736.

Oertner TG, Sabatine BL, Nimchinsky EA and Svoboda K (2002) Facilitation at single synapses probed with optical quantal analysis. *Nature Neuroscience* **5**, 657-664.

Pavlidis P and Madison DV (1999) Synaptic transmission in pair recordings from CA3 pyramidal cells in organotypic culture. *J. Neurophysiol.* **81**, 2787-2797.

Redman S (1990) Quantal analysis of synaptic potentials in neurons of the central nervous system. *Physiol. Rev.* **70**, 165-198.

Sayer RJ, Friedlander MJ and Redman SJ (1990) The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. *J. Neurosci.* **10**, 826 - 836.

- Smith AFM and Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser B.* **55**, 3-23.
- Sokolov MV, Rossokhin AV, Astrelin AV, Frey JU and Voronin LL (2002) Quantal analysis suggests strong involvement of presynaptic mechanisms during the initial 3 hr maintenance of long-term potentiation in rat hippocampal CA1 area in vitro. *Brain Res.* **957**, 61-75.
- Spruston N, Jaffe DB, Williams SH, and Johnston D. (1993). Voltage- and space-clamp errors associated with the measurement of electrotonically remote synaptic events. *J. Neurophysiol.* **70**,781-802.
- Staley KJ (1999) Quantal GABA release: noise or not? *Nature Neuroscience* **2**, 494-495.
- Stricker C, Field AC and Redman SJ (1996a) Statistical analysis of amplitude fluctuations in EPSCs evoked in rat CA1 pyramidal neurones in vitro. *J. Physiol. (Lond.)* **490**, 419-441.
- Stricker C, Field AC and Redman SJ (1996b) Changes in quantal parameters of EPSCs in rat CA1 neurones in vitro after the induction of long-term potentiation. *J. Physiol. (Lond.)* **490**, 443-454.
- Tierney LJ (1994) Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701-1728.
- Turner, DA (1984). Conductance transients onto dendritic spines in a segmental cable model of CA1 and dentate hippocampal neurons. *Biophys. J.* **46**, 85-96.
- Turner DA, Chen Y, Isaac J, West M and Wheal HV (1997) Synaptic site heterogeneity in paired-pulse plasticity in CA1 pyramidal cells. *J. Physiol. (Lond.)* **500**, 441-462.
- Turner DA and West M (1993) Bayesian analysis of mixtures applied to post-synaptic potential fluctuations. *J. Neurosci. Meth.* **47**, 1-21.
- Uteshev VV, Patlak JB and Pennefather PS (2000) Analysis and implications of equivalent uniform approximations of nonuniform unitary synaptic systems. *Biophys. J.* **79**, 2825-2839.
- Walmsley B (1995) Interpretation of 'quantal' peaks in distributions of evoked synaptic transmission at central synapses. *Proc. R. Soc. Lond. B.* **261**, 245-250.
- West M (1997) Hierarchical mixture models in neurological transmission analysis. *In press, J. Am. Stat. Assoc.*
- West M and Harrison PJ (1997) Bayesian Forecasting and Dynamic Models. (2nd Ed.) New York, Springer Verlag.
- West M and Turner DA (1994) Deconvolution of mixtures in analysis of neural synaptic transmission. *Statistician* **43**, 31-43.
- Williams SR and Stuart GJ (2003) Role of dendritic synapse location in the control of action potential output. *Trends Neurosciences* **26**, 147-154.

FIGURE LEGENDS

Figure 1 - Diagram of Synaptic Site

This diagram shows the presynaptic axon (to the left) and the presynaptic terminal containing vesicles and neurotransmitter. The presynaptic action potential (AP) conducted along the axon leads to calcium release and vesicle fusion and release occurs, though as a stochastic function. This leads to release probability (π), which is simply the number of releases occurring in comparison to the number of presynaptic action potentials. The neurotransmitter then binds to receptors on the postsynaptic side of the synaptic cleft (usually a dendritic spine in the hippocampus, which is attached to a dendrite), resulting in ionic channel opening and a postsynaptic current or voltage (EPSP) being generated in the dendrite, with peak amplitude μ . On any given neuron there may be multiple such release sites (k), acting in concert, but likely with different parameters. The EPSP signals are summated in time and space from these multiple synapses to determine the postsynaptic activity of the neuron.

Figure 2 - Data Set U2N1 with 1 Site

Figure 2A shows the mean trace from U2N1 data set ($n=244$ responses). Figure 2B shows the scatter plot for stationarity, indicating no change in the degree of fluctuation pattern over the course of the ensemble. The open circles show the responses and the closed circles the noise samples over the course of the ensemble. Figure 2C illustrates the response histograms and Bayesian analysis results. The smooth function (thick line) in Fig. 2C shows the noise function used in the analysis, overlapping the failure peak. The smooth line indicates the predictive Bayesian pdf and the dotted lines the error ranges around the pdf. Figure 2D shows the individual synaptic site parameters, with a dual plot comparing site amplitude on the abscissa versus site probability on the ordinate. Note the single site at approximately 10 pA, corresponding well with the histogram in Figure 2C. This ensemble shows an example of a simple response with either failures or occurrences.

Figure 3 - Data Set U4N1 with 2 Sites

Figure 3A shows the mean trace from U4N1 data set ($n=386$ responses). Figure 3B shows the scatter plot for stationarity, indicating no change in the degree of fluctuation pattern over the course of the ensemble. The open circles show the responses and the closed circles the noise samples over the course of the ensemble. Figure 3C illustrates the response histograms and Bayesian analysis results. The smooth function (thick line) in Fig. 3C shows the noise function used in the analysis, overlapping the failure peak. The smooth line indicates the predictive Bayesian pdf and the dotted lines the error ranges around the pdf. Figure 3D shows the individual synaptic site parameters, with a dual plot comparing site amplitude on the abscissa versus site probability on the ordinate. Note the dual sites, with the first at approximately 4 pA and the second at approximately 18 pA, corresponding well with the histogram in Figure 3C.

Diagram of Synaptic Site

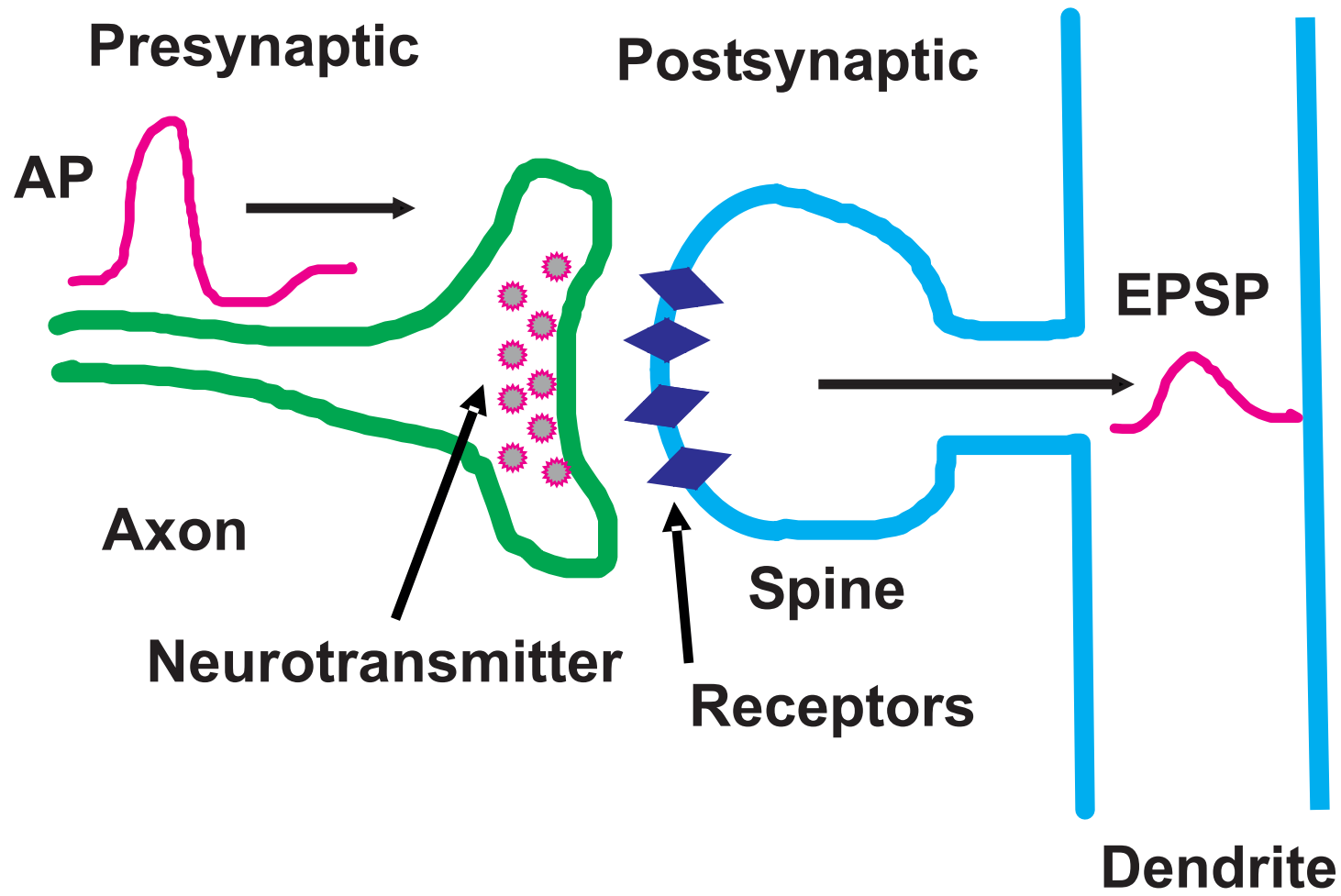
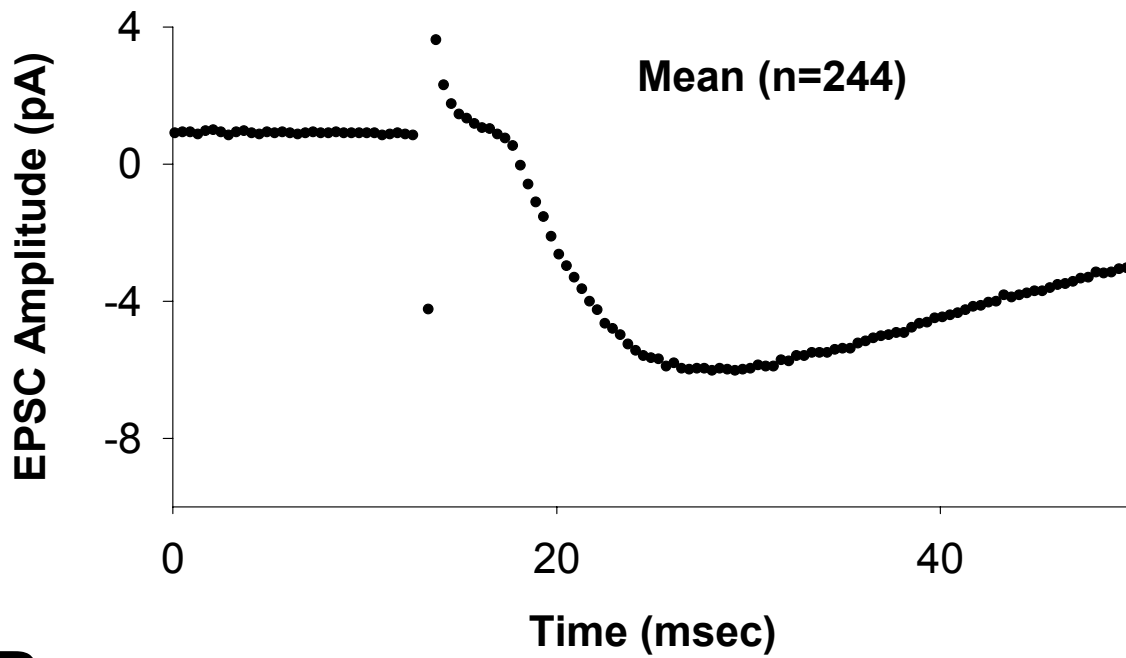
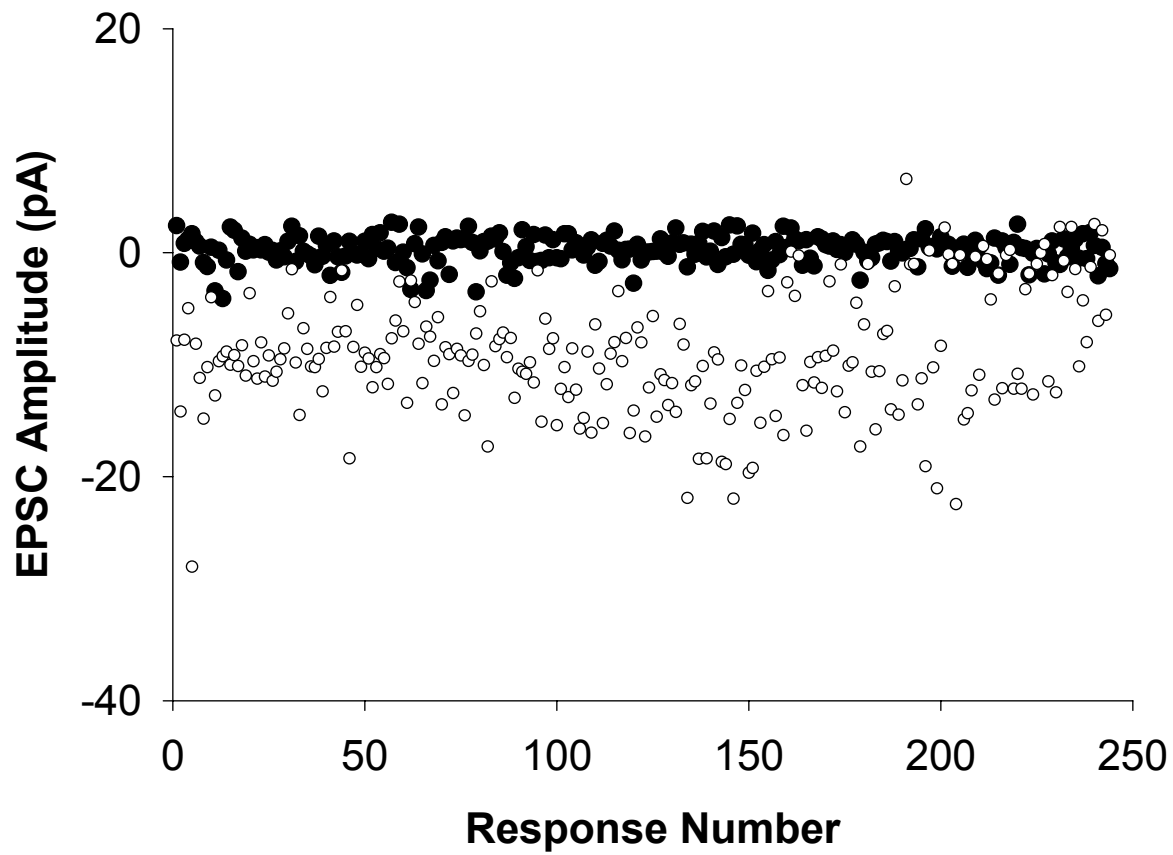
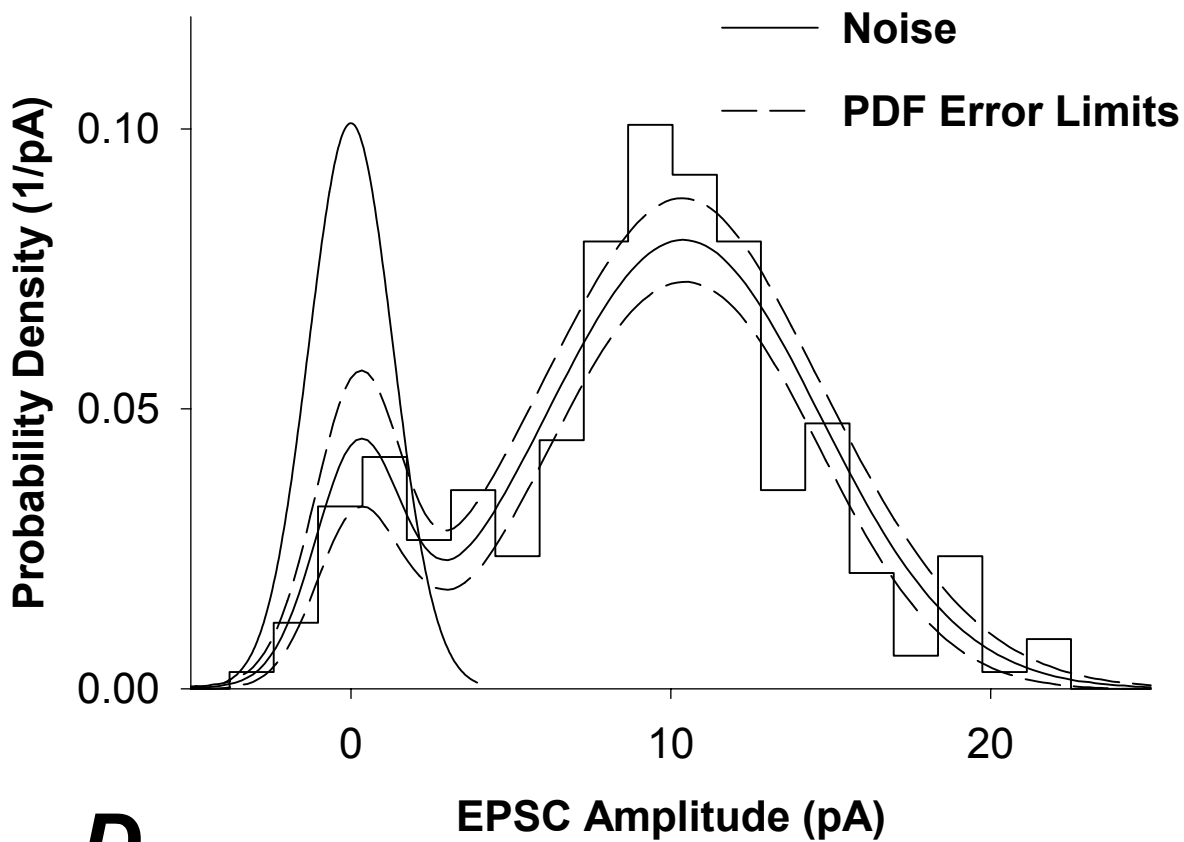
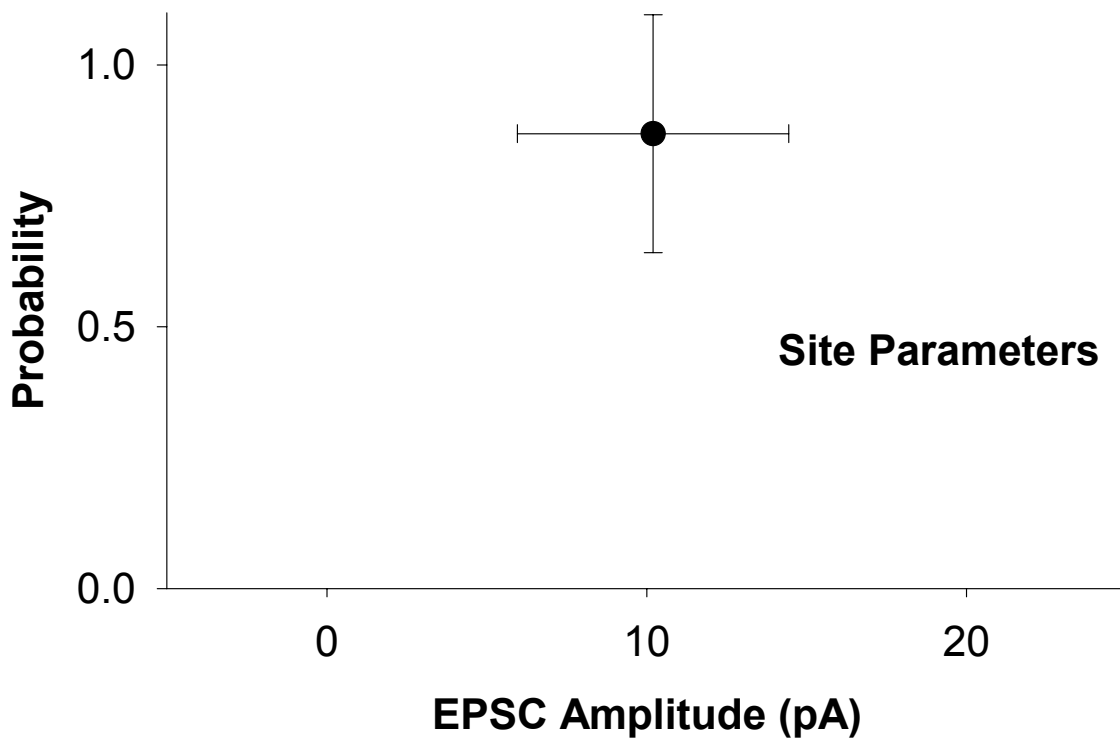
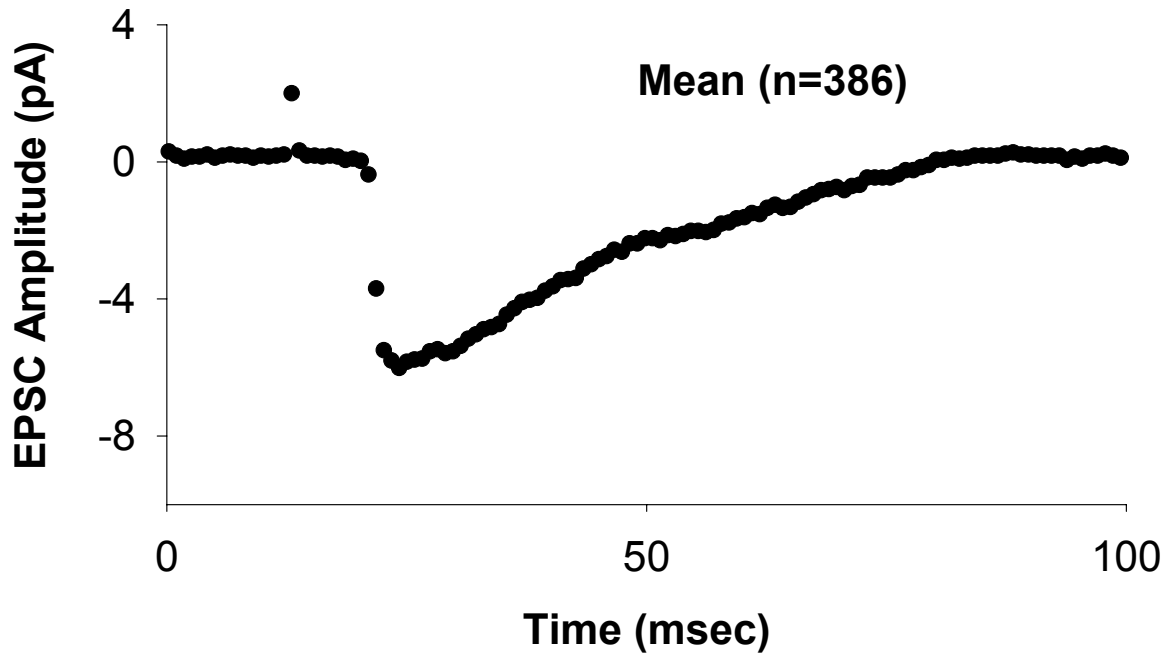


Fig. 1

A**Data Set U2N1****B****Fig. 2**

C**Data Set U2N1****D****Fig. 2**

A Data Set U4N1



B

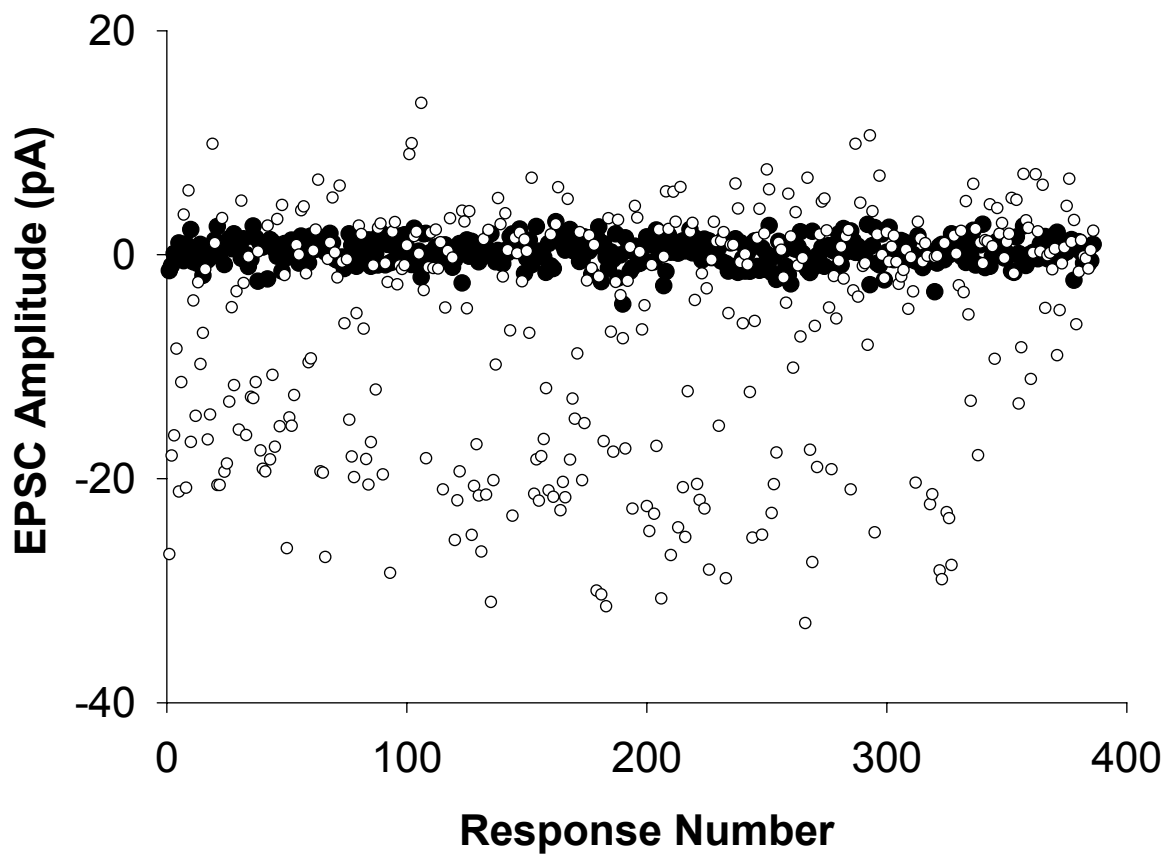
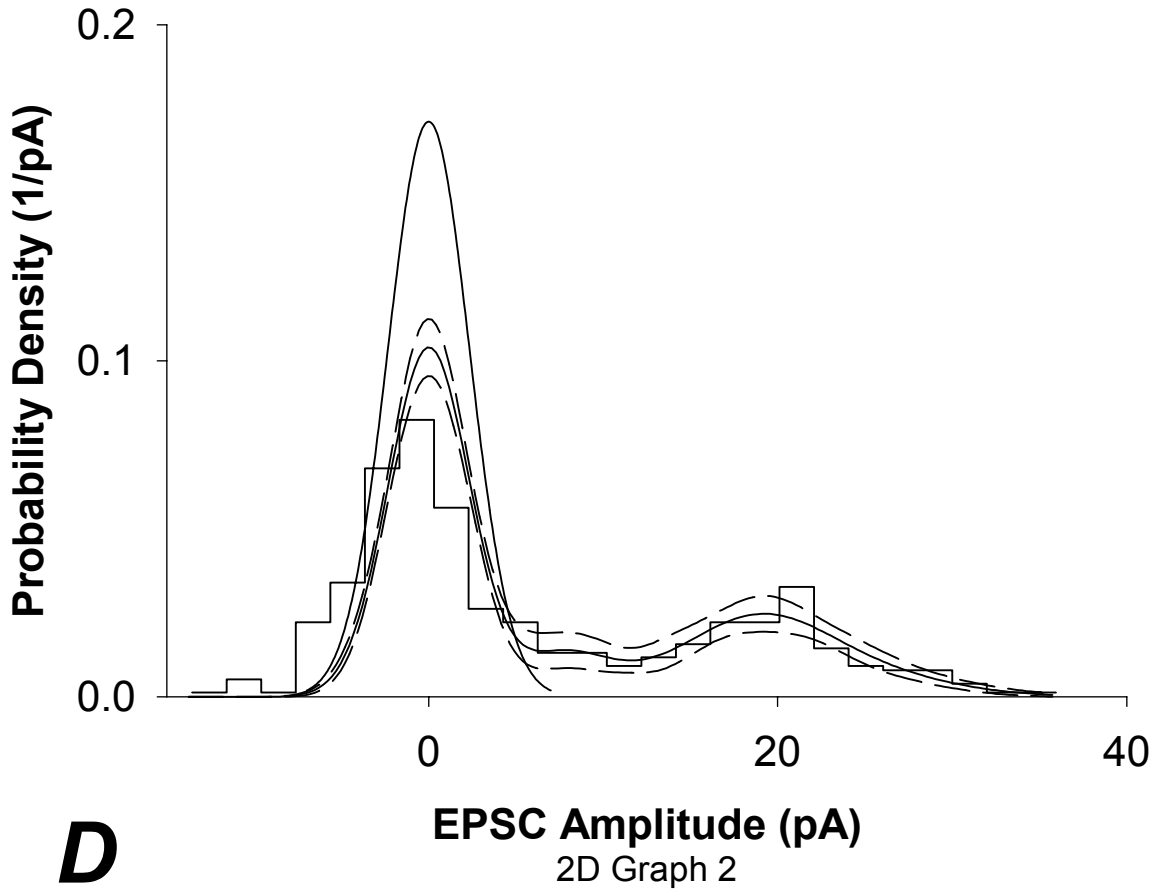
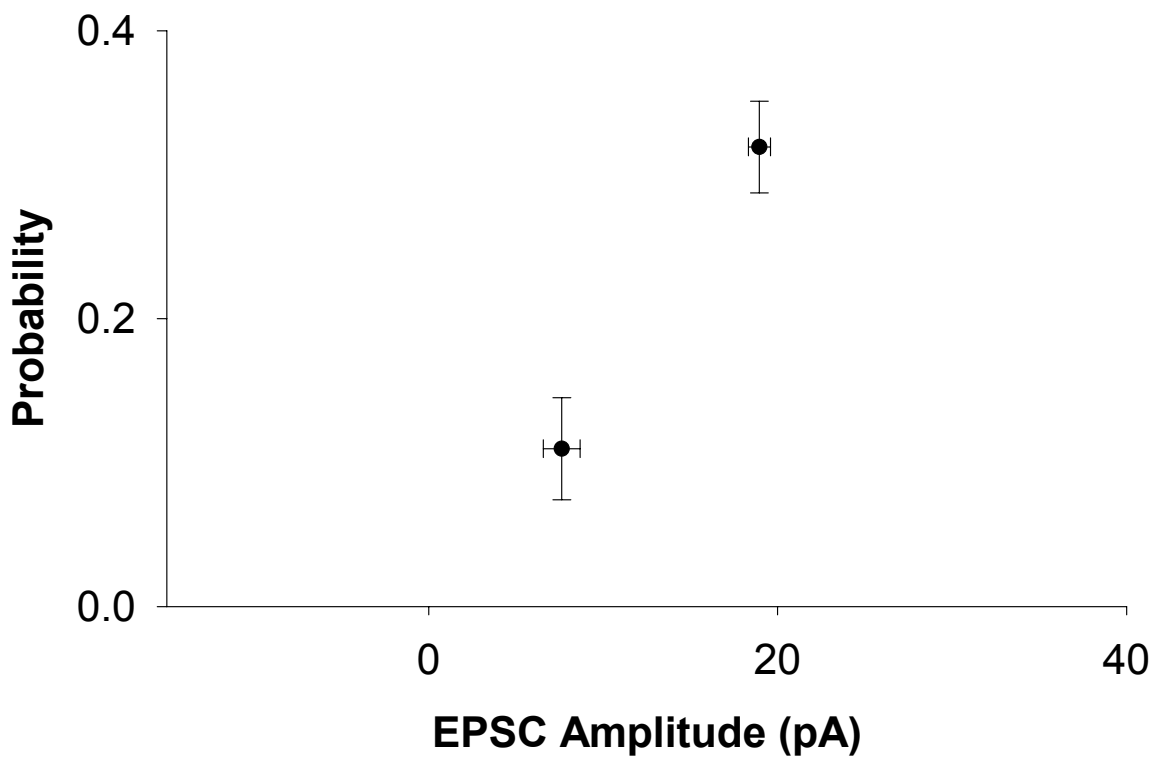


Fig. 3

C**Data Set U4N1****D****Fig. 3**