Statistical, Stochastic, and Dynamical Models of Neural Decision Making

Nicholas Cain

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2012

Reading Committee:
Eric Shea-Brown, Chair
Hong Qian
Michael Shadlen
Emanuil Todorov

Program Authorized to Offer Degree:
Department of Applied Mathematics

Dedication

Especially to my family, but to all those who have: listened, consoled, commiserated, counseled, advised, encouraged, and celebrated, (repeating these steps as many times as necessary), I dedicate this dissertation with my deepest gratitude.

University of Washington

Abstract

Statistical, Stochastic, and Dynamical Models of Neural Decision Making

Nicholas Cain

Chair of the Supervisory Committee:
Dr. Eric Shea-Brown
Department of Applied Mathematics

Models of decision making provide a direct link between behavior and neurobiology. How does the encoding and accumulation of evidence by neural circuits impact decision making performance? Through data-driven and biophysically-based modeling, abstract models for simple decisions can be made more realistic, and may eventually explain how biological organisms can robustly make more complicated decisions. This dissertation investigates several models of this type, spanning a wide range of model abstraction. Deriving, parameterizing, and analyzing these models requires techniques from signal processing, mathematical statistics, stochastic processes, and dynamical systems.

In many cases, we have found surprising roles for nonlinearities in the circuits that accumulate sensory evidence. In simple models, linear integration of evidence-encoding stimuli enables optimal decision making. This integration may be unstable in practice, however a nonlinear thresholding mechanism can ameliorate this deficit while retaining nearly optimal performance. Moreover, when sensory evidence is encoded in a correlated population of spiking neurons, pre-

processing nonlinearities are exactly the prescription for optimal inference. We also examine a reduced model of a network of spiking neurons, in order to determine how nonlinear dynamics and noise combine to dictate performance. Our results suggest nonlinear dynamics may not significantly diminish performance, compared to the noise sources in biophysically motivated models of evidence accumulation. Combined with experimental studies, the exciting, and sometimes counterintuitive, effects of nonlinear computations may eventually elucidate the neural mechanisms of decision making.

Acknowledgements

The author wishes to acknowledge the generous contributions of several individuals and institutions that have made this dissertation possible. Foremost is the tireless advice and constructive criticism of Dr. Eric Shea-Brown, who's role as advisor and mentor has been instrumental on all fronts of this work. The National Science Foundation (VIGRE, GK-12, and TeraGrid programs), the Burroughs-Wellcome Fund, the Northwest Center for Neural Engineering, and the University of Washington eScience Institute have all provided generous financial support. The Department of Applied Mathematics has provided outstanding Teaching Assistant and Instructor opportunities that simultaneously provided financial support and opportunities to gain experience in undergraduate instruction. Finally, many scientists and friends have given selflessly of their expertise and support, including: Mike Shadlen, Andrea Barreiro, Sander Keemink, Shin Kira, Adrienne Fairhall, Fred Rieke, Hong Qian, Emo Todorov, Yu Hu, Josh Goldwyn, Guillaume Lajoie, Alex Cayco-Gajic, David Leen, Evan Thilo, and the faculty and students of Applied Mathematics department.

Contents

List of Figures

CHAPTER 1

Introduction and Background

Mathematical modeling is the science and art of approximating complex systems with a system of mathematical equations. Often, the goal of a mathematical model is to reduce the complexity of a system through a well-motivated elimination of factors, so that the more fundamental principles governing the system can be understood. Currently, the most complicated physical system in the known universe is the central nervous system; the staggering number of synaptic connections in the brains of even simple organisms underly the richly fascinating variety of behaviors in the natural environment.

The complexity of the organ responsible for these behaviors necessitates reduced mathematical models of their function. Decision making is a natural target for this analysis, and will be the focus of this dissertation. Here we formulate dynamical, stochastic, probabilistic, and statistical models of neural activity, based on principled simplifications of known neuroanatomy and neurophysiology. We then explore how these models might instantiate the mathematical principles of decision making theory, and what the consequences of these models are for behavioral performance.

Sensory discrimination tasks provide an ideal starting point for modeling decision making, because instantaneous snapshots of sensory input represent single, noisy samples that alone carry limited evidence. Together, however, these samples can provide enough clues for the organism to make a sound judgment. A prototypical example is the *random dots task*, in which the subject must deter-

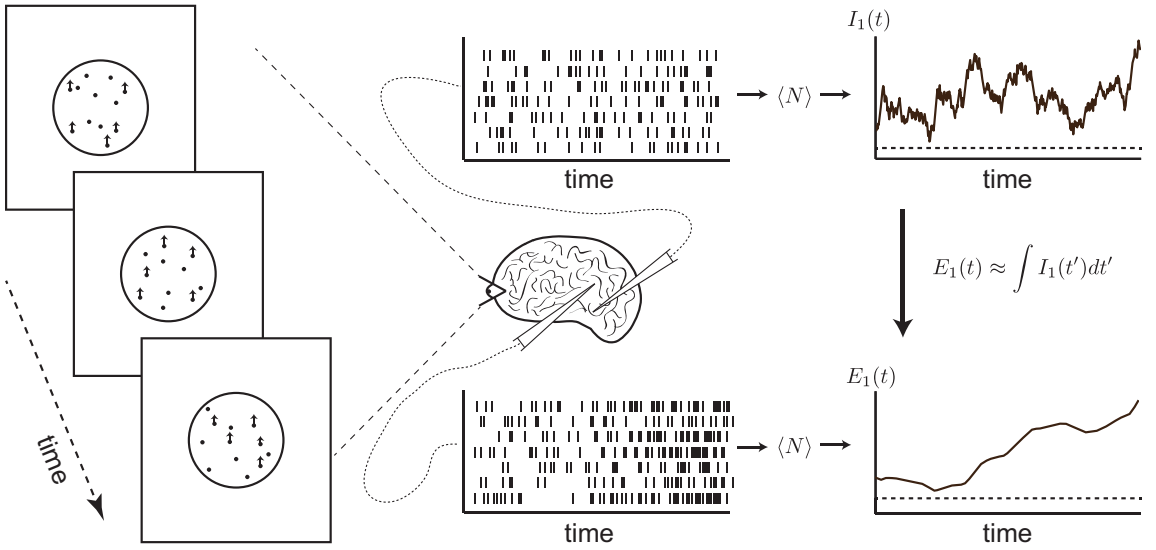


Figure 1.1: Schematic of sensory evidence accumulation over time. In the random dots task, a field of irregularly moving dots are presented to the subject, with motion biased toward one favored direction. While some neurons represent "instantaneous" values of this motion evidence (top rastergram), other neurons appear to represent its accumulated value over time (bottom rastergram). According to the equation shown, the relationship between the average activity of these neural pools approximates temporal integration.

mine the average direction of motion in a field of randomly moving dots (Figure 2.2) [103, 117, 59, 72, 88, 121, 9].

In this chapter, we review recent progress on how this type of accumulation might be implemented by neural circuits. We begin by describing how optimal mechanisms for evidence encoding and decision making might arise from the "neural hardware". We then discuss the stability and variability of these circuits, their adaptability in uncertain decision making environments, and how they might operate with irregular and noisy components.

1.1 Integration, optimality, and functionality

What is the optimal way to accumulate evidence over time? Many different definitions of "optimal" – from maximizing the rate at which correct responses are

made [42, 9] to maximizing guaranteed payoff [132] to lossless Bayesian decoding [4] – depend on the same key computation. This is to update the statistical likelihood of each decision alternative being correct, relative to the other alternatives, when each increment of evidence arrives [125, 43, 111, 42]. In general this is a tall order, requiring computation of an arbitrary function of the momentary neural activity that encodes incoming evidence, and then combining this with the evidence already accumulated. Making matters worse, even knowing what function to compute depends on full knowledge of the statistics of that activity [111].

Fortunately, in a number of circumstances, the likelihood calculation is equivalent to directly summing, or integrating, the evidence stream over time (see Figure 2.2). In the simplest case there are two task alternatives, and two independent streams of incoming evidence $I_1(t)$ and $I_2(t)$ – for example, representing the output of neurons selective for upward vs. downward motion. Perhaps the best known result says that, if $I_1(t)$ and $I_2(t)$ are uncorrelated gaussian signals with identical variance, then the relative likelihood of each alternative can be computed by integrating the signals over time – that is, via the accumulated inputs

$$E_1(t) = \int_0^t I_1(t')dt' \tag{1.1.1}$$

$$E_2(t) = \int_0^t I_2(t')dt' \tag{1.1.2}$$

[125, 43, 111, 42].

If $I_1(t)$ and $I_2(t)$ are Poisson processes – i.e., spike trains [133, 4, 64, 109] – the optimal computation is again integration over time. In fact, this setting is much more flexible, as there is no longer a requirement that all inputs have the same variance ([133, 109] identify limits in which the Poisson and gaussian cases agree). Moreover, for decisions involving multiple alternatives, where motion could follow any direction, the same result extends. In this case, there will be N different input streams $I_i(t)$, whose integrals produce N accumulated variables $E_i(t)$. Additionally, for the two alternative case, [78] extends these results to compute estimates of the quality of evidence streams for each task alternative.

When the statistics of incoming evidence vary over the course of a trial, in some cases the relative likelihood of task alternatives can still be computed based on the

integral of the inputs. For two gaussian input streams, this holds when the ratio of the "signal" $\langle I_1(t) - I_2(t) \rangle$ in the inputs to the variance of these inputs is constant over time [14]. For N Poisson input streams, there are stronger results: if the time dependence is via a *gain* term that uniformly scales up or scales down the firing rate of each input stream, then integration once again enables a direct computation of the relative likelihoods; similar considerations hold for variation across trials [4].

Can integration as a neural computation be observed in the activity of neural circuits? Churchland et al. [19] presents compelling evidence, devising a novel method to measure the time-evolution of the *variance* of spike rates believed to represent the integrals $E_j(t)$. The key finding is that this variance grows linearly in time, consistent with what is expected from an integrating process.

Nevertheless, computing an integral is not the end of the story: integrated evidence must trigger a decision. For the two alternative case, the highest accuracy at any given speed is achieved by making a decision when the difference $E_1(t) - E_2(t)$ crosses a preset bound [125, 42, 111, 54]. (We note that this still leaves open precisely what tradeoff between speed and accuracy the decision maker will select, a distinct and interesting question [69, 32, 132, 9, 133, 42].) For multiple alternatives, the $E_i(t)$ can be used to implement a "MSPRT" rule which can also optimize speed and accuracy [8, 133], or to find the maximum likelihood alternatives at any given time [4]. While in principle these decision rules could be computed in "downstream" areas, recurrent inhibition within the same networks that integrate the evidence streams $I_j(t)$ can often implement or closely approximate the required differencing operations [121, 14, 130, 9, 73].

1.2 Precision, thresholding, and the neural ratchet

While "pure" integration of incoming evidence is optimal in many settings, it poses a challenge for neural circuitry. Evidence is integrated over hundreds of milliseconds or longer as decisions develop, but the activity of individual neurons and synapses tend to decay with timescales that are several orders of magnitude faster. How can this rapid decay be countered? A classical solution is via feedback connections tuned to balance – and hence cancel – passive voltage leak and synaptic

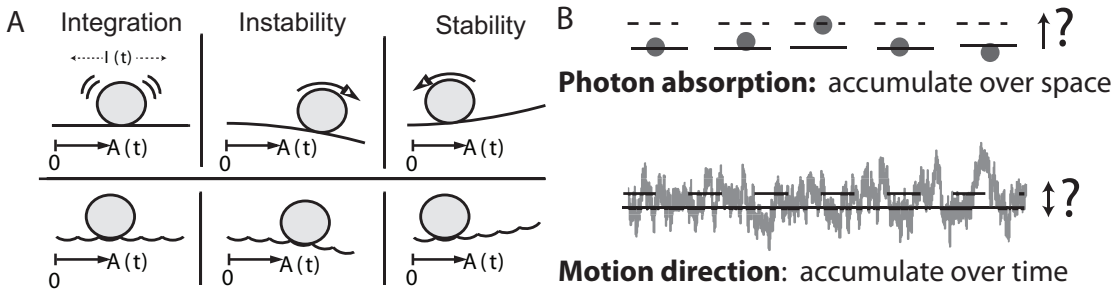


Figure 1.2: (A) Visualizing integration via an energy surface [84, 68, 46], where the location of the ball indicates integrated evidence $E(t)$. A robust integrator can "fixate" at a range of discrete values, indicated by a sequence of potential wells, despite imprecise tuning of circuit feedback. Without these wells (the non-robust case), activity in a mistuned integrator would either exponentially grow or decay, as in the top panels. Perturbing the robust integrator from one well to the next, however, requires sufficiently strong momentary input $I(t)$. (B) Relationship of robust integration to the detection of sparse visual signals. Top: Activity of five photoreceptors is shown as a dot, and the dotted line represents photon absorption. Light that arrives at only a tiny fraction of photoreceptors (here receptor # 3) must be detected. Bottom: Motion evidence fluctuates over time, and the direction of the time-average must be discriminated among two possibilities – i.e., up (dotted line) vs. down (solid line).

decay [46, 92]. This process is illustrated in Figure 1.2(a) via motion of a ball on an energy surface. The position $E(t)$ represents the total activity of a circuit (say, average firing rate relative to a baseline marked 0). If decay dominates (upper-right), then $E(t)$ always has a tendency to "roll back" to baseline values; we say the integrator circuit is *stable*. Conversely, if feedback connections are in excess, then activity will grow away from the baseline value and thus the circuit is *unstable*. Pure integration is achieved (upper-left) with feedback that is precisely tuned to match decay, so that changes in $E(t)$ represent integrated evidence alone [100, 46].

Biologically, this feedback could occur in many forms. Beyond excitatory connections, positive feedback can also occur through opponent inhibition [68, 121, 15, 9, 126, 8, 46]; recently, a generically valid "normal form" equation that describes the resulting dynamics was presented and its predictions favorably compared with behavioral data [95]. Interestingly, unstructured networks also appear able to accumulate inputs over time: a single "integration mode" may emerge naturally as the connection statistics are varied in large, randomly coupled networks [36]. Recent studies have also raised the possibility of achieving integration without depending on feedback per se, rather by passing inputs through extended sequences of "functionally feedforward" states within a large network [37, 44, 62]. This mechanism is one of several that produce high-dimensional dynamics – that is, with time constants that differ from cell to cell – in contrast to classical line attractor models. Intriguingly, such cell-to-cell heterogeneity has recently been found in the zebrafish oculomotor integrator circuit [76].

However they are implemented, must feedback effects always be precisely tuned to produce integration over time? The *robust integrator* mechanism of [56] presents an alternative. Here, the circuit as a whole "ratchets" among many stable states, as illustrated in the bottom row of Figure 1.2(a) by introducing a series of wells into the energy surface [84, 68, 46]. Importantly, even with imprecisely tuned feedback, the network now avoids the tendency to intrinsically decay or grow without incoming evidence. Thus, the activity $E(t)$ approximates the integral of $I(t)$, without the effects of decay or runaway excitation.

Nevertheless, such robustness impacts how the evidence streams $I_j(t)$ are processed. As the energy wells illustrate, instantaneous inputs $I_j(t)$ below a thresh-

old value will fail to perturb the state from one well to another [56, 45], and are therefore ignored by the integrator. The idea of thresholding inputs has a long history [47, 35], and several studies explore its implications for decision making over time. In [112], the authors show that a closely related "interval of uncertainty" model can match distributions of reaction times and correct vs. incorrect responses better than an allied model without a threshold. Moreover, [86] develops a model with a threshold-like "gating" effect, finding improved fits to single unit recordings while simultaneously matching behavioral data (see also [99]). Several key statistical and dynamical properties of robust integrators that contribute to the quality of this match are studied in [74].

Chapter 2 asks how robust integration impacts the optimality of decision performance. For example, consider the case introduced above where instantaneous evidence is carried by gaussian signals $I_1(t)$ and $I_2(t)$. Rather than accumulate the entire signal, a robust integrator would ignore all samples below a preset threshold. Intuitively, one might expect this to diminish the speed and accuracy with which decisions can be made. However, results to date [16] suggest that the loss is minimal, even when more than half of the signal is thresholded away. This suggests that most of the evidence can be gleaned from large deviations from the mean, in the "tails" of the signals.

Moving beyond the case of gaussian inputs, analogy with other fields suggests some settings in which thresholding actually improves decision performance. For example, Chapter 6 considers the case of detecting photons at low light levels, with noisy photoreceptors (see Figure 1.2(b)). If the output of all photoreceptors is simply summed – i.e., integrated – this sparse signal could easily be lost in noise. Passing the photoreceptor signal first through a nonlinear thresholding function provides a solution [35], analogous to integrating only those inputs strong enough to move the robust integrator state from well to well. For decisions based on accumulating *temporally* sparse evidence over a noisy background – such as the auditory and visual impulses recently studied in rats and humans [87] – thresholding integrators would then be expected to improve decision performance. Taken together, these observations point to a potentially favorable role for robust integrators in decision making, beyond an insurance policy against mistuning. Ignoring

the weaker parts of the input signal is sometimes the best strategy, and when it is not, the cost of doing so can be surprisingly low.

1.3 Integration, background, and circuit-driven noise

Above, we have treated the transformation of the incoming evidence $I(t)$ into its integral $E(t)$ as a deterministic mathematical operation. However, $E(t)$ is presumably computed by irregularly spiking neurons (Figure 2.2, bottom raster), perhaps distributed throughout the brain – and possibly incorporating "background" neurons that are untuned or poorly tuned to a specific decision. Do integrator circuits with these noisy components contribute additional variability to decisions?

Empirical work suggests two constraints. First, well-established patterns of accuracy and reaction time distributions [93, 111, 43] must be compatible with whatever type of circuit-based noise arises. Second, for the moving dots task, a number of studies have precisely measured the signal and noise properties of motion evidence at a stage (area MT) that appears to precede its accumulation over time [134, 12, 13, 23]. Recently, building on prior analyses [102, 70], [23] showed that a noiseless accumulation, or sum, of MT spikes over an empirically estimated distribution of reaction times predicts levels of decision performance that closely match experiment. This only leaves room for a relatively small amount of additional variability to be contributed by the integrator circuit itself.

Spiking neuron models of integrator circuits have met some of these constraints. A number of studies show that such circuits can reproduce behavioral statistics as well as key features of neural activity [126, 4, 110, 27, 70]. In addition, [4] shows that additional properties of noiseless integration – including how confidence in task alternatives grows over time – are preserved in a simple integrator circuit with Poisson-like firing (see also [70, 107]). For biologically detailed models – especially those in which the long time constants of integration arise through recurrent network interactions [126, 127, 27, 128] – future work will be needed to understand how much variability arises from the fluctuating inputs $I(t)$, from background neurons, and from the irregular dynamics of the circuits themselves.

Three recent advances in theoretical neuroscience lay important groundwork.

The first is the model of [110], in which past evidence is represented by spikes that recur at randomly spaced times, as in a Poisson process – allowing them to continue to contribute variability to the decision process. While this is an abstract model, it does separate the variability driven by the evidence stream and that generated by mechanisms within the integrator circuit itself, a perspective that will remain important in future studies.

The second, related development concerns how integrator circuits process both noise in incoming signals and noise local to each "cell" in the circuit [62, 37, 83]. In particular, functionally feedforward architectures amplify incoming signals faster than internally occurring noise. This could clearly improve decision making performance, although constraints such as saturation of firing rates may limit when this strategy can be applied [62, 37]. Finally, [7] shows how lossless statistical inference over time can be performed by the (integrate-and-fire) dynamics of spiking neural networks. Importantly, this study shows how irregular, variable spiking and optimal neural integration can coexist, and suggests that – at least in principle – all of the truly detrimental noise in integrator circuits could be sourced to variability upstream.

1.4 Summary

Models of decision making have an enduring theoretical foundation in the accumulation of evidence streams over time. In more and more settings, we have learned that this computation can be implemented by simply integrating neural activity over time. We have also learned about when and why we need to reach beyond this concept. For example, exact integration over long timescales may be difficult to achieve in realistic circuits. In Chapter 3 I derive a reduced model of a model capable of this long timescale integration, that can accumulate evidence for a decision at the cost of ignoring weak fluctuations. However, in Chapter 2 I describe how this apparent "bug" does not necessarily impact decision making ability. In the end, this aspect may become a "feature" of decision making circuits, enabling good performance in unpredictable decision making environments.

Chapter 4 demonstrates that nonlinearities in integration circuits may in fact

hold the key to optimal inference when faced with evidence encoded in correlated input pools. Adding to the possibilities, different neural integrator circuits vary not only in how they process fluctuating evidence, but also in whether and how they may contribute additional noise to the decision computation. Chapter 5 attempts to explain the decision making performance of a complicated neural integration model by appreciating exactly the extent to which this additional noise reduces decision making accuracy. Finally, Chapter 6 presents a model for flash detection by the retina that applies decision making theory to optimal signal detection tasks. The future is bright for careful combinations of experiment, modeling, and theory that will link increasingly realistic circuit models with optimal algorithms for decision making in uncertain environments.

CHAPTER 2

Analysis of a Robust Neural Integrator

2.1 Introduction

Many decisions are based on the balance of evidence that arrives at different points in time. This process is quantified via simple perceptual discrimination tasks, in which the momentary value of a sensory signal carries negligible evidence but correct responses arise from summation of this signal over the duration of a trial. At the core of such decision making must lie neural mechanisms that integrate signals over time [43, 127, 9]. The function of these circuits is intriguing, because perceptual decisions develop over hundreds of milliseconds to seconds, while individual neuronal and synaptic activity often decays on timescales of several to tens of milliseconds – a difference of at least an order of magnitude. A mechanism that bridges this gap is feedback connectivity tuned to balance – and hence cancel – inherent voltage leak and synaptic decay [18, 121].

The tuning of recurrent connections to achieve this balance presents a challenge [100, 101], illustrated in Figure 2.1A (top) via motion of a ball on a smooth energy surface. Here, the ball position $E(t)$ represents the total activity of a circuit (relative to a baseline marked 0); momentary sensory input perturbs $E(t)$ to increase or decrease. If decay dominates (upper-right), then $E(t)$ always has a tendency to "roll back" to baseline values, thus forgetting accumulated sensory

input. Conversely, if feedback connections are in excess, then activity will grow away from the baseline value (center). If balance is perfectly achieved via fine-tuning, (left) temporal integration can occur. That is, inputs can then smoothly perturb network activity back and forth, so that the network state at any given time represents the time-integral of past inputs.

[56] proposed an alternate model: a ratchet-like accumulator, equivalent to movement along a scalloped energy surface (Figure 2.1A, bottom) [84, 46]. Importantly, even without finely-tuned connectivity, network states can hold prior values without decay or growth, allowing integration of inputs over time. Thus, this mechanism is called a *robust integrator*. Energy wells can be spaced arbitrarily close together while maintaining their depth, so that the robust integrator can represent a practically continuous range of values. However, the energy wells imply a minimum input strength to transition between adjacent states, with inputs below this limit effectively ignored.

The two models just introduced present a tradeoff between robustness to parameter mistuning and sensitivity to inputs. Here, we ask how this tradeoff impacts behavioral performance in perceptual decision making. We find that the tradeoff is favorable: decision speed and accuracy is lost when the integrator circuit is mistuned, but this loss is partially recovered by making the network dynamics robust. Thus, although the robust integrator discards the weakest portions of the evidence stream, enough evidence is retained to produce decisions that are faster and more accurate than would occur with unchecked over- or under-tuning of the circuit (Figure 2.1). Moreover, we establish that the robust integrator model is consistent with established empirical data. The implication is that robust integrators may be remarkably well suited to subserve a variety of decision making computations.

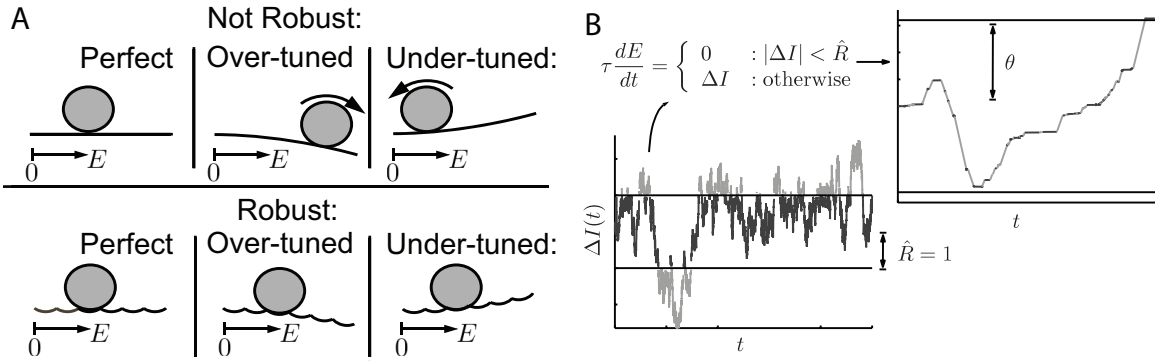


Figure 2.1: Schematic of neural integrator models. (A) Visualizing integration via an energy surface [84, 46]. The robust integrator can "fixate" at a range of discrete values, indicated by a sequence of potential wells, despite mistuning of circuit feedback. Without these wells (the non-robust case), activity in a mistuned integrator would either exponentially grow or decay, as in the top panels. Perturbing the robust integrator from one well to the next, however, requires sufficiently strong momentary input. (B) As a consequence, low-amplitude segments in the input signal $\Delta I(t)$, below a robustness limit R, are not accumulated by a robust integrator: only the high-amplitude segments are. The piecewise definition of Equation 2.2.7 captures this robustness behavior, resulting in the accumulated activity shown, and may be related to, e.g., a detailed bistable-subpopulation model. A decision is expressed when the accumulated value $E(t)$ crosses the decision threshold θ.

2.2 Materials and methods

2.2.1 Model and task overview

To explore the consequences of the robust integrator mechanism for decision performance, we begin by constructing a two-alternative decision making model similar to that proposed by [70]. For concreteness, we concentrate on the forced choice motion discrimination task [93, 70, 43, 20, 103, 104]. Here, subjects are presented with a field of random dots, of which a subset move coherently in one direction; the remainder are relocated randomly in each frame. The task is to correctly choose the direction of coherent motion from two alternatives (i.e., left vs. right).

As in [70] (see also [109]), we first simulate a population of neurons that represent the sensory input to be integrated over time. This population is a rough model of cells in extrastriate cortex (Area MT) which encode momentary information about motion direction [13, 12, 97]. We pool spikes from model MT cells that are selective for each of the two possible directions into separate streams, labeled according to their preferred “left" and “right" motion selectivity (see Figure 2.2).

Two corresponding integrators then accumulate the difference between these streams, left-less-right or vice-versa. Each integrator therefore accumulates the evidence for one alternative over the other. Depending on the task paradigm, different criteria may be used to terminate accumulation and give a decision. In the *reaction time* task, accumulation continues until activity crosses a decision threshold: if the leftward evidence integrator reaches threshold first, a decision that overall motion favored the leftward alternative is registered. In a second task paradigm, the *controlled duration* task, motion viewing duration is set in advance by the experimenter. A choice is made in favor of the integrator with greater activity at the end of the stimulus duration.

Accuracy is defined as the fraction of trials that reach a correct decision. Speed is measured by the time taken to cross threshold starting from stimulus onset. Reaction Time (RT) is then defined as the time until threshold (decision time) plus 350 ms of non-decision time, accounting for other delays that add to the time taken to select an alternative (e.g. visual latencies, or motor preparation time, cf. [70, 67]). The exact value of this parameter was not critical to our results. Task difficulty is

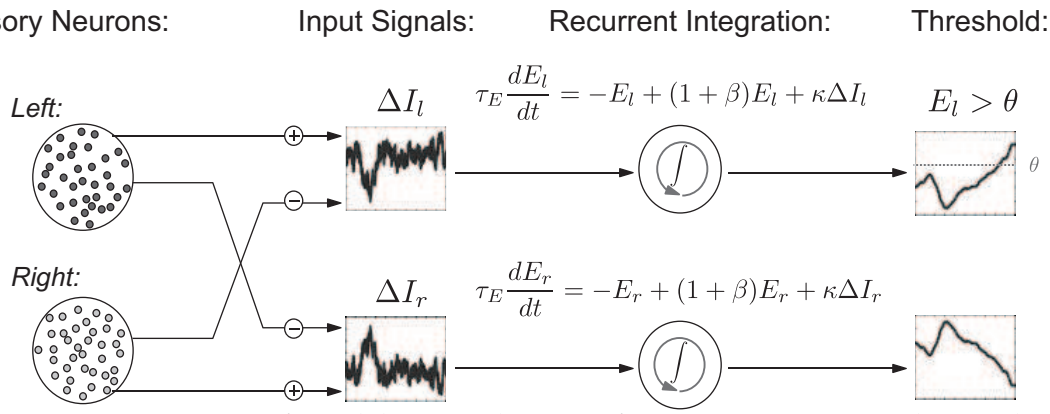


Figure 2.2: Overview of model. Simulations of sensory neurons and neural recordings are used to define the left and right inputs $\Delta I_l(t)$, $\Delta I_r(t)$ to neural integrators. These inputs are modeled by Gaussian (OU) processes, which capture noise in the encoding of the motion strength by each pool of spiking neurons (see Equations 2.2.1-2.2.3 for definition of input signals). Similar to [70], the activity levels of the left and right integrators $E_l(t)$ and $E_r(t)$ encode accumulated evidence for each alternative. In the reaction time task, $E_l(t)$ and $E_r(t)$ race to thresholds in order to determine choice on each trial. In the controlled duration task the choice is made in favor of the integrator with higher activity at the end of the stimulus presentation.

determined by the fraction of coherently moving dots C [12, 70, 93]. Accuracy and RT across multiple levels of task difficulty define the accuracy and chronometric functions in the reaction time task, and together can be used to assess model performance. When necessary, these two numbers can be collapsed into a single metric, such as the reward per unit time or *reward rate*. In the controlled duration task, the only measure of task performance is the accuracy function.

2.2.2 Sensory input

We now describe in detail the signals that are accumulated by the integrators corresponding to the “left" and “right" alternatives. First, we model the pools of leftward or rightward direction-selective sensory (MT) neurons as 100 weakly correlated (Pearson’s correlation $\rho = .11$ [134, 3]) spiking cells (see Figure 2.3). As in [71], neural spikes are modeled via unbiased random walks to a spiking threshold, which are correlated for neurons in the same pool. Increasing the variance of each step in the random walk increases the firing rate of each model neuron; it was

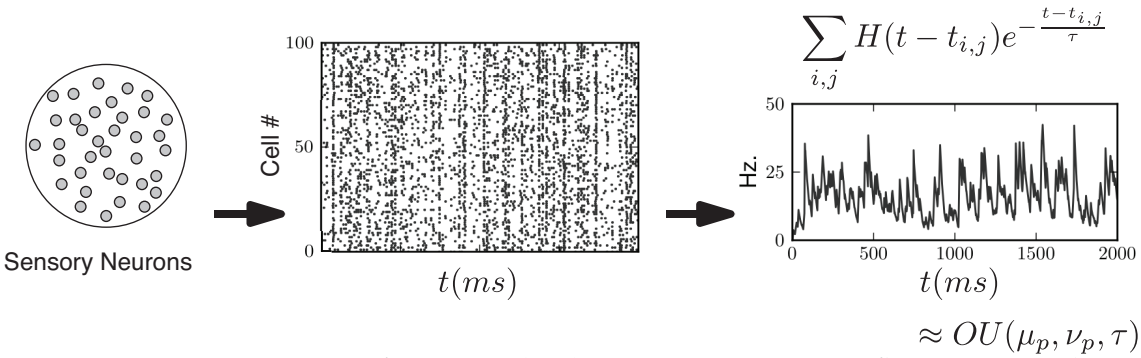


Figure 2.3: Construction of Gaussian (OU) processes to represent fluctuating, trial-by-trial firing rate of a pool of weakly correlated MT neurons [3, 134]. As in [71], these motion sensitive neurons provide direct input to our model integrator circuits. Simulated spike trains from weakly-correlated, direction selective pools of neurons are shown as a rastergram. All spikes prior to time t – a sum over the j^{th} spike from the i^{th} neuron, for all i and j – are convolved with an exponential filter, and then summed to create a continuous stochastic output (right); here, $H(t)$ is the Heaviside function. We approximated this output by a simpler Gaussian (OU) process in order to simplify numerical and analytical computations that follow.

therefore chosen at each coherence value to reproduce the linear relationship between coherence C and mean firing rate $\mu_{l,r}$ of the left and right selective neurons observed in MT recordings:

$$\mu_{l,r}(C) = r_0 + b_{l,r}C \quad . \tag{2.2.1}$$

Here the parameters r_0, b_l, and b_r approximate recordings from MT [13]. $r_0 = 20$, and if evidence favors the left alternative, $b_l = .4$ and $b_r = -.2$; if the right alternative is favored, these values are exchanged.

Next, the output of each spiking pool was aggregated. Each spike emitted from a neuron in the pool was convolved with an exponential filter with time constant 20 ms, an approximate model of the smoothing effect of synaptic transmission. These smoothed responses were then summed to form a single stochastic process for each pool (see Figure 2.3, right, and [109]).

We then approximated the smoothed output of each spiking pool by a simpler stochastic process that captures the mean, variance, and temporal correlation of this output as a function of dot coherence. We used Gaussian processes $I_l(t)$ and

$I_r(t)$ for the rightward- and leftward-selective pools (See Figure 2.3). Specifically, we chose Ornstein-Uhlenbeck (OU) processes, which are continuous Gaussian process generated by the stochastic differential equations

$$dI_{l,r} = \frac{\mu_{l,r}(C) - I_{l,r}}{\tau}dt + \sqrt{\frac{2\nu_{l,r}(C)}{\tau}}dW_t \tag{2.2.2}$$

with mean $\mu_{l,r}(C)$ as dictated by Equation 2.2.1, and noise contributed by the Wiener process W_t. The variance $\nu_{l,r}(C)$ and timescale τ were chosen to match the steady-state variance and autocorrelation function of the smoothed spiking process. As shown in Results, this timescale affects the speed and accuracy of decisions under robust integration.

Our construction so far accounts for variability in output from left vs. right direction selective neurons. We now incorporate an additional noise source into the output of each pool. These noise terms ($\eta_l(t)$ and $\eta_r(t)$, respectively) could represent, for example, neurons added to each pool that are nonselective to direction or intrinsic variability in the integrating circuit. Each noise source is modeled as an independent OU process with mean 0, timescale 20 ms as above, and a strength (variance) $\nu_\gamma/2$. This noise strength is a free parameter that we vary to match behavioral data (see "A robust integrator circuit" and Figure 2.16). We note that previous studies [102, 70, 23] also found that performance based on the direction-sensitive cells alone can be more accurate than behavior, and therefore incorporated variability in addition to the output of "left" and "right" direction selective MT cells.

Finally, the signals that are accumulated by the left and right neural integrators are constructed by differencing the outputs of the two neural pools:

$$\begin{aligned} \Delta I_l(t) &= [I_l(t) + \eta_l(t)] - [I_r(t) + \eta_r(t)] \\ \Delta I_r(t) &= -\Delta I_l(t) \ . \end{aligned} \tag{2.2.3}$$

2.2.3 Neural integrator circuit and feedback mistuning

A central focus of our paper is variability in the relative tuning of recurrent feedback vs. decay in an integrator circuit. Below, we will introduce the *mistuning*

parameter β, which determines the extent to which feedback and decay fail to perfectly balance. We first define the dynamics of the integrator circuit on which our studies are based. This is described by the firing rates $E_{l,r}(t)$ of integrators that receive outputs from left-selective or right-selective pools $\Delta I_{l,r}(t)$ respectively. The firing rates $E_{l,r}(t)$ increase as evidence for the corresponding task alternative is accumulated over time:

$$\tau_E \frac{dE_{l,r}}{dt} = -E_{l,r} + (1+\beta)E_{l,r} + \kappa \Delta I_{l,r}(t). \tag{2.2.4}$$

The three terms in this equation account for leak, feedback excitation, and the sensory input (scaled by a weight $\kappa = 1/9$), with time constant $\tau_E = 20$ ms. When the mistuning parameter $\beta = 0$, leak and self-excitation exactly cancel, and hence the integrator is *perfectly tuned*. An integrator with $\beta \neq 0$ is said to be *mistuned*, with either exponential growth or decay of activity (in the absence of input). Imprecise feedback tuning is modeled by randomly setting β to different values from trial to trial (but constant during a given trial), with a mean value $\bar{\beta}$ and a precision given by a standard deviation σ_β. We assume that $\bar{\beta} = 0$ for most of the study. Thus the spread of β, which we take to be Gaussian, represents the intrinsic variability in the balance between circuit-level feedback and decay. Perfect tuning corresponds to $\sigma_\beta = \bar{\beta} = 0$, while $\sigma_\beta \neq 0$ or $\bar{\beta} \neq 0$ corresponds to a mistuned integrator. Finally, we set initial activity in the integrators to zero ($E_{l,r}(0) = 0$), and impose reflecting boundaries at $E_r = 0$, $E_l = 0$ (as in, e.g., [111]) so that firing rates never become negative.

2.2.4 A robust integrator circuit

One method to construct a robust integrator with the properties described in Figure 2.1 is from a series of bistable subpopulations, which sequentially activate in order to represent accumulated evidence (for other approaches see [56, 81, 45]). The many equations that describe the evolution of these systems can be closely approximated with reduced models, as demonstrated in [45]. We derived a single piecewise-defined differential equation model that approximates the dynamics of a robust integrator constructed from bistable pools.

We begin with a series of N subpopulations with the firing rate of the i^{th} population r_i defined by:

$$\tau_E \frac{dr_i}{dt} = -r_i + r^- + (r^+ - r^-)H\left[p * r_i + q(1+\beta)\sum_{i\neq j}^{N} r_j + a\Delta I_{l,r} - b_i\right]$$

$$\hat{E} = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{2.2.5}$$

In this equation, r^- and r^+ define the minimum and maximum firing rates of the pools, p, q, and a are local, global, and input coupling strengths, b_i is an individualized bias term that allows sequential activation of the subunits, and $\Delta I_{l,r}$ represents the input signal to be integrated (H represents unit-step function, or some other suitable sigmoidal transfer function). Importantly β provides a fractional mistuning of global excitation.

The average dynamics $\hat{E}$ of the pool can be approximated by a piecewise-defined ODE model E that reduces the system of N equations to a single global equation (see Figure 2.19, and also [45]):

$$\tau_E \frac{dE_{l,r}}{dt} \approx \begin{cases} 0 & \left|\beta E_{l,r} + \frac{a\Delta I_{l,r}}{Nq} - \beta\frac{(r^+ + r^-)}{2N}\right| < \frac{(r^+ - r^-)(p-\beta q)}{2Nq} \\ \beta E_{l,r} + \frac{a\Delta I_{l,r}}{Nq} - \beta\frac{(r^+ + r^-)}{2N} & \text{otherwise} \end{cases} \tag{2.2.6}$$

We simplify this model by relabeling $\kappa = \frac{a}{Nq}$ and $R = \frac{(r^+ - r^-)p}{2a}$; if a and p are increased as the number of subpopulations N becomes large, the remaining terms $\beta\frac{(r^+ \pm r^-)}{2N}$ will vanish. Thus, the model further simplifies to:

$$\tau_E \frac{dE_{l,r}}{dt} = \begin{cases} 0 & |\beta E_{l,r} + \kappa\Delta I_{l,r}| \leq \kappa R \\ \beta E_{l,r} + \kappa\Delta I_{l,r} & \text{otherwise} \end{cases} \tag{2.2.7}$$

All subsequent results are based on this simplified model, which captures the essence of the robust integration computation. The first line is analogous to the series of potential wells depicted in Figure 2.1: if the sum of the mistuned integrator feedback and the input falls below the robustness limit R, the activity of the integrator remains fixed. If this summed input exceeds R, the activity evolves as

for the non-robust integrator in Equation 2.2.4. To interpret the robustness limit R, it is convenient to normalize by the standard deviation of the input signal:

$$\hat{R} = \frac{R}{\sqrt{Var\left[\Delta I_{l,r}(t)\right]}}. \tag{2.2.8}$$

In this way, $\hat{R}$ can be interpreted in units of standard deviations of input OU process that are “ignored" by the integrator.

Figure 2.4 compares the integrators defined by Equations 3.2.1 and 2.2.7 at three different values of the normalized robustness limit $\hat{R}$, in response to a fixed realization of $\Delta I(t)$ for comparison. As $\hat{R}$ increases, the extent to which the effective model tracks the full model decreases. The quality of the reduction can be quantified by examining the relative error between the full and effective models:

$$\epsilon_t = \frac{\hat{E}(t) - E(t)}{\hat{E}(t)} \tag{2.2.9}$$

Histograms of ϵ_t evaluated at $t = 500$ ms are included as insets. The average agreement of the two models is within roughly 20% across a range of robustness values $\hat{R}$. This agreement on the basis of individual trials is sufficient for the purpose of demonstrating the connection between the simplified integrator model we analyze, and one of its many possible neural substrates. A performance comparison between these models in included in Figure 2.19.

To summarize, Equation 2.2.7 defines a parameterized family of neural integrators, distinguished by the robustness limit $\hat{R}$. As $\hat{R} \to 0$, the model reduces to Equation 2.2.4. When additionally $\beta = 0$, the (perfectly tuned) integrator computes an exact integral of its input: Equation 2.2.7 then yields $E_{l,r}(t) \propto \int_0^t \Delta I_{l,r}(t')dt'$. We analyze this robust integrator model below.

2.2.5 Computational methods

Monte Carlo simulations of Equation 2.2.7 were performed using the Euler-Maruyama method [50], with dt= 0.1 ms. For a fixed choice of input statistics and threshold θ, a minimum of $10,000$ trials were simulated to estimate accuracy and RT values.

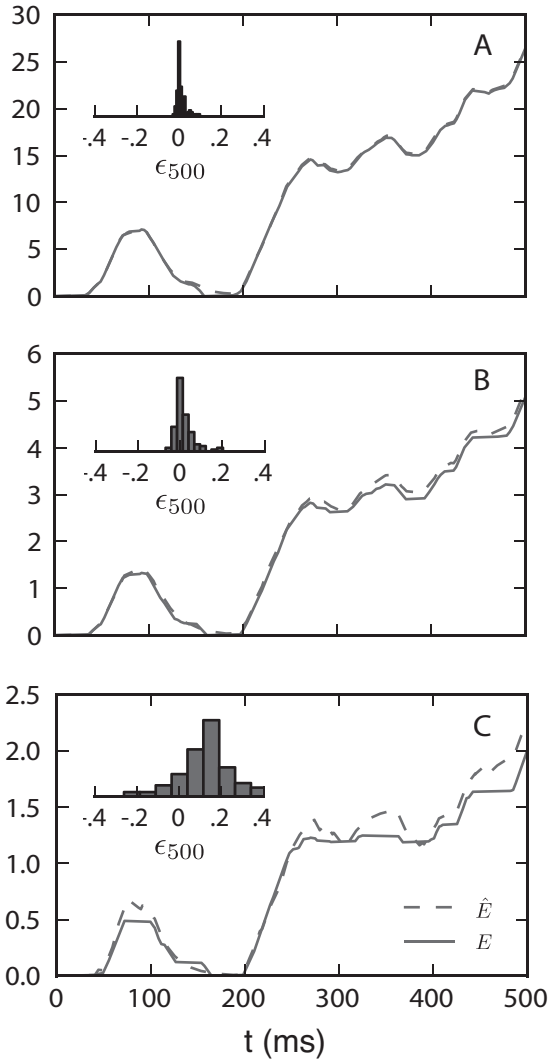


Figure 2.4: Comparison of integration by Equation 3.2.1 and 2.2.7, $\hat{R} = .1$ (Panel A), $\hat{R} = .5$ (Panel B), and $\hat{R} = 11$ (Panel C). Histograms of the relative error between the values of $E(t)$ and $\hat{E}(t)$ at $t = 500$ ms are plotted as insets (See Equation 2.2.9). The mean and standard deviation (μ,σ) for each distribution are, respectively, (.0097, .022) (.0183, .0453), and (.1342, .1358). At these levels of $\hat{R}$, E approximates $\hat{E}$ low to moderate error.

In simulations where $\sigma_\beta > 0$, results were generated across a range of β values and then weighted according to a normal distribution. The range of values was chosen with no less than 19 linearly spaced points, across a range of $\pm$ 3 standard deviations around the mean $\bar{\beta}$. Simulations were performed on NSF Teragrid clusters, and the UW Hyak cluster.

Reward rate values presented in "Reward rate and the robustness-sensitivity tradeoff" are presented as maximized by varying the free parameter θ; values were computed by simulating across a range of θ values. The range and spacing of these values were chosen dependent on the values of $\hat{R}$ and β for the simulation; the range was adjusted to capture the relative maximum of reward rate as a function of θ, while the spacing was adjusted to find the optimal θ value with a resolution of $\pm$ 0.1. Values of θ and ν_γ in the table included in Figure 2.16 were chosen to best match accuracy and chronometric functions to behavioral data reported in [93]. This was accomplished by minimizing the sum-squared error in data vs. model accuracy and chronometric curves across a discrete grid of θ and ν_γ values, with a resolution of 0.1.

Autocovariance functions of integrator input presented in "Analysis: robust integrators and decision performance" were computed by simulating an Ornstein-Uhlenbeck process using the exact numerical technique in [41] with $dt = 0.1$ ms, to obtain a total of 2^{27} sample values. Sample values of the process less than the specified robustness limit $\hat{R}$ were set to 0, and the autocovariance function was computed using standard Fourier transform techniques.

2.3 Results

2.3.1 Do robustness and mistuning affect decision performance?

In the Methods we define a general neural integrator model (Equation 2.2.7) that accumulates signals representing the output of motion sensitive neurons (Equation 2.2.3). The integrator model includes two key parameters. The first is σ_β, which represents standard deviation of β from the ideal value $\bar{\beta} = 0$. The second is the robustness limit $\hat{R}$. We emphasize dual effects of $\hat{R}$: as $\hat{R}$ increases, the

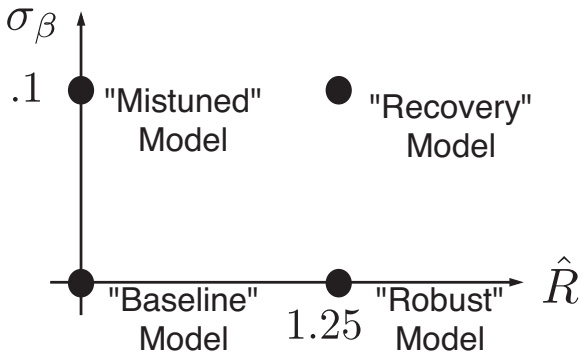


Figure 2.5: Parameter space view of four integrator models, with different values of the robustness limit $\hat{R}$ and feedback mistuning variability σ_β. The impact of transitioning from one model to another by changing parameters is either to enhance or diminish performance, or to have a neutral effect (see text).

integrator becomes able to produce a range of graded persistent activity for ever-increasing levels of mistuning (see Figure 2.1A, where $\hat{R}$ corresponds to the depth of energy wells). This prevents runaway increase or decay of activity when integrators are mistuned; intuitively, this might lead to better performance on sensory accumulation tasks. At the same time, as $\hat{R}$ increases, a larger proportion of the evidence fails to affect the integrator (see Figure 2.1B, where $\hat{R}$ specifies a limit within which inputs are "ignored"). Such sensitivity loss should lead to worse performance. This implies a fundamental tradeoff between competing effects: (i) one would prefer to integrate all relevant input, favoring small $\hat{R}$, and (ii) one would prefer an integrator robust to mistuning (e.g., $\sigma_\beta > 0$), favoring large $\hat{R}$. Thus it makes sense to assess the effect of robustness under different degrees of mistuning, as represented schematically in Figure 2.5.

To assess this performance, we consider relationships between decision speed and accuracy in both controlled duration and reaction time tasks. In the controlled duration task, we simply vary the stimulus presentation duration, and plot accuracy vs. experimenter-controlled stimulus duration. In the reaction time task, we vary the decision threshold θ — treated as a free parameter — over a range of values, thus tracing out the parametric curve for all possible pairs of speed and accuracy values. Here, speed is measured by reaction time (RT, see Materials and Methods). For both cases, we use a single representative dot coherence (C=12.8

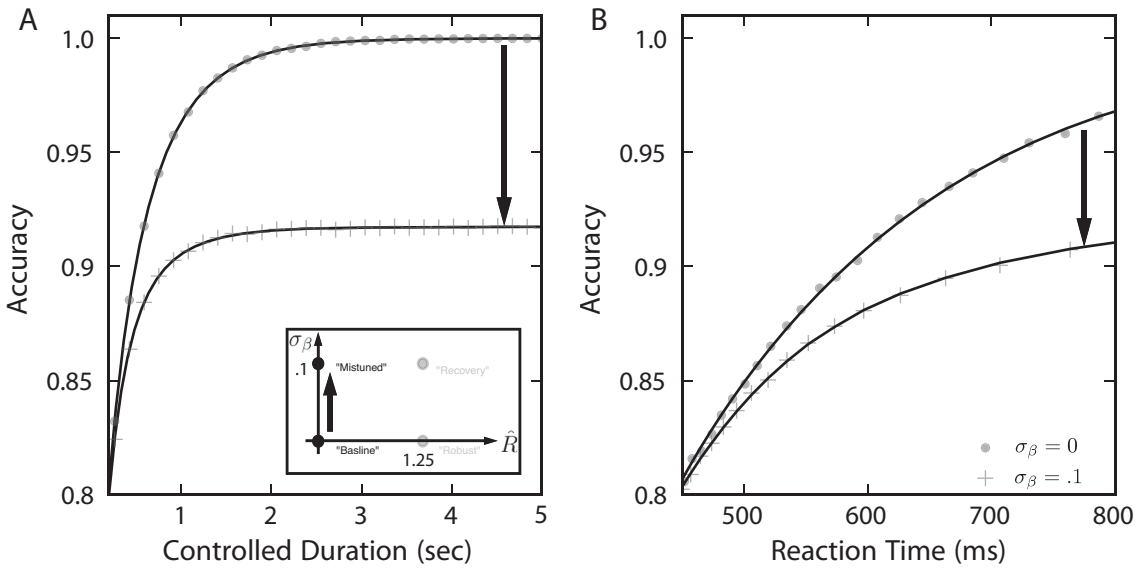


Figure 2.6: Mistuned feedback diminishes decision performance. (Inset) The plots depict a move in parameter space from the "baseline" model to the "mistuned" model by changing $\sigma_\beta = 0 \rightarrow 0.1$. In this and subsequent plots, simulation results are given with markers; lines are rational polynomial fits. (A) In the controlled duration task, accuracy is lower for the "mistuned" model than for the "baseline" model at every trial duration, indicating a loss of performance when σ_β increases. (B) In the reaction time task, we parametrically plot all (RT, accuracy) pairs attained by varying the decision threshold θ. Once again, accuracy is diminished by mistuning.

in Equation 2.2.1); similar results were obtained using other values for motion strength (see Figure 2.19).

We first study a case we call the "baseline" model (Figure 2.5), for which there is no mistuning or robustness: $\sigma_\beta = \hat{R} = 0$. Speed accuracy plots for this model are shown as filled dots in Figures 2.6A and B, for the controlled duration and reaction time tasks respectively. We compare the "baseline" model with the "mistuned" model, indicated by crosses, for which the feedback parameter has a standard deviation of $\sigma_\beta = 0.1$ (i.e., 10% of the mean feedback) and robustness $\hat{R} = 0$ remains unchanged. In the controlled duration task (Figure 2.6A) we observe that mistuning diminishes accuracy by as much as 10%, and this effect is sustained even for arbitrarily long viewing windows [121, 9]. The RT task (Figure 2.6B) produces a similar effect: for a fixed RT, the corresponding accuracy is decreased.

While maintaining feedback mistuning, we next increase the robustness limit

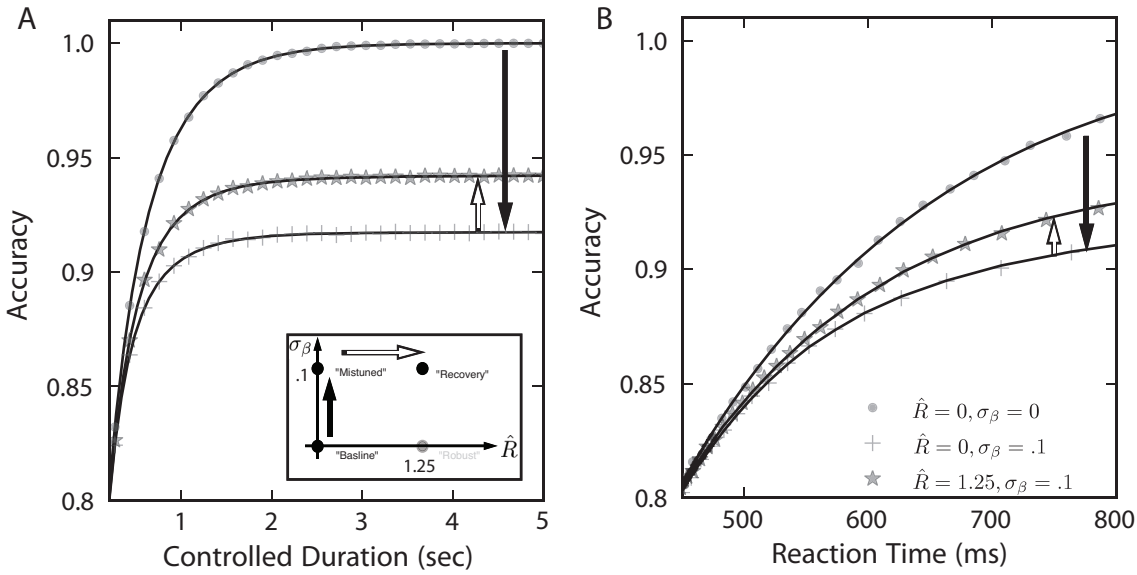


Figure 2.7: Increasing the robustness limit $\hat{R}$ helps recover performance lost due to feedback mistuning. Results of simulation are plotted with fitting lines. (Inset) We illustrate this by moving in parameter space from the "mistuned" model to the "recovery" model, by changing $\hat{R} = 0 \rightarrow 1.25$. The impact on decision performance is shown for both the controlled duration (A) and reaction time (B) tasks. We find that $\hat{R} > 0$ yields a performance gain for the "recovery" model in comparison with the "mistuned" model.

to $\hat{R} = 1.25$ (so that approximately 75% of the input stream is "ignored" by the integrators). We call this case the "recovery" model because robustness compensates in part for the performance loss due to feedback mistuning: the speed accuracy plots in Figure 2.7 for the recovery case, indicated by stars, lie above those for the "mistuned" model. For example, at the longer controlled task durations (Figure 2.7A) and reaction times (Figure 2.7B) plotted, 30% of the accuracy lost due to integrator mistuning is recovered via the robustness limit $\hat{R} = 1.25$.

Finally, we study the remaining possibility, when the robustness parameter $\hat{R}$ is increased from zero in a perfectly tuned integrator ($\sigma_\beta = 0$); this is the "robust" case in Figure 2.5. We expected performance to be substantially diminished as a consequence of lost sensitivity to inputs. However, Figure 2.8 demonstrates that this is not the case: speed accuracy curves for $\hat{R} = 1.25$ almost coincide with those for the "baseline" case of $\hat{R} = 0$. We note that since $\hat{R}$ measures ignored input in units of the standard deviation, the integrator circuit disregards the weakest 75%

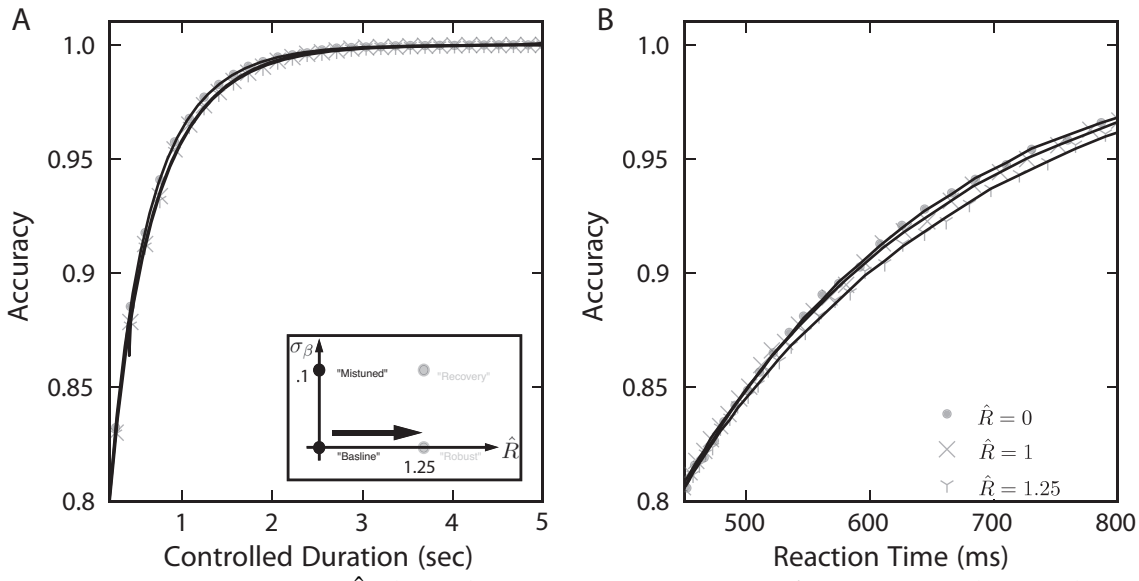


Figure 2.8: Increasing $\hat{R}$ alone does not compromise performance. Only simulation results, without fitting lines, are plotted for clarity. (Inset) We illustrate this by moving in parameter space directly from the "baseline" to the "robust" model. For both (A) controlled duration and (B) reaction time tasks, we plot the relationship between speed and accuracy. Circles give results for the "baseline" model, and 'x' and 'y' markers for the robust model at $\hat{R} = 1$ and 1.25 respectively. These curves are very similar in the "baseline" case, indicating little change in decision performance due to the robustness limit $\hat{R} = 1.25$.

of the input stimulus. Given this large amount of ignored stimulus, the fact that the robust integrator produces nearly the same accuracy and speed as the "baseline" case is surprising. This implies that the "robust" model can protect against feedback mistuning, without substantially sacrificing performance when feedback is perfectly tuned. An even larger performance recovery (approaching 75%) can be achieved by the robust integrator model described in Equation 3.2.1 (See Figure 2.19).

To summarize, the ratchet-like mechanism of the robust integrator appears well-suited to the decision tasks at hand. This mechanism counteracts some of the performance lost when feedback is mistuned. Moreover, even without mistuning, a robust integrator still performs as well as the "baseline" case that integrates all information in the input signal. In the next section, we begin to explain this observation by constructing several simplified models and employing results from statistical decision making theory.

2.3.2 Analysis: robust integrators and decision performance

Controlled duration task: Discrete time analysis

We can begin to understand the effect of the robustness limit on decision performance by formulating a simplified version of the evidence accumulation process. We focus first on the controlled duration task, where the analysis is somewhat simpler.

Our first simplification is to consider a single accumulator E which receives evidence for or against a task alternative in discrete time. The value of E on the i^{th} time step, E_i, is allowed to be either positive or negative, corresponding to accumulated evidence favoring the leftward or rightward alternatives, respectively. On each time step, E_i increments by an independent, random value Z_i with a probability density function (PDF) $f_Z(Z)$. We first describe an analog of the "baseline" model above (i.e., in the absence of robustness, $\hat{R} = 0$). Here, we take the increments Z_i to be independent, identical, and Gaussian distributed, with a mean $\mu > 0$ (establishing the preferred alternative) and standard deviation σ: that is, $Z_i \sim N\left(\mu, \sigma^2\right)$.

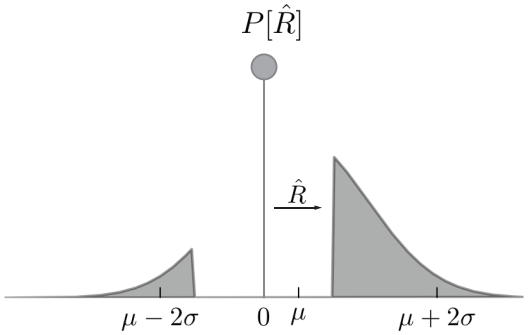


Figure 2.9: $\hat{R}$ affects the discrete time increment distribution. The PDF of the random variable $Z_{\hat{R}}$, with probability mass for values between the robustness limit $\hat{R}$ re-allocated as a delta function centered at zero (Here $\hat{R} = 1$).

After the n^{th} step, we have

$$E_n = \sum_{i=1}^{n} Z_i \ .$$

In the controlled duration task, a decision is rendered after a fixed number of time steps N, (i.e. $n = N$) and a correct decision (i.e., in favor of the preferred alternative) occurs when $E_N > 0$. By construction, $E_n \sim N\left(n\mu, n\sigma^2\right)$, which implies that accuracy can be computed as a function of the signal-to-noise ratio (SNR) $s = \frac{\mu}{\sigma}$ of a sample:

$$\text{Accuracy} = \int_0^\infty \frac{1}{\sqrt{2\pi N\sigma^2}} e^{-\frac{(x-N\mu)^2}{2N\sigma^2}} dx = \frac{1 + \text{Erf}\left(\sqrt{\frac{N}{2}}s\right)}{2} \ . \tag{2.3.1}$$

Next, we change the distribution of the accumulated increments Z_i to construct a discrete time analog of the robust integrator. Specifically, increasing the robustness parameter to $R > 0$ affects increments Z_i by redefining the PDF $f_Z(Z)$ so that weak samples do not add to the total accumulated "evidence", precisely as in Equation 2.2.7. (Models where such a central "region of uncertainty" of the sampling distribution is ignored were previously studied in a race-to-bound model [112]; see Discussion). This requires reallocating probability mass below the robustness

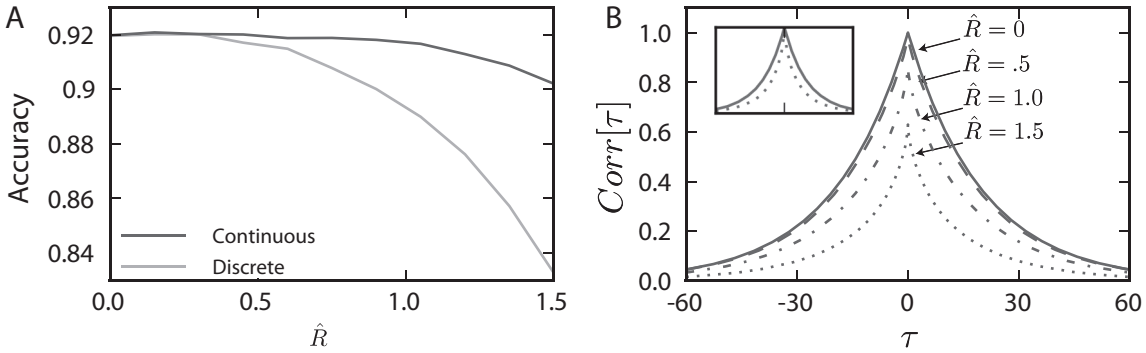


Figure 2.10: Accuracy of the discrete and continuous time models for the controlled duration task, $T = 500$ ms. (A) Performance of the continuous time model (Equation 2.3.10) is plotted as a black curve, and that predicted by the discrete time model with identical signal increments is plotted as a gray curve. (B) The disparity between the continuous-time and discrete-time model performance can be partially understood by observing the uncorrelating effect of $\hat{R}$ on the autocorrelation function for evidence stream in the continuous-time model (Inset) Two of these same autocovariance functions (for $\hat{R} = 0$ and $\hat{R} = 1.5$) are plotted normalized to their peak value.

limit to a weighted delta function at zero (Figure 2.9A). Specifically:

$$f_{Z_R}(Z) = \delta(Z) \int_{-R}^{R} f_Z(Z')\, dZ' + \begin{cases} 0 & |Z| < R \\ f_Z(Z) & \text{otherwise} \end{cases} \tag{2.3.2}$$

To compute accuracy with robustness $\hat{R} > 0$, we sum N random increments from this distribution forming the cumulative sum $E_{N_{\hat{R}}}$. As above, a correct decision occurs on trials where $E_{N_{\hat{R}}} > 0$. As shown in Figure 2.10, the replacement of increments with zeros has negligible effect on accuracy when $\hat{R}$ is less than $\sim .5$. For larger values of $\hat{R}$, accuracy diminishes faster for the discrete model (gray curve) in comparison to the continuous model (black curve). Loss of accuracy is expected for both models as robustness effectively prevents stimulus information from affecting the decision. However, the continuous model suffers less than the discrete approximation, owing to the one important difference: the presence of temporal correlations in the evidence stream. We will return to this matter below. First, we give an explanation of the negligible effect values of $\hat{R}$ on decision accuracy for either model.

The central limit theorem allows us to approximate the new cumulative sum $E_{N_{\hat{R}}}$ as a normal distribution (for sufficiently large N), with μ and σ in Equation 2.3.1 replaced by the mean and standard deviation of the PDF defined by Equation 2.3.2. As before, we normalize R by the standard deviation of the increment, $\hat{R} = \frac{R}{\sigma}$, and then express the fraction correct $\text{Accuracy}_{\hat{R}}$ as a function of $\hat{R}$ and s. One can think of $\hat{R}$ as perturbing the original accuracy function given in Equation 2.3.1. Although this perturbation has a complicated form, we can understand its behavior by observing that its Taylor expansion does not have first or second-order contributions in $\hat{R}$:

$$\text{Accuracy}_{\hat{R}}(N) = \text{Accuracy}(N) - \frac{\sqrt{N}s\left(1+2s^2\right)e^{-\frac{(1+N)s^2}{2}}}{6\pi}\hat{R}^3 + O(\hat{R}^5). \quad (2.3.3)$$

Thus, for small values of $\hat{R}$ (giving very small $\hat{R}^3$), there will be little impact on accuracy. Equation 2.3.3 can therefore *partially* explain the key observation in Fig 2.8A that $\hat{R}$ can be substantially increased while incurring very little performance loss.

Now we return to the comparison between the discrete and continuous time models, by setting the signal-to-noise ratio of the sampling distribution identical to the steady state distribution of the input signal to the neural integrator model (See Sensory Input). The interval between samples is set to match accuracy performance of the continuous time model at $\hat{R} = 0$. The gray line in Figure 2.10A show the accuracy for the discrete time model, as the robustness limit $\hat{R}$ is increased. The discrete time model predicts a decrease in accuracy at $\hat{R} = .5$ that is not seen in the performance of the full continuous time model. In the next section, we explain how this discrepancy can be resolved with a more complete model that accounts for the temporal structure of the continuous time signal.

Controlled duration task: Continuous time analysis

We next extend the analysis of the controlled duration task in the previous section to signal integration in continuous time. We follow the method developed in [41] to describe the mean and variance of the integral of a continuous input signal, while these quantities increase over time. The challenge here lies in the temporal

correlations in the Gaussian OU input signal (see Methods, Sensory Input). As in the previous section, we describe the distribution of the integrated signal at the final time T, which determines accuracy in the controlled duration task.

We first replace the discrete input samples Z_i from the previous section with a continuous signal $Z(t)$, which we take to be a Gaussian process with a correlation timescale derived from our model sensory neurons (see Methods). We define the integrated process

$$\frac{dE}{dt} = Z(t) \quad \rightarrow \quad E(t) = \int_0^t Z(t')dt' \tag{2.3.4}$$

with initial condition $E(0) = 0$.

Assuming that $Z(t)$ satisfies certain technical conditions that are easily verified for the OU process (wide-sense stationarity, α-stability, and continuity of sample paths [39, 5, 41]), we can construct differential equations for the first and second moments $\langle E(t) \rangle$ and $\langle E^2(t) \rangle$ evolving in time. We start by taking averages on both sides of our definition of $E(t)$, and, noting that $E(0) = 0$, compute the time-varying mean:

$$\frac{d \langle E(t) \rangle}{dt} = \langle Z(t) \rangle \implies \langle E(t) \rangle = t \langle Z(t) \rangle \quad . \tag{2.3.5}$$

Similarly, we can derive a differential equation for the second moment of $E(t)$:

$$\frac{d \langle E^2(t) \rangle}{dt} = 2 \langle Z(t)E(t) \rangle \, . \tag{2.3.6}$$

The righthand side of this equation can be related to the area under the autocovariance function $A(\tau) \equiv \langle Z(t)Z(t+\tau) \rangle - \langle Z(t) \rangle^2$ of the process $Z(t)$:

$$\begin{aligned} \langle Z(t)E(t) \rangle = \left\langle Z(t) \int_0^t Z(s)ds \right\rangle &= \int_0^t \langle Z(t)Z(s) \rangle \, ds \\ &= \int_0^t \langle Z(t)Z(t-\tau) \rangle \, d\tau \\ &= \int_0^t A(\tau) + \langle Z(t) \rangle^2 \, d\tau \end{aligned} \tag{2.3.7}$$

We now have an expression for how the second moment evolves in time. We

can simplify the result via integration by parts:

$$\begin{aligned}\left\langle E^2(t)\right\rangle &= 2\int_0^t\int_0^s A(\tau)+\langle Z(t)\rangle^2\,d\tau ds = 2\int_0^t (t-\tau)A(\tau)\,d\tau + t^2\langle Z(t)\rangle^2 \\ &\implies \mathrm{Var}[E(t)] = 2\int_0^t (t-\tau)A(\tau)\,d\tau. \end{aligned} \tag{2.3.8}$$

Because $E(t)$ is an accumulation of Gaussian random samples $Z(t)$, it will also be normally distributed, and hence fully described by the mean (Equation 2.3.5) and variance (Equation 2.3.8) [5].

To model a non-robust integrator, we take $Z(t)$ to be a OU process with steady-state mean and variance μ and σ^2, and time constant τ. For the robust case, we can follow Equation 2.2.7 and parameterize a family of processes $Z_{\hat{R}}(t)$ with momentary values below the robustness limit $\hat{R}$ set to zero. (Here, we again normalize the robustness limit by the standard deviation of the OU process.) We numerically compute the autocovariance functions $A_{\hat{R}}(\tau)$ of these processes, and use the result to compute the required mean and variance, and hence time-dependent signal-to-noise ratio SNR(t), for the integrated process $E(t)$. This yields

$$SNR_{\hat{R}}(t) = \frac{\langle E(t)\rangle}{\sqrt{Var[E(t)]}} = \frac{tE[Z_{\hat{R}}(t)]}{\sqrt{2\int_0^t (t-\tau)A_{\hat{R}}(\tau)\,d\tau}}\ . \tag{2.3.9}$$

Under the assumption that $E(T)$ is approximately Gaussian for sufficiently long T (which can be verified numerically), we use this SNR to compute decision accuracy at T:

$$Accuracy_{\hat{R}}(T) \approx \frac{1+\mathrm{Erf}\left(\frac{1}{\sqrt{2}}SNR_{\hat{R}}(T)\right)}{2}\ . \tag{2.3.10}$$

This function is plotted for $T = 500$ ms as the black line in Figure 2.10A. The plot shows that accuracy remains relatively constant until the robustness limit $\hat{R}$ exceeds ≈ 1.25, a longer range of $\hat{R}$ values than for the discrete time case (compare gray curve vs. black curve in Figure 2.10A).

Why does the robustness limit appear to have a milder effect on degrading decision accuracy for our continuous vs. discrete time input signals? We can get some insight into the answer by examining the autocovariance functions $A_{\hat{R}}(\tau)$, which

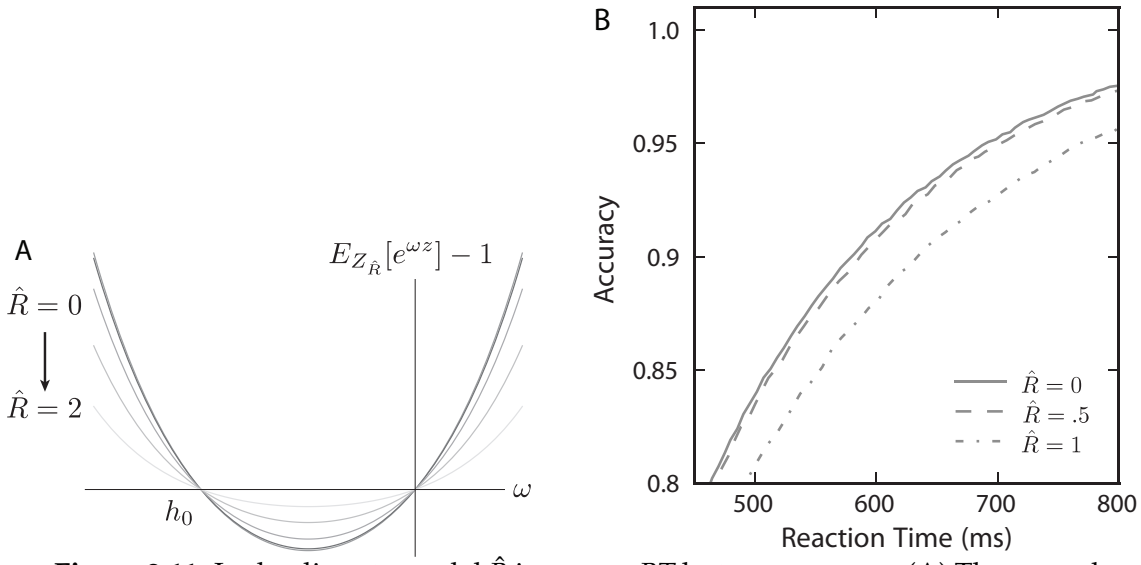


Figure 2.11: In the discrete model $\hat{R}$ increases RT but not accuracy. (A) The second real root h_0 of $M_{Z_{\hat{R}}}(s)$ remains unchanged as $\hat{R}$ increases from $0 \rightarrow 2$. (Lines are uniformly distributed in this range.) This implies that in the reaction time task, no changes in the accuracy will be observed (See Equation 5.3.1). (B) However, the speed accuracy tradeoff will be affected, once $E[Z_R]$ begins to diminish (See Equation 5.3.2). This performance loss begins for $\hat{R} > .5$, in contrast to the performance of the continuous time model (see Figure 2.8B).

we present in Figure 2.10B. When normalized by their peak value, the autocovariance for $\hat{R} > 0.5$ falls off more quickly vs. the time lag τ (see inset in Figure 2.10B), indicating that subsequent samples become less correlated in time. Thus, there are effectively more "independent" samples that are drawn over a given time range T, improving the fidelity of the signal and hence decision accuracy. This effect is not present in our discrete time model.

Summary: Our analysis of decision performance for the controlled duration task shows that two factors contribute to the preservation of decision performance for robust integrators. The first is that the momentary SNR of the inputs is barely changed for robustness limits up to $\hat{R} \approx 0.5$. The second is that, as $\hat{R}$ increases, the signal $Z_{\hat{R}}(t)$ being integrated becomes less correlated in time. This means that (roughly) more independent samples will arrive over a given time period.

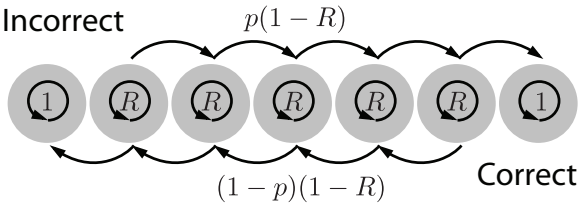


Figure 2.12: Biased random walk between two absorbing boundaries. The **Incorrect** and **Correct** states act as "sinks" of the discrete time discrete space Markov chain. The intermediate states are analogous to the potential wells in the robust integrator model. Here, a particle will remain in the current state at the next time step with probability R. The final probability of ending up in either sink is independent of R

Reaction time task: Discrete analysis

We begin our analysis of the reaction time task by introducing a discrete time, discrete space random walk model. In this model, schematized in Figure 2.12 with five intermediate states, a particle representing the accumulated value E starts at a state balanced between two absorbing "sink" states. At every time step, the particle moves towards the "correct" (i.e. preferred) sink with probability $p(1-R)$, and towards the "incorrect" (null) sink with probability $(1-p)(1-R)$ (we consider $p > 0.5$, biasing the random walk toward the "correct" sink). There is also the possibility that the particle might remain in the current state, with probability R.

We now draw an analogy between the states in this random walk model and the ratcheting dynamics among energy wells in a robust integrator (see Figure 2.1 and Introduction). Here, the position of the particle represents the accumulated evidence for the left vs. right alternatives, and the absorbing states represent crossing of the corresponding decision thresholds. When the robustness limit R is increased, the wells – each of which could represent a bistable neural subpopulation (see Methods) – act to hold the particle in a given state, with a probability set by R.

As R is increased in the random walk model, the probability of transitioning out of a given state similarly decreases. Standard results on Markov chains (see, for example, [53]) provide formulas for the probability that the particle will end in one vs. the other sink, as well as the expected number of time steps until this occurs, based on the transition matrix associated with the random walk. The probability of ending in the "correct" sink corresponds to decision accuracy, and is found at

the middle entry in the solution vector $\boldsymbol{x}$ of the matrix equation

$$(\boldsymbol{I} - \boldsymbol{Q})\,\boldsymbol{x} = (1-R)p\boldsymbol{e_1}\ . \tag{2.3.11}$$

Here $\boldsymbol{Q}$ is a tridiagonal matrix with R on the main diagonal, $p(1-R)$ on the lower diagonal, and $(1-p)(1-R)$ on the upper diagonal; $\boldsymbol{e_1}$ is the canonical basis vector with $e_1^{(1)} = 1$, and all other entries equal to 0. After some factoring, we find a common factor of $(1-R)$ on both sides of the equation; thus the solution to $\boldsymbol{x}$ is independent of R. This implies that the probability of ending up in the correct state is unchanged by increasing R from the non-robust case ($R = 0$). Intuitively this makes sense: if one conditions on the fact that one will leave the current state on the next time step, the probability of moving toward the correct and incorrect states are independent of R.

The same is not true for the expected number of steps necessary to reach a sink (by analogy, the reaction time). This is because the matrix system that yields reaction times is:

$$(\boldsymbol{I} - \boldsymbol{Q})\,\boldsymbol{t} = \boldsymbol{1}. \tag{2.3.12}$$

Here the right-hand side of this equation is the vector of all ones, and therefore no equivalent cancellation can occur. However, we do notice that the reaction time with $R \neq 0$ is just a rescaling of the original reaction time with $R = 0$. Specifically, if t_R is the expected number of steps required to reach an absorbing state, then

$$t_R = t_0 \frac{1}{1-R}\ . \tag{2.3.13}$$

Thus, the only effect of the robustness limit R is to delay arrival at the sinks.

$$P[X_i = x] = \begin{cases} p(1-R) & : x = 1 \\ R & : x = 0 \\ (1-p)(1-R) & : x = -1 \end{cases} \tag{2.3.14}$$

We can give this markov chain the notation of a stochastic process by defining X_i as an incremental displacement of the particle, and Y_i as the number of states that the particle has moved from the initial condition after i time steps. Here Y_i is a

signed quantity, with positive sign indicating displacement in the direction of the "correct" alternative. Y_i is itself a stochastic process indexed by i, such that:

$$Y_n = \sum_{i=1}^{n} X_i \tag{2.3.15}$$

Wald [123] provides a description of how to compute the probability that the particle will arrive in the "correct" state in terms of the roots of the moment generating function (MGF) of X_i, set to unity. The essence of the technique can be interpreted as solving Equation 2.3.11 via Cramer's rule.

$$1 = E[e^{tX_i}] = p(1-R)e^{(-1)t} + Re^{0t} + (1-p)(1-R)e^{t} \tag{2.3.16}$$

$$\iff 0 = p(e^{-t} + (1-p)e^{t} \tag{2.3.17}$$

As can be seen, the roots of this expression in t do not depend on R, the parameter that causes the particle to remain in it's current state. In this way, we see that the accuracy of this decision making model is unaffected by robustness precisely because the roots of the MGF of the increment distribution remain fixed as R increases.

Summary: We have used a simplified random walk model to gain intuition about the effect of the robustness limit in the reaction time task, and to show that adding a robustness limit only affects decision latency, but not accuracy. In the next section, we will derive a similar result for continuous sample distributions.

Reaction time task: Continuous analysis

We return to the continuous sampling distribution introduced in "Controlled duration task: Discrete time analysis", but now in the context of threshold crossing in the reaction time task. The accumulation of these increments toward decision thresholds can be understood as the sequential probability ratio test, where the log-odds for each alternative are summed until a predefined threshold is reached [124, 42, 66, 59]. [123] provides an elegant method of computing decision accuracy and speed (RT). The key quantity is given by the moment generating function (MGF, denoted $M_Z(\omega)$ and defined in Equation 2.3.20) for the samples Z (see [67],

[63], and [29]). Under the assumption that thresholds are crossed with minimal overshoot, we have the following expressions:

$$\text{Accuracy} = \frac{1}{1+e^{\theta h_0}} \tag{2.3.18}$$

$$RT = \frac{\theta}{E[Z]} \tanh\left[-\frac{\theta}{2}h_0\right] \tag{2.3.19}$$

where h_0 is the nontrivial two real root of the equation $M_Z(\omega) = 1$ and θ is the decision threshold.

We first consider the case of a non-robust integrator, for which the samples Z are again normally distributed. In this case, we must solve the following equation to find $\omega = h_0$:

$$M_Z(\omega) = E_Z\left[e^{\omega z}\right] = \int_{-\infty}^{\infty} f_Z(z)\, e^{\omega * z} dz = e^{\frac{\omega^2\sigma^2}{2}+\omega\mu} = 1. \tag{2.3.20}$$

It follows that $\omega = 0$ and $\omega = h_0 = -2\frac{\mu}{\sigma^2}$ provide the two real solutions of this equation. (Wald's Lemma ensures that there are exactly two such real roots, for any sampling distribution meeting easily satisfied technical criteria.)

When the robustness limit $\hat{R} > 0$, we can again compute the two real roots of the associated MGF. Here, we use the increment distribution $f_{Z_R}(Z)$ given by Equation 2.3.2, for which all probability mass within R of 0 is reassigned to 0. Surprisingly, upon plugging this distribution into the expression $M_Z(\omega) = 1$, we find that $\omega = 0, h_0$ continue to provide the two real solutions to this equation *regardless of* R, as depicted in Figure 2.11A.

This observation implies that (i) accuracies (Equation 5.3.1) are unchanged as R is increased, and (ii) reaction times (Equation 5.3.2) only change when $E[Z_R]$ changes. In other words, the integrator can ignore inputs below an arbitrary robustness limit at no cost to accuracy, and a penalty in terms of reaction time will only be observed when $E[Z_R]$ changes appreciably. This result holds for any distribution for which:

$$f_Z(z) = f_Z(-z)e^{-h_0 z}; \tag{2.3.21}$$

it is straightforward to verify that the Gaussian satisfies this property.

How much of an increase in R is necessary to decrease $E[Z_R]$, the key quantity that alone controls performance loss? After substituting $\hat{R} = \frac{R}{\sigma}$, we again find only one term up to fifth order in $\hat{R}$,

$$E[Z_{\hat{R}}] = E[Z] - \mu\sqrt{\frac{2}{9\pi}}e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}\hat{R}^3 + O(\hat{R}^5) + ..., \tag{2.3.22}$$

indicating that modest amounts of robustness lead to only small changes in $E[Z_{\hat{R}}]$ (This is similar to the controlled duration case, where small values of $\hat{R}$ will have little effect on $Accuracy_{\hat{R}}(N)$, cf. Equation 2.3.3). This property, in combination with the constancy of h_0, allows us to reason about the trade off speed versus accuracy under robustness. Under symmetrically bounded drift-diffusion, accuracy is determined by θ and h_0, whereas the mean decision time is determined by θ, h_0, and $E[\hat{Z}]$. Since h_0 is fixed for all values of $\hat{R}$, the predicted effect of robustness is a slowing of the decision time owing to the small change in $E[Z_{\hat{R}}]$ (Equation 2.3.22).

This effect is depicted in Figure 10B. The solid curve is a locus of speed-accuracy combinations achieved by varying θ under no robustness. As in the previous section, the signal-to-noise ratio of the independent sampling distribution is identical to the continuous time model (and performance at $\hat{R} = 0$ is matched to Figure 2.8B by varying the time inter-sampling time, here 37 ms). Performance begins to decrease at $\hat{R} = .5$, and is much lower at $\hat{R} = 1$ than the continuous time model with the autocorrelated evidence streams. Precisely as in the preceding section, robustness serves to decorrelate this input stream, effectively giving more independent samples and preserving performance beyond that predicted by the independent sampling theory.

Summary of analysis: In the preceding three sections we have analyzed the impact of the robustness limit on decision performance. For both the controlled duration and reaction time tasks, we first studied the effect of this limit on the evidence carried by momentary values of sensory inputs. In each task, this effect was more favorable than might have been expected. In the controlled duration case, the signal-to-noise ratio of momentary inputs was preserved for a fairly broad range of R, while in the reaction time task, R affected speed but not accuracy at fixed decision threshold. These results provided a partial explanation for the impact of

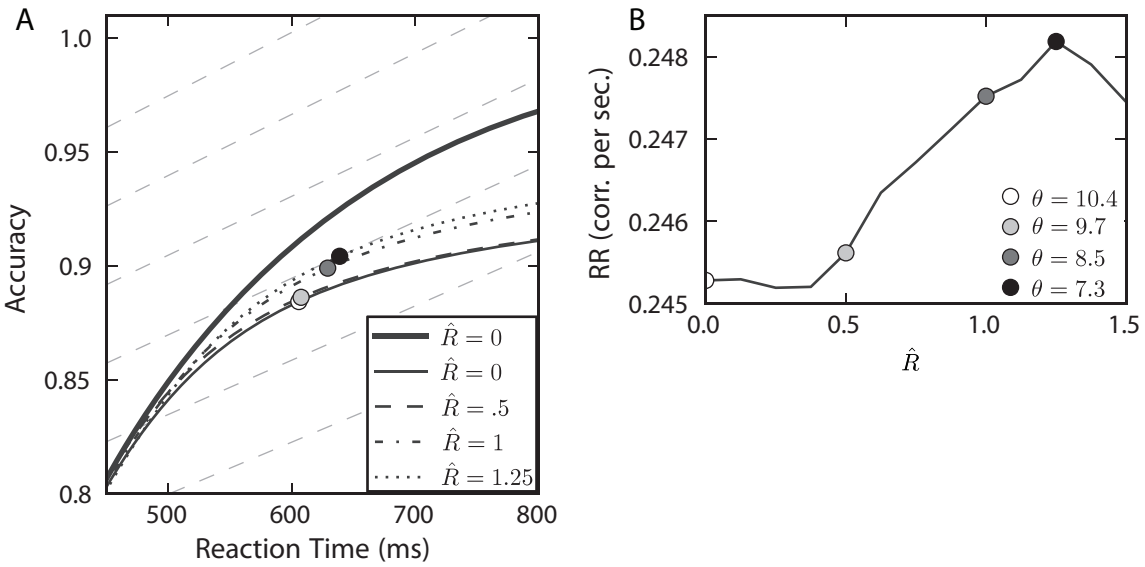


Figure 2.13: Robustness improves reward rate under mistuning. (A) Speed accuracy curves plotted for multiple values of $\hat{R}$; as in previous figures, the greater accuracies found at fixed reaction times indicate that performance improves as $\hat{R}$ increases. The heavy line indicates the "baseline" case of a perfectly tuned, non-robust integrator (repeated from Figure 2.6B). RR level curves are plotted in background (dotted lines; see text), and points along speed accuracy curves that maximize RR are shown as circles. These maximal values of reward rate are plotted in (B), demonstrating the non-monotonic relationship between $\hat{R}$ and the best achievable RR.

robustness on decisions. The rest of the effect was attributed to the fact that the robustness mechanism serves to decorrelate input signals in time, further preserving decision performance by providing the equivalent of more independent evidence samples in a given time window.

2.3.3 Reward rate and the robustness-sensitivity tradeoff

Up to now, we have examined performance in the reaction time task by plotting the full range of attainable speed and accuracy values. The advantage of this approach is that it demonstrates decision performance in a general way. An alternative, more compact approach, is to assume a specific method of combining speed and accuracy into a single performance metric. This approach is useful in quantifying decision performance, and rapidly comparing a wide range of models.

Specifically, we use the reward rate (RR) [9]:

$$RR = \frac{\text{Accuracy}}{\langle RT \rangle + T_{del}}, \tag{2.3.23}$$

the number of correct responses made per unit time, where a delay T_{del} imposed between responses to penalizes rapid guessing. Implicitly, this assumes a motivation on the part of the subject which may not be true; in general, human subjects seldom achieve optimality under this definition as they tend to favor accuracy over speed in two-alternative forced choice trials [132]. Here, we simply use this quantity to formulate a scalar performance metric that provides a clear, compact interpretation of reaction time data.

Figure 2.13A shows accuracy vs. speed curves at 4 levels of $\hat{R}$. The heavy solid line corresponds to the "baseline" model with robustness and mistuning set to zero (see Figure 2.5). The lighter solid line corresponds to the "mistuned" model with $\sigma_\beta = .1$. The remaining broken lines correspond to the "recovery" model for three increasing levels of the robustness limit $\hat{R}$. Also plotted in the background as dashed lines are RR level curves – that is, lines along which RR takes a constant value, with $T_{del} = 3$ sec. On each accuracy vs. speed curve, there exists a RR-maximizing (RT, accuracy) pair. This corresponds to a tangency with one RR level curve, and is plotted as a filled circle. In general, each model achieves maximal RR via a different threshold θ; values are specified in the legend of Panel B. (A general treatment of RR-maximizing thresholds for drift-diffusion models is given in [9].)

In sum, we see that mistuned integrators with a range of increasing robustness limits $\hat{R}$ achieve greater RR, as long as their thresholds are adjusted in concert. The optimal values of RR for a range of robustness limits $\hat{R}$ are plotted in Figure 2.13B. This figure illustrates the fundamental tradeoff between robustness and sensitivity. If there is variability in feedback mistuning ($\sigma_\beta > 0$), increasing $\hat{R}$ can help recover performance. However, beyond a certain point increasing $\hat{R}$ further starts to diminish performance, as too much of the input signal is ignored.

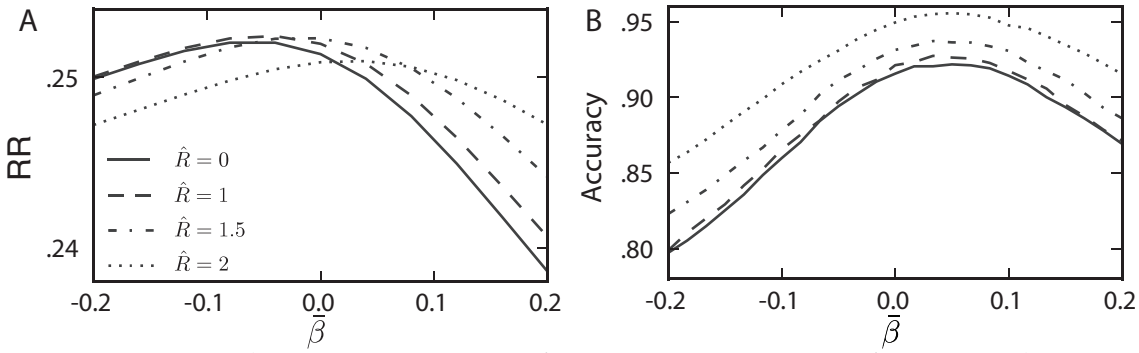


Figure 2.14: Robustness improves performance across a range of mistuning biases $\bar{\beta}$. In both the reaction time (A) and controlled duration (B) tasks, robustness helps improve performance when $\beta \sim N\left(\bar{\beta}, .1^2\right)$, for all values of $\bar{\beta}$ shown. As in previous figures, the coherence of the sensory input is $C = 12.8$. In the reaction time task (A), θ is varied for each value of $\bar{\beta}$ to find the maximal possible reward rate RR, and performance gains are largest for $\bar{\beta} > 0$. In the controlled duration task, substantial gains are possible across the range of $\bar{\beta}$ values.

2.3.4 Biased mistuning towards leak or excitation

We next consider the possibility that variation in mistuning from trial to trial could occur with a systematic bias in favor of either leak or excitation, and ask whether the robustness limit has qualitatively similar effects on decision performance as for the unbiased case studied above. Specifically, we draw the mistuning parameter β from a Gaussian distribution with standard deviation $\sigma_\beta = 0.1$ as above, but with various mean values $\bar{\beta}$ (see Methods). In Figure 2.14A we show reward rates as a function of the bias $\bar{\beta}$, for several different levels of the robustness limit $\hat{R}$. At each value of $\bar{\beta}$, the highest reward rate is achieved for a value of $\hat{R} > 0$; that is, regardless of the mistuning bias, there exists an $\hat{R} > 0$ that will improve performance vs. the non-robust case ($\hat{R} = 0$). We note that this improvement appears minimal for substantially negative mistuning biases (i.e., severe leaky integration), but is significant for the values of $\bar{\beta}$ that yield the highest RR. Finally, the ordering of the curves in Figure 2.14A shows that, for many values of $\bar{\beta}$, this optimal robustness limit is an intermediate value less than $\hat{R} = 2$.

While Figure 2.14 only assesses performance via a particular performance rule (RR, $T_{del} = 3$ sec.), the analysis in "Reward rate and the robustness-sensitivity tradeoff" suggests that the result will hold for other performance metrics as well.

Moreover, Figure 2.14B demonstrates the analogous effect for the controlled duration task: for each mistuning bias $\bar{\beta}$, decision accuracy increases over the range of robustness limits shown.

2.3.5 Bounded integration as a model of the fixed duration task

We have demonstrated that increasing the robustness limit $\hat{R}$ can improve performance for mistuned integrators, in both the reaction time and controlled duration tasks. In the latter, a decision was made by examining which integrator had accumulated more evidence at the end of the time interval. In contrast, [54] argue that decisions in the controlled duration task may actually be made with a decision threshold, much like the reaction time task. That is, evidence accumulates until an absorbing bound is reached, causing the subject to ignore any further evidence and simply wait until the end of the trial to report the decision.

Figure 2.15 demonstrates that our observations about how the robustness limit can recover performance lost to mistuned feedback carry over to this model of decision making as well. Specifically, Panel 2.15A shows how setting $\hat{R} > 0$ improves performance in a mistuned integrator. In fact, more of the lost performance is recovered than in the previous model of the controlled duration task (cf. Figure 2.7A). Panel 2.15B extends this result to show that some value of $\hat{R} > 0$ will recover lost performance over a wide range of mistuning biases $\bar{\beta}$ (cf. Figure 2.14B).

2.3.6 Compatibility of robust integration with behavioral data

We have demonstrated that robustness serves to protect an integrator against the hazards of run away excitation and leak, and that the cost of doing so is surprisingly small. Yet it is hard to know whether the range of effects is compatible with known physiology and behavior. Without better knowledge of the actual neural mechanisms that support integration and decision making, it is not possible to reconcile the parameters of our analysis with physiology. However, we can at least test whether or not the cost of robustness on performance is compatible with behavior.

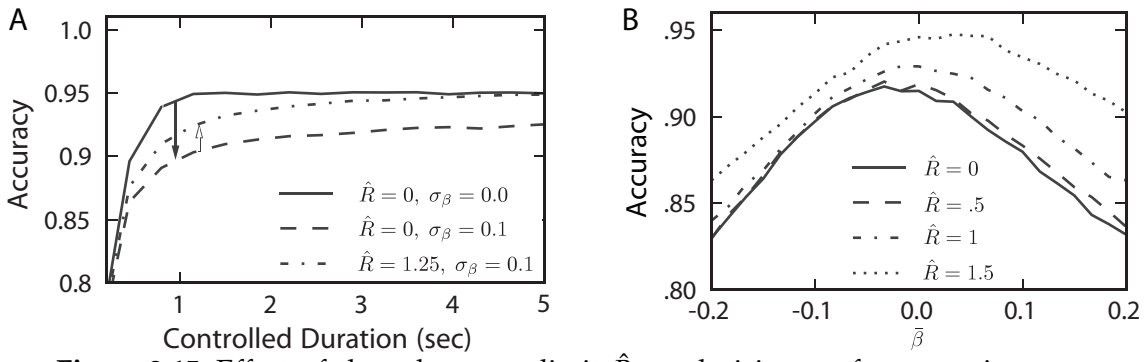
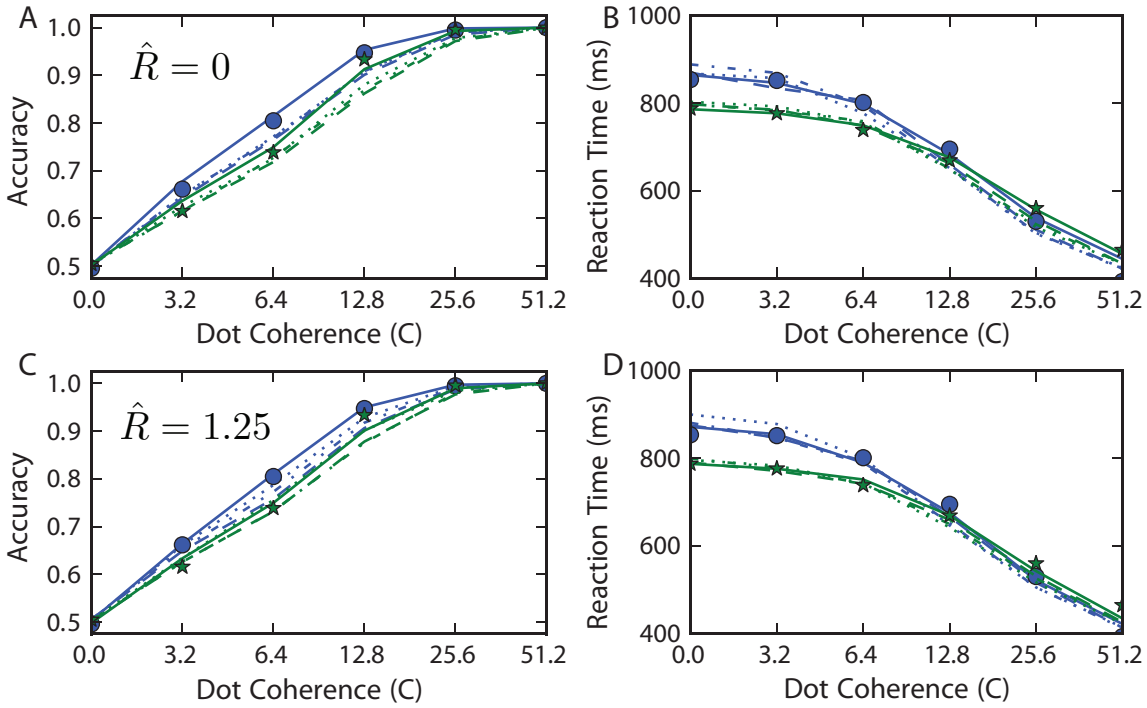


Figure 2.15: Effect of the robustness limit $\hat{R}$ on decision performance in a controlled duration task, under the bounded integration model of Kiani et al (2008). Dot coherence $C = 12.8$. (A) Increasing the robustness limit $\hat{R}$ helps recover performance lost to mistuning at multiple reaction times in the controlled duration task. Specifically, moving from the "baseline" model to the "mistuned" model decreases decision accuracy, but this lost accuracy can be partially or fully recovered for $\hat{R} > 0$. (B) When allowing for biased mistuning ($\bar{\beta} \neq 0$, $\sigma = .1$), $\hat{R}$ still allows for recovery of performance; effects are most pronounced when $\bar{\beta} > 0$.

To address this, we fit accuracy and chronometric functions from robust integrator models to reaction time psychophysics data reported in [93]. This fit is via least squares across the range of coherence values, and requires two free parameters: additive noise variance ν_γ and the decision bound θ (see Methods). Figure 2.16 shows the results. Panels A and B display accuracy and chronometric data (dots) together with fits for various integrator models, with $\hat{R} = 0$. First, the solid line gives the fit for the "baseline" model (i.e., with no feedback mistuning or robustness, see Figure 2.5). The close match between model and data agrees with findings of prior studies [70]. Next, the broken lines give fits for mistuned models ($\sigma_\beta = 0.1$), with three values of bias in feedback mistuning ($\bar{\beta}$). To obtain these fits, both ν_γ and θ are changed from their values for the baseline case.

Panels C and D show analogous results for robust integrators. For all cases in these panels, we take the robustness limit $\hat{R} = 1.25$. We *fix* levels of additive noise to values found for the non-robust case above, in order to demonstrate that by adjusting the decision threshold, one can obtain approximate fits to the same data. This is expected from our results above: Figure 2.7 shows that, while accuracies at given reaction times are higher for mistuned robust vs. non-robust models, the



			Not Robust ($\hat{R} = 0$)	Robust ($\hat{R} = 1.25$)
	β		$(\theta, \sqrt{\nu_\gamma})$	
Perfect Tuning, Subject N	$\sim \delta(0)$	—	**(14.9, 11.7)**	**(9.1, 11.7)**
Mistuning, Subject N	$\sim N(0, .1^2)$	- -	**(10.5, 12.3)**	**(7.3, 12.3)**
	$\sim N(-.05, .1^2)$	-.-	**(9.5, 14.0)**	**(6.7, 14.0)**
	$\sim N(.05, .1^2)$	...	**(12.0, 11.1)**	**(8.6, 11.1)**
Perfect Tuning, Subject B	$\sim \delta(0)$	—	**(16.4, 17.3)**	**(9.8, 17.3)**
Mistuning, Subject B	$\sim N(0, .1^2)$	- -	**(11.8, 17.2)**	**(8.0, 17.2)**
	$\sim N(-.05, .1^2)$	-.-	**(10.7, 18.4)**	**(7.3, 18.4)**
	$\sim N(.05, .1^2)$	...	**(13.3, 15.4)**	**(9.0, 15.4)**

Figure 2.16: Accuracy (A,C) and chronometric (B,D) functions: data and model predictions. Solid dots and stars are behavioral data for a rhesus monkey (Subject "N" and "B" respectively, [93]). In each panel, the accuracy and chronometric functions are fit to behavioral data via least-squares, over the free parameters θ and ν_γ. In Panels (A,B), the robustness threshold $\hat{R} = 0$, and results are shown for "baseline" and exemplar "mistuned" models (see legend in table). In Panels (C,D), results are shown for the "robust" and "recovery" models (R is fixed across the range of coherence values so that $\hat{R} = 1.25$ at $C = 0$). The close matches to data points indicate that these models can be reconciled with the psychophysical performance of individual subjects by varying few parameters. Parameter values for each curve are summarized in the table.

effect is modest on the scale of the full range of values traced over an accuracy curve. Moreover, for the perfectly tuned case, accuracies at given reaction times are very similar for robust and non-robust integrators (Figure 2.8, with a slightly lower value of $\hat{R}$). Thus, comparable pairs of accuracy and RT values are achieved for robust and non-robust models, leading to similar matches with data. In sum, the accuracy and chronometric functions in Figure 2.16 show that all of the models schematized in Figure 2.5 —"baseline", "mistuned", "robust", and "recovery" — are generally compatible with the chronometric and accuracy functions reported in [93].

2.3.7 Reaction time distributions

A limitation of our analysis concerns the distribution of RT. Above we considered the effects of robustness on mean decision time, but not on the shape of the distributions. The standard DDM predicts a longer tail to the RT distribution than is seen in data, thereby necessitating modifications to the simple model [28, 90, 20]. We did not attempt a rigorous fit of the RT distributions from the experiments, however Figure 2.17 depicts a qualitative matching of these experimental data with the RT distributions generated from our model, with parameters chosen from the psychometric and chronometric fitting procedure indicated in Figure 2.16. As can be seen, without additional fitting parameters, the model does not precisely predict the experimental observations. The most compelling modification in our view is a time dependent reduction in the decision threshold θ, the results of which are depicted in Figure 2.18 for Subject "B" from [93]. Our experience suggests that such modifications can also be implemented under robustness with and without mistuning, and leave the matter for future investigation.

2.4 Discussion

A wide range of cognitive functions require the brain to process information over time scales that are at least an order of magnitude greater than values supported by membrane time constants, synaptic integration, and the like. Integration of ev-

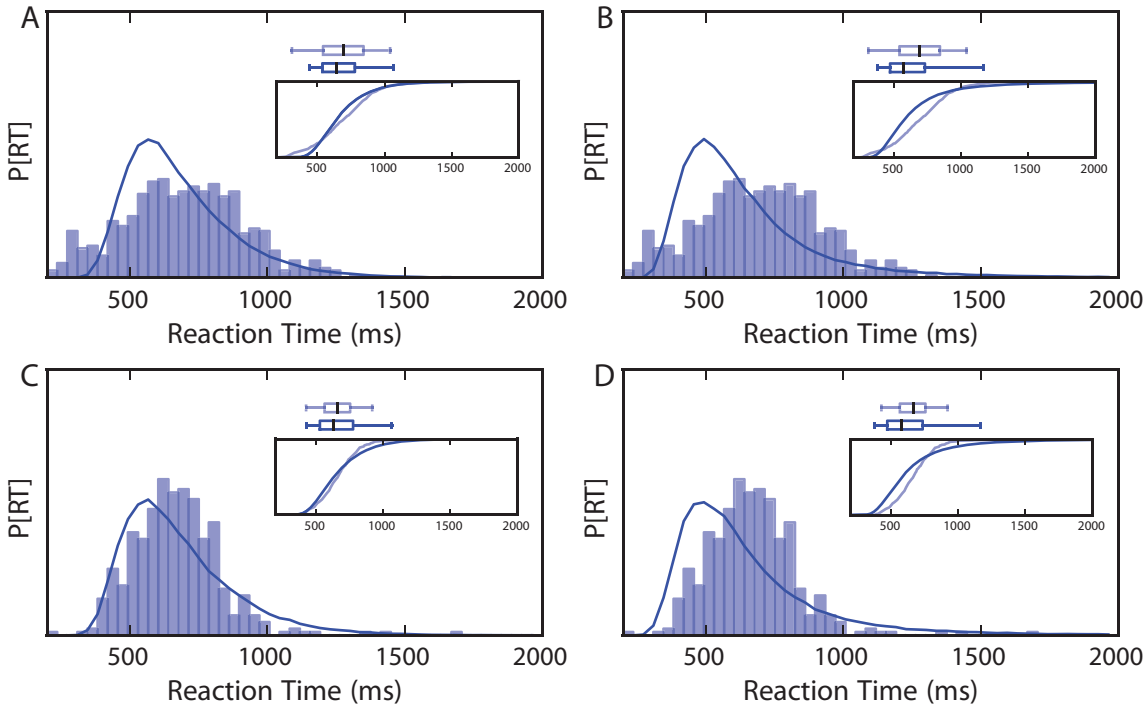


Figure 2.17: Reaction time histograms, with decision bounds held constant in time and dot coherence C=12.8. Probability densities for reaction times for our model with both $\hat{R} = 0, \sigma_\beta = 0$ (A,C) and $\hat{R} = 1.25, \sigma_\beta = .1$ (B,B) are plotted as solid lines. Overlaid are the reaction time histograms for Subject "N" (A,B) Subject "B" in (C,D) from [93]. Box-and-whisker plots indicate the quartiles for each data set. In each panel, both histograms have nearly identical means (owing to selection of model parameters that are taken from the table in Figure 2.16), but clearly differ in basic shape (i.e., the model produces longer tails). This mismatch is a property of our basic integration to bound model, regardless of the value of the robustness limit $\hat{R}$.

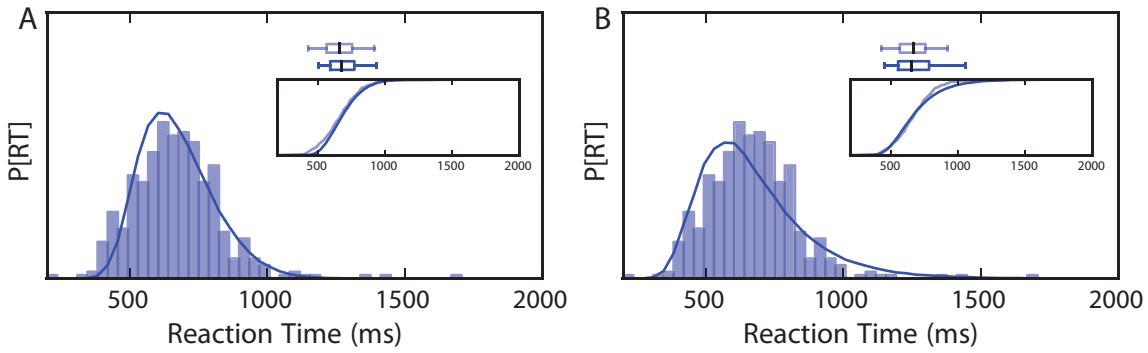


Figure 2.18: Reaction time histograms with collapsing decision bounds, Subject B. The introduction of collapsing bounds produces simulated RT histograms that are a closer match to data, for both non-robust (Panel A, $\hat{R} = 0$) and robust (Panel B, $\hat{R} = 1.15$) integrators. Following [20], we used the functional form for a collapsing bound: $\theta(t) = \theta_0 - (\theta_0 - \theta_{ss})\frac{t}{t+t_{\frac{1}{2}}}$, with $t_{\frac{1}{2}} = 500$, $\theta_0 = 25$, and $\theta_{ss} = 0$.

idence in time, as occurs in simple perceptual decisions, is one such well studied example, whereby evidence bearing on one or another alternative is gradually accumulated over time. This is formally modeled as a bounded random walk or drift-diffusion process in which the state (or decision) variable is the accumulated evidence for one choice and against the alternative(s). Such formal models explain both the speed and accuracy of a variety of decision-making tasks studied in both humans and nonhuman primates [89, 67, 43, 82], and neural correlates have been identified in the firing rates of neurons in the parietal and prefrontal association cortex [70, 43, 20, 103, 98, 104, 55]. The obvious implication is that neurons must somehow integrate evidence supplied by the visual cortex, but there is mystery as to how.

This is a challenging problem because the biological building blocks operate on relatively short time scales. From a broad perspective, the challenge is to assemble neural circuits that that can sustain a stable level of activity (i.e., firing rate) and yet retain the capability to increase or decrease firing rate when perturbed with new input (e.g., momentary evidence). A well known solution is to suppose that recurrent excitation might balance perfectly the decay modes of membranes and synapses [18, 121]. However, this balance must be fine tuned [100, 101], or else the signal will either dissipate or grow exponentially (Figure 2.1A, top). Several investigators have proposed biologically plausible mechanisms that mitigate somewhat

the need for such fine tuning [65, 45, 44, 94, 75, 56]. These are important theoretical advances because they link basic neural mechanism to an important element of cognition and thus provide grist for experiment.

Although they differ in important details, many of the proposed mechanisms can be depicted as if operating on a scalloped energy landscape with relatively stable (low energy) values, which are robust to noise and mistuning in that they require some activation energy to move the system to a larger or smaller value (Figure 2.1A, bottom; cf. [84, 46]). The energy landscape is a convenient way to view such mechanisms – which we refer to as robust integrators – because it also draws attention to a potential cost. The very same effect that renders a location on the landscape stable also implies that the mechanism must ignore information in the incoming signal (i.e., evidence). Here, we have attempted to quantify the costs inherent in this loss. How much loss is tolerable before the circuit misses substantial information in the input? How much loss is consistent with known behavior and physiology?

We focused our analyses on a particular well-studied task because it offers critical benchmarks to assess both the potential costs of robustness to behavior and a gauge of the degree of robustness that might be required to mimic neurophysiological recordings with neural network models. Moreover, the key statistical properties of the signal and noise (to be accumulated over) can be estimated from neural recordings.

Our central finding is that ignoring a surprisingly large part of the motion evidence had almost negligible impact on performance. Indeed, we found that speed and accuracy are preserved even when more than a full standard deviation of the input distribution is ignored. We also found that a similar degree of robustness provides protection of performance against mistuning of recurrent excitation. Although in general this protection is only partial (Figure 2.7), for the controlled duration task it can be nearly complete (Figure 2.15A, controlled duration > 3) depending on the presence of a decision bound.

We can appreciate the impact of robust integration intuitively by considering the distribution of random values that would increment the stochastic process of integrated evidence. Instead of imagining a scalloped energy surface, we simply

replace all the small perturbations in integrated evidence with zeros. Put simply, if a standard integrator would undergo a small step in the positive or negative direction, a robust integrator instead stays exactly where it was. In the setting of drift-diffusion, this is like removing a portion of the distribution of momentary evidence (the part that lies symmetrically about zero) and replacing the mass with a delta function at 0. At first glance this appears to be a dramatic effect – see the illustration of the distributions in Figure 2.9 – and it is surprising that it would not result in strong changes in accuracy or reaction time or both.

Three factors appear to mitigate this loss of momentary evidence. First, we showed that setting weak values of the input signal to zero can reduce both its mean and its standard deviation by a similar amount, creating compensatory effects that result in a small change to the input signal-to-noise ratio. Second, we showed that, surprisingly, the small loss of signal-to-noise that does occur would not result in any loss of accuracy if the accumulation were to the same bound as for a standard integrator. The cost would be to decision time, but mainly in the regime that is dominated by drift – that is, the shorter decision times – hence not a large cost overall. Third, even this slowing is mitigated by the temporal dynamics of the input. Unlike for idealized drift diffusion processes, real input streams possess finite temporal correlation. Left unchecked, this would imply greater variability in the integrated signal. Interestingly, removing the weakest momentary inputs reduces the temporal correlation of the noise component of the input stream. This can be thought of as allowing more independent samples in a given time period, thereby improving accuracy at a given response time.

While we used a simplified characterization of the robust integration operation in our study, we noted that there are many different ways in which this could be realized biologically [56, 81, 45, 46]. In Methods we suggest that a circuit based on bistable neural pools, Equation 3.2.1, can implement a robustness mechanism similar to Equation 2.2.7. Figure 2.19 compares predicted speed and accuracy in the reaction time task for these two models, at three different coherence values. The robustness mechanism provided by the circuit-based bistable model produces an even more favorable effect of robustness for decision making than the simplified model in the main text. This demonstrates the generality of our results and

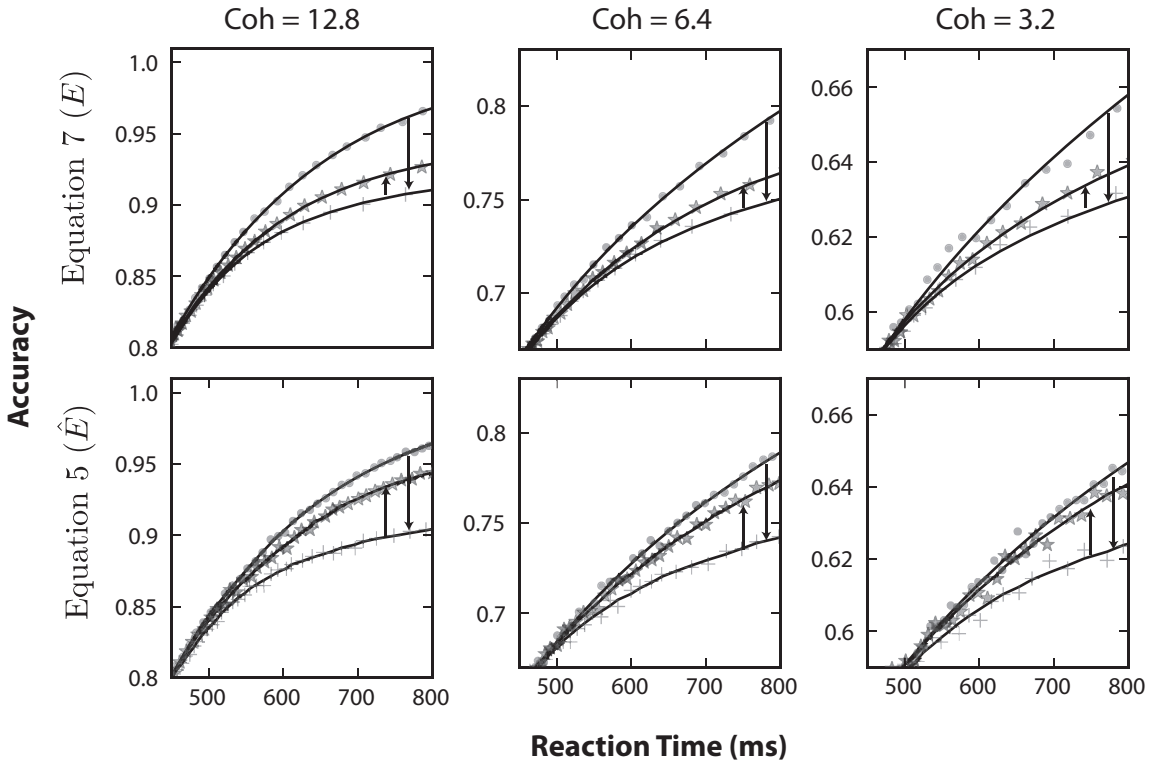


Figure 2.19: Performance comparison of two models of robust integration, reaction time task. In each panel, the lines are the same as in the figure legend in Figure 2.7. The downward arrows indicates the impact of mistuning: increasing σ_β from 0 to .1. The upward arrows show the effect of robustness: increasing $\hat{R}$ from 0 to 1.25. Overall, we see that robustness has an even greater positive impact on the performance of the circuit-based model (Here $q = 1$, $r^- = 0$, $r^+ = 50$, $N = 250$, $\kappa = 1/9$; p is then adjusted to give the required robustness level, see Equation 2.2.6). Data were generated from 50000 numerical trials per value.

points to an intriguing area of future study, focusing on the impact of more detailed circuit-level dynamics.

Our robust integrator framework shares features with existing models in sensory discrimination. The interval of uncertainty model of [112] and the gating model of [86] ignore part of the incoming evidence stream, yet they can explain both behavioral and neural data. We suspect that the analyses developed here might also reveal favorable properties of these models. Notably, some early theories of signal detection also featured a threshold, below which weaker inputs fail to be registered – the so called high threshold theory (reviewed in [118]). The primary difference in the current work is to consider single decisions made based on an accumulation of many such thresholded samples (or a continuous stream of them).

Although they are presented at a general level, our analyses make testable predictions. For example, they predict that pulses of motion evidence added to random dot stimulus would affect decisions in a nonlinear fashion consistent with a soft threshold. Such pulses are known to affect decisions in a manner consistent with bounded drift diffusion [52] and its implementation in a recurrent network [129]. A robust integration mechanism further predicts that brief, stronger pulses will have greater impact on decision accuracy than longer, weaker pulses containing the same total evidence.

However, we believe that the most exciting application of our findings will be to cases in which the strength of evidence changes over time, as expected in almost any natural setting. One simple example is for task stimuli that have an unpredictable onset time, and whose onset is not immediately obvious. For example, in the moving dots task, this would correspond to subtle increases in coherence from a ÒbaselineÓ of zero coherence. Our preliminary calculations agree with intuition that robust integrator mechanism will improve performance: in the period before the onset of coherence, less baseline noise would be accumulated; after the onset of coherence, the present results suggest that inputs will be processed with minimal loss to decision performance – despite the continued ignoring of weak components. This intuition can be generalized to apply to a variety of settings with non-stationary sensory streams.

Many cognitive functions evolve over time scales that are much longer than the perceptual decisions we consider in this paper. Although we have focused on neural integration, it seems likely that many other neural mechanisms are also prone to drift and instability. Hence, the need for robustness may be more general. Yet, it is difficult to see how any mechanism can achieve robustness without ignoring information. If so, our finding may provide some optimism. Although we would not propose that ignorance is bliss, it may be less costly than one would expect.

CHAPTER 3

Derivation of an Effective Robust Integrator

3.1 Introduction

Here we describe a family of neural circuit models, parameterized by a robustness value R, that produce dynamics similar to that of the central robust integrator model of Chapter 2. In particular, both as $R \to 0$ and $\sigma_\beta \to 0$, the dynamics reduce to a perfect integrator (referred to in the main text as the "baseline" model).

Our construction follows closely that of [56] and [45]. Accordingly, robustness arises from the bistability of multiple self-excitatory subpopulations (or subunits). The dynamics of each subunit are governed by one of N differential equations. Depending on the activity in the rest of the network, each individual subunit can be become bistable, so that its eventual steady-state value (i.e., "On" or "Off") depends on its past. This effect, known as hysteresis, underlies several robust integrator models [56, 45].

The circuit integrates inputs via sequential activation of subunits, in an order determined by graded levels of "background" inputs (or biases) to each subunit. Following [45], we collapse the N differential equations that describe individual subunits into an equation that approximates the dynamics of the entire integrator. This expression for the total firing rate $\hat{E}(t)$ averaged over all subpopulations reduces to the robust integrator analyzed in the main text (Equation 7).

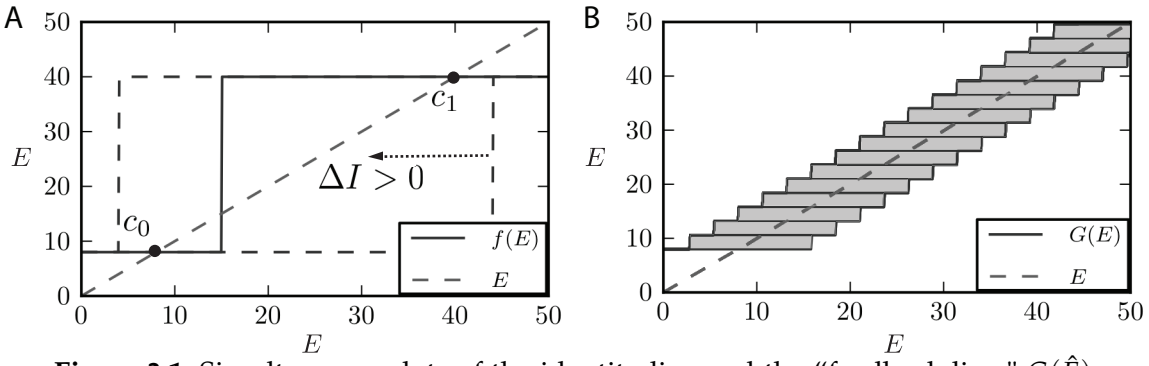


Figure 3.1: Simultaneous plots of the identity line and the "feedback line," $G(\hat{E})$, for two circuits with differing numbers of subunits. (A) Here $N = 1$, and so the feedback line $G(\hat{E}) = f(\hat{E})$ is exactly determined as a function of $\hat{E} = r_1$ (see Equation 3.2.4). The two intersections c_0 and c_1 are stable fixed points. In this way, the subunit's firing rate is bistable, and the value attained will depend on the history of the circuit activity. As ΔI is changed, this translates $f(\hat{E})$, eventually eliminating either c_0 or c_1 and forcing the subunit to the remaining stable fixed point (here, c_0 corresponds to $\hat{E} = 8$ Hz. and c_1 to 40 Hz). In dashed curves, the feedback line is plotted for two such values of ΔI. (B) Now $N > 1$ and so the feedback line $G(\hat{E})$ is no longer unambiguously specified as a function of its argument. The function is instead the sum of N potentially bi-valued functions, whose actual values will depend on the stimulus history. We represent this fact by plotting the feedback line as a set of stacked boxes, representing the potential contribution of the i^{th} subunit to the total integrator dynamics [45].

3.2 Firing rate model

The firing rate $r_i(t)$ of the ith bistable subunit ($i \in \{1, 2, ..., N\}$) is modeled by a firing rate equation:

$$\tau_E \frac{dr_i}{dt} = -r_i + r^- + (r^+ - r^-)H\left[pr_i + q(1+\beta)\sum_{i\neq j}^{N} r_j + a\Delta I - b_i\right] \tag{3.2.1}$$

$$\hat{E} = \frac{1}{N}\sum_{i=1}^{N} r_i \tag{3.2.2}$$

Figure 3.1A demonstrates the firing rate dynamics of a circuit composed of a single subunit ($N = 1$). We plot the identity line, corresponding to the first "decay" term in Equation 3.2.1, and the "feedback" line, corresponding to the second.

Since $N = 1$, this simplifies to $f(r_1)$. The two intersections marked c_0 and c_1 are stable fixed points (which we refer to as "On" and "Off" respectively). Thus, the subunit shown is bistable. Importantly, however, the location of the step in $f(\hat{E})$ varies with changes in the input signal (as per Equation 3.2.1). In particular, substantial values of $\Delta I(t)$ will (perhaps transiently) eliminate one of the fixed points, forcing the subunit into either the "On" or the "Off" state with $r_i = r_+$ or $r_i = r_-$, respectively. Moreover, the change is self-reinforcing via the recurrent excitation pr_i. The range over which a given subunit displays bistability is affected by the mistuning parameter β, which scales the total recurrent excitation from the rest of the circuit.

The firing rate dynamics, $\hat{E}(t)$, are obtained by summing both sides of Equation 3.2.1 over i:

$$\tau_E \frac{d\hat{E}}{dt} = -\hat{E} + G(\hat{E}) \tag{3.2.3}$$

where

$$G(\hat{E}) = r^- + \frac{(r^+ - r^-)}{N} \sum_{i=1}^{N} H\left[(p - q(1+\beta))r_i + Nq(1+\beta)\hat{E} + a\Delta I - b_i\right] \quad . \tag{3.2.4}$$

At this point, we almost have a differential equation for a single variable, $\hat{E}(t)$. However, Equation 3.2.4 still depends on the N activities r_i of the individual subunits, and at any particular time their values are not uniquely determined by the value of $\hat{E}$; we can only bound their values as $r^- \leq r_i \leq r^+$.

3.3 Bias term

The bias term for the i^{th} subunit, b_i, is set by determining the range of values of $\hat{E}$ for which the exact value of the feedback function $f(r_i)$ is unknown. In the case of an integrator composed of only a single subunit, the bias term causes the positive input needed to force the unit to be on, and the negative input needed to force the unit to be off, to take the same values. This yields $b_1 = \frac{p(r^+ + r^-)}{2}$.

The general case of N subunits is more complicated. Now the feedback contribution of the i^{th} unit, $f(r_i)$, is no longer a simple function of the population activity

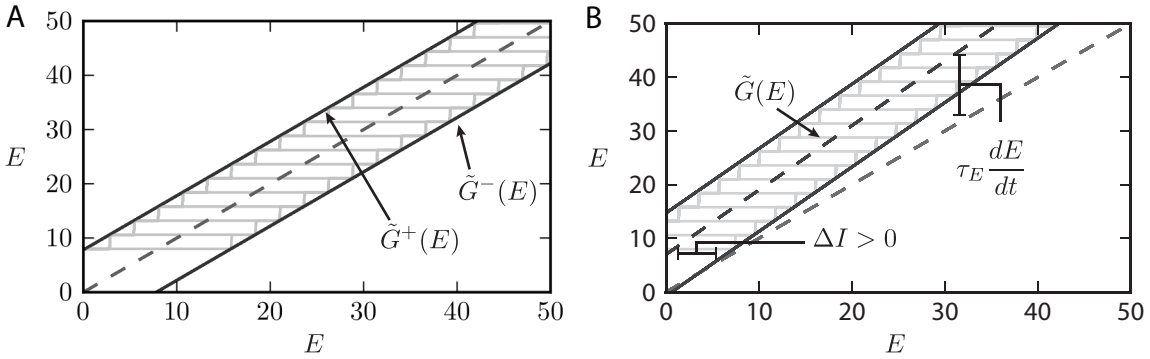


Figure 3.2: Plot of possible equilibria for Equation 3.2.3. (A) The extent of each multivalued feedback function defines the minimum input necessary to perturb the system away from equilibrium, defining the "fixation" lines. As $N \to \infty$, the stable fixed points become more tightly packed on the interval (r^-, r^+). (B) When the integrator is mistuned, the fixation lines and the feedback line are no longer parallel. The rate that the integrator accumulates input is approximated by the distance between the center line of the feedback subunits ($\tilde{G}(\hat{E})$), and the feedback line. However, integration only occurs when the "fixation condition" is no longer satisfied, i.e. when the feedback line is no longer bounded by the fixation lines at the current value of $\hat{E}$.

$\hat{E}$. Instead, it has additional dependence on its own activity r_i. We see this clearly in Equation 3.2.4, where the values of r_i that contribute to the definition of $G(\hat{E})$ are unspecified. However, we do know that each r_i is trapped between r^+ and r^-. Therefore, we can plot $G(\hat{E})$ as the sum of a sequence of bivalued functions of $\hat{E}$; see Figure 3.1B and [45]. The contribution from each pool is then represented by the shaded region. Finally, the bias terms are chosen to center these shaded boxes over the identity line:

$$b_i = p\left(\frac{r^- + r^+}{2}\right) + q((i-1)r^+ + (N-i)r^-)\,. \tag{3.3.1}$$

3.4 Fixation lines

We next define the "fixation" lines ($\tilde{G}^+(\hat{E})$ and $\tilde{G}^-(\hat{E})$), which are the consequence of the multi-valued property of the integrator. These lines define a region (the fixation region) that runs across the outermost corners of the "stacked boxes" in

Figure 3.2.

$$\tilde{G}^{+}(\hat{E}) = \hat{E}(1+\beta) + \frac{a\Delta I}{Nq} - \frac{\beta r^{+}}{N} + \frac{p(r^{+}-r^{-})}{2Nq} \tag{3.4.1}$$

$$\tilde{G}^{-}(\hat{E}) = \hat{E}(1+\beta) + \frac{a\Delta I}{Nq} - \frac{\beta r^{-}}{N} - \frac{p(r^{+}-r^{-})}{2Nq}\,. \tag{3.4.2}$$

The term fixation region refers to the following property: if the input ΔI is such that the identity line lies within the fixation region, then the integrator will possess a range of closely spaced fixed points (where $\hat{E} = G(\hat{E})$). Thus, $\hat{E}$ is not expected to change from its current value and integration of ΔI will not occur. Recall that ΔI acts to shift these boxes leftward or rightward relative to the identity line, just as in the analysis of Figure 3.1. As a consequence, it is weak inputs that fail to be integrated.

From this analysis, we can see that integration by the system as a whole relies on two concepts. The first is a condition on ΔI necessary to eliminate fixed points; we call this the "fixation condition." The second is the question of how quickly to integrate once this condition is no longer satisfied.

3.5 Integration

Based on the analysis above, we derive a reduced model that approximately captures the dynamics of the "full" model indicated by Equation 3.2.3. We call this the "effective" model. The rate of change of $\hat{E}$ – i.e., the rate of integration – is given by the distance between the current value of $G(\hat{E})$ and the identity line. We approximate this by the distance between the *middle* of the fixation lines, which we define as $\tilde{G}(\hat{E})$, and the identity line. This is pictured in Figure 3.2B, and yields:

$$\tau_E \frac{d\hat{E}}{dt} = -\hat{E} + G(\hat{E}) \approx -\hat{E} + \tilde{G}(\hat{E}) \tag{3.5.1}$$

$$\tilde{G}(\hat{E}) = (1+\beta)\hat{E} + \frac{a\Delta I}{Nq} - \beta\frac{(r^{+}+r^{-})}{2N} \tag{3.5.2}$$

We emphasize that integration by this equation only occurs when the "fixation" condition is no longer satisfied, i.e. when the fixation lines no longer bound the

identity line.

3.6 Fixation condition

The last step in defining the 1-dimensional "effective" model is determining the fixation condition. We must solve for the values of $\hat{E}$ that cause the feedback line to lie between the two fixation lines:

$$\text{No Change in } \hat{E} \iff \tilde{G}^-(\hat{E}) < \hat{E} < \tilde{G}^+(\hat{E}) \tag{3.6.1}$$

$$\iff \tilde{G}(\hat{E}) - \frac{(r^+ - r^-)(p - q\beta)}{2Nq} < \hat{E} < \tilde{G}(\hat{E}) + \frac{(r^+ - r^-)(p - q\beta)}{2Nq} \tag{3.6.2}$$

$$\iff \left| \beta\hat{E} + \frac{a\Delta I}{Nq} - \beta\frac{(r^+ + r^-)}{2N} \right| < \frac{(r^+ - r^-)(p - \beta q)}{2Nq} \tag{3.6.3}$$

If this condition is violated with $\Delta I = 0$, the integrator displays runaway integration ($\beta > 0$) or leak ($\beta < 0$). If it is satisfied when $\Delta I = 0$, we have a condition on the level of ΔI that must be present for integration to occur. This yields a piecewise-defined differential equation, corresponding to when integration can and cannot occur:

$$\tau_E \frac{d\hat{E}}{dt} \approx \begin{cases} 0 & \left| \beta\hat{E} + \frac{a\Delta I}{Nq} - \beta\frac{(r^+ + r^-)}{2N} \right| < \frac{(r^+ - r^-)(p - \beta q)}{2Nq} \\ \beta\hat{E} + \frac{a\Delta I}{Nq} - \beta\frac{(r^+ + r^-)}{2N} & \text{otherwise} \end{cases} \tag{3.6.4}$$

We now simplify this equation; we assume that $p \sim O(N)$ and $a \sim O(N)$, i.e. the local feedback term and input weight can be increased as N is increased. With all other terms held constant, we relabel $\kappa = \frac{a}{Nq}$ and $R = \frac{(r^+ - r^-)p}{2a}$; here R is the robustness parameter in the main text, and can be decreased or increased by adjusting p. This yields the central relationship of Chapter 2:

$$\tau_E \frac{dE}{dt} = \begin{cases} 0 & |\beta E + \kappa\Delta I| \leq \kappa R \\ \beta E + \kappa\Delta I & \text{otherwise} \end{cases} \tag{3.6.5}$$

CHAPTER 4

Impact of correlated neural activity on decision making performance

4.1 Introduction

Sensory information is often encoded in irregularly spiking neural populations. One well-studied example is given by direction-selective cells in area MT, whose firing rates depend on the degree and direction of coherent motion in the visual field [13, 79, 12, 97]. Individual neurons in MT – as in many other brain areas – exhibit noisy and variable spiking [79], as can be modeled by Poisson point processes [114, 119]. Moreover, this variable spiking is generally not independent from cell to cell. Returning to our example, a number of studies have measured pairwise correlations in MT during direction discrimination tasks as well as smooth-pursuit eye movements [51, 3, 134, 22]; while this measurement is a subtle endeavor experimentally, a number of studies suggest a value near $\rho \approx .1 - .15$ ([21] summarizes these observations, for a number of brain areas.)

What are the consequences of correlated spike variability for the speed and accuracy of sensory decisions? The role of pairwise correlations in stimulus encoding has been the subject of many prior studies [96, 60, 2]. The results are rich, showing that correlations can have positive, negative, or neutral effects on levels of encoded information. The present study serves to extend this body of work in two ways. First, as done in a different context by [38, 77], we contrast the impact

of correlations that have the same pairwise level but a different structure at higher orders.

Second, as in [22, 4], we consider the impact of correlations on decisions that unfold over time, by combining a sequence of samples observed over time in the sensory populations. A classical example that we will use to describe and motivate our studies is the *moving dots* direction discrimination task. Here, a fraction of dots in a visual display move coherently in a given direction, while the remainder display random motion; the task is to identify the direction from two possible alternatives. Decisions become increasingly accurate as subjects take (or are given) longer to make the decision.

In analyzing decisions that develop over time, we utilize a central result from sequential analysis. This is the Sequential Probability Ratio Test (SPRT) [125, 42], which linearly sums the log-odds of independent observations from a sampling distribution until a predetermined evidence threshold is reached. The SPRT is the optimal statistical test in that it gives the minimum expected number of samples for a required level of accuracy in deciding among two task alternatives.

We pose two related questions based on the SPRT. First, how does the presence of correlated spiking in the sampled pools impact the speed and accuracy of decisions produced by the SPRT? Our focus is on how the structure of population-wide correlations determines the answer. Second, how does the presence of correlated spiking impact the computations that are necessary to perform the SPRT? This question is intriguing, because the SPRT may be performed via the simple, linear computation of integrating spikes over time and across the populations for a surprisingly broad class of inputs, including independent Poisson spike trains [133, 9]. Thus, in this setting optimal decisions can be made by integrator circuits [9, 46, 17]. Our goal here is to determine whether and when this continues to hold true for correlated neural populations.

We answer these questions for two illustrative models of correlated, Poissonian spiking. We emphasize that the spikes that these models produce are indistinguishable at the level of both single cells and pairs of cells. However, they differ in higher-order correlations, in that they can only be distinguished by examining the statistics of three or more neurons. In the first model, correlations are introduced

via shared spike events across the entire pool. In this case optimal inference via the SPRT produces fast and accurate decisions, but depends on a nonlinear computation. As a result, the simpler computation of spike integration requires, on average, longer times to reach the same level of accuracy. In contrast, when shared spiking events are more frequent but are common to fewer neurons within a pool, performance under the SPRT is significantly diminished. However, in this case both SPRT and spike integration perform comparably, so a linear computation can produce decisions that are close to optimal.

4.2 Models of evidence accumulation and encoding

4.2.1 Model neural populations and the decision task

We begin by introducing the notation for the two decision making models that will be compared. In this study we consider the case of discrimination between two alternatives, and therefore model two populations of neurons that encode the strength of evidence for each alternative. Returning to the moving dots task for illustration, each population could be the set of MT cells that are selective for motion in a given direction. Here, the firing rates in each population represents the dot motion C via their firing rates λ_p and λ_n; here the subscripts indicate the "preferred" and "null" populations, which correspond to the motion direction of the visual stimulus versus the alternate direction. In this way, the firing rate of neurons encoding the preferred direction will be higher than the null direction, $\lambda_p > \lambda_n$. Following [126] (see also [70, 13]), we model this relationship as linear:

$$\lambda_p = 40 + .4C \text{ Hz} \tag{4.2.1}$$

$$\lambda_n = 40 - .4C \text{ Hz}. \tag{4.2.2}$$

Throughout the text we consider present results at $C = 6.4$, however the results do not depend on this particular value of dot motion or its precise relationship firing rate.

In our model, we assume that each population consists of N neurons firing

spikes via a homogenous Poisson process, with rate λ_p or λ_n. We use the notation $x_k(t)$ to each spike train. Integrating these processes over a time interval ΔT provides two time series of $N-$dimensional vectors of Poisson random variables; these independent vectors provide the input to the decision making models. Specifically, for the k^{th} neuron in a pool, on the i^{th} time step,

$$S_k^i = \int_{i\Delta T}^{(i+1)\Delta T} x_k(t)dt \sim \text{Poiss}(\lambda \Delta T). \tag{4.2.3}$$

The properties of Poisson processes imply that S_k^i is independent from S_k^j $(i \neq j)$, i.e. for different time steps.

However, the outputs of different neurons in the same time are not, in general, independent. Following experimental observations that neurons with similar directional tuning tend to be correlated, while those with very different tuning are not [134, 22], we model neurons from different pools as independent and those within a single pool as correlated with a correlation coefficient ρ:

$$\rho = \frac{\text{Cov}[S_k^i, S_l^i]}{\sqrt{\text{Var}[S_k^i]\text{Var}[S_l^i]}}, \quad k \neq l. \tag{4.2.4}$$

This implies that, with vector notation for the probability distribution of spike counts for each pool,

$$P[\boldsymbol{S_p^i}, \boldsymbol{S_n^i}] = P[\boldsymbol{S_p^i}]P[\boldsymbol{S_n^i}]. \tag{4.2.5}$$

Next, we introduce notation for decision making between the two task alternatives. The task of determining, e.g., direction in the moving dots task is that of determining which of the two pools fires spikes with the higher firing rate. We frame this as decision making between the hypotheses

$$H_1 \ : \ \lambda_p > \lambda_n \tag{4.2.6}$$

$$H_0 \ : \ \lambda_p < \lambda_n, \tag{4.2.7}$$

where each alternative corresponds to a decision as to the motion direction. This formalism allows us to define accuracy as the fraction of trials on which the correct

hypothesis H_1 is accepted. In this study we consider decision making tasks at a fixed level of difficulty, so that λ_p and λ_n do not vary from trial to trial (i.e., this hypothesis test is simple and not composite).

4.2.2 Accumulating spikes and evidence over time

We relate the decision making task to a discrete random walk, which follows in turn from the sequential accumulation of independent and identically distributed (IID) realizations from the sampling distribution W_i. We will specify this distribution below; for now, we note that the random walk takes the general form:

$$E_0 = 0 \tag{4.2.8}$$

$$E_{n+1} = E_n + W_n, \tag{4.2.9}$$

In a drift-diffusion model of decision making, accumulation continues as long as $|E_n| < \theta$, the decision threshold [89, 42, 9]. The number of increments necessary to cross one of the two increments multiplied by its duration ΔT defines the decision time; this is a random variable, as it varies from trial to trial. Crossing the threshold corresponding to H_1 is interpreted as a correct trial; the fraction of correct (FC) trials defines the accuracy of a the model. Together, the expected (mean) decision time (DT) and accuracy (FC) determine the performance of a decision making model.

Formulas for the mean decision time and accuracy are given in Wald [123] as a function of the sampling distribution and the decision threshold. Importantly, these formulas are exact under the assumption that the final increment in E_n does not overshoot the threshold, a point we return to below. Given the moment generating function for the sampling distribution:

$$\phi(s) = E[e^{Ws}], \tag{4.2.10}$$

Speed and accuracy are given by:

$$FC \approx \frac{1}{1+e^{h_0\theta}} \tag{4.2.11}$$

$$DT \approx \frac{\theta\Delta T}{E[W]}\tanh\left(\frac{-h_0\theta}{2}\right) \tag{4.2.12}$$

where h_0 is the nontrivial root of $\phi(s)-1$, i.e.

$$\phi(h_0)=1, \quad h_0\neq 0. \tag{4.2.13}$$

We notice here that as θ increases (and assuming $h_0 < 0$), both FC and RT will increase.

We now return to the definition of the random increments W_i. We consider two different ways in which this can be done. First, in the spike integration (SI) model, increments are constructed by counting the spikes emitted in a ΔT window by the preferred pool, and subtracting the number emitted by the null pool. This is equivalent to the time evolution of a neural integrator model that receives spikes as impulses with opposite signs from the preferred and null populations. This integrate-to-bound model is an analog of drift-diffusion model (DDM) with inputs that are not "white noise", but rather Poisson spikes:

$$W_i = \sum_{k=1}^{N} S^i_{k,p} - \sum_{k=1}^{N} S^i_{k,n} \tag{4.2.14}$$

[89, 9, 133, 4], cf. [70].

Second, in the Sequential Probability Ratio Test (SPRT), the increment is defined as the log-odds ratio of observing the spike count from both of the pools, under each of the two competing hypothesis:

$$W_i = \log\left[\frac{P[\boldsymbol{S^i_p}|H_1]P[\boldsymbol{S^i_n}|H_1]}{P[\boldsymbol{S^i_p}|H_0]P[\boldsymbol{S^i_n}|H_0]}\right] \tag{4.2.15}$$

4.2.3 The case of independent neurons

[133] present an analysis of speed and accuracy of decision making based on independent neural pools; for completeness, and to help contrast this result with the correlated case, we give the key calculations in Sections 4.7.1, 4.8.1. Here, choosing increments via the SPRT yields:

$$h_0 = -1 \tag{4.2.16}$$

$$E[W] = \Delta TN\left(\lambda_p - \lambda_n\right)\log\left(\frac{\lambda_p}{\lambda_n}\right). \tag{4.2.17}$$

Under the spike integration model, Zhang and Bogacz [133] (See also Section 4.8.1) find that:

$$h_0 = -\log\left(\frac{\lambda_p}{\lambda_n}\right). \tag{4.2.18}$$

$$E[W] = \Delta TN\left(\lambda_p - \lambda_n\right). \tag{4.2.19}$$

Therefore, by applying a change of variables $\theta \rightarrow \theta\log\left(\frac{\lambda_p}{\lambda_n}\right)$ in Equations 5.3.1 and 5.3.2, spike integration can implement the SPRT. The implication is that simply counting spikes, positive for one pool and negative for the other, can implement statistically optimal decisions for when the neural pools are independent [133].

4.2.4 Correlated neural populations: SIP and MIP models

We next describe two models for introducing correlations into the Poisson spike trains of each neural population.Both models are studied in [58, 116], and rely on shared input from a single correlating process to generate the correlations in each pool. These authors termed the two model SIP and MIP for single- and multiple-interaction process; here we use the added descriptors “additive" and “subtractive." In both models, a realization of correlated spike trains that provide the input to the accumulation models is achieved via a common correlating train.

Before describing the models in detail, we note that in this study, these models are statistical approaches chosen to illustrate a range of impacts that correlations can have on decision making (see also [48, 80]). In contrast, in neurobiological

networks, correlated spiking arises as through a complex interplay of many mechanisms, including recurrent connectivity and shared feedforward interactions (For example, [1, 105, 108]). While beyond the scope of the present paper, avenues for bridging the gap between statistical and network-based models of correlations in the context of decision making are considered in the Discussion.

The first case is the additive (SIP) model, in which the spike train for each neuron is generated as the sum of two homogenous Poisson point processes. The first Poisson train is generated with an overall firing rate of $(1-\rho)\lambda$, where λ is the intended firing rate of the neuron, and ρ is the intended pairwise spike count correlation between any two neurons in the pool. The second train, with a rate of $\rho\lambda$, is added to every neuron in the pool, and serves as the common source of correlations. An example of this model of spike train generation is depicted in the rastergrams in Figure 4.1A and B; the common spike events are evident as shared spikes across the entire population.

The second case is the subtractive (MIP) model, in which correlated spikes are generated through random, independent deletions from an original “mother" spike-train; we refer to this as the correlating spike train [58]. There is a separate correlating spike train for each of the two independent populations. In order to achieve an overall firing rate for the pool of λ spikes per second, with a pairwise correlation ρ between any two individual neurons, the correlating train has a rate of λ/ρ spikes per second. Then, for each neuron in the pool, a spike is included from this train IID with a probability of ρ. An example of this model of spike train generation is depicted in the rastergrams in Figure 4.1D and E.

In summary, the two models both include correlated spike events that originate in from a single “mother." Although they produce identical correlations among cell pairs, these events are distributed in different ways across the entire population. We note that the results of [133] can be seen as a limiting case as $\rho \to 0$ of either the additive (SIP) or subtractive (MIP) models.

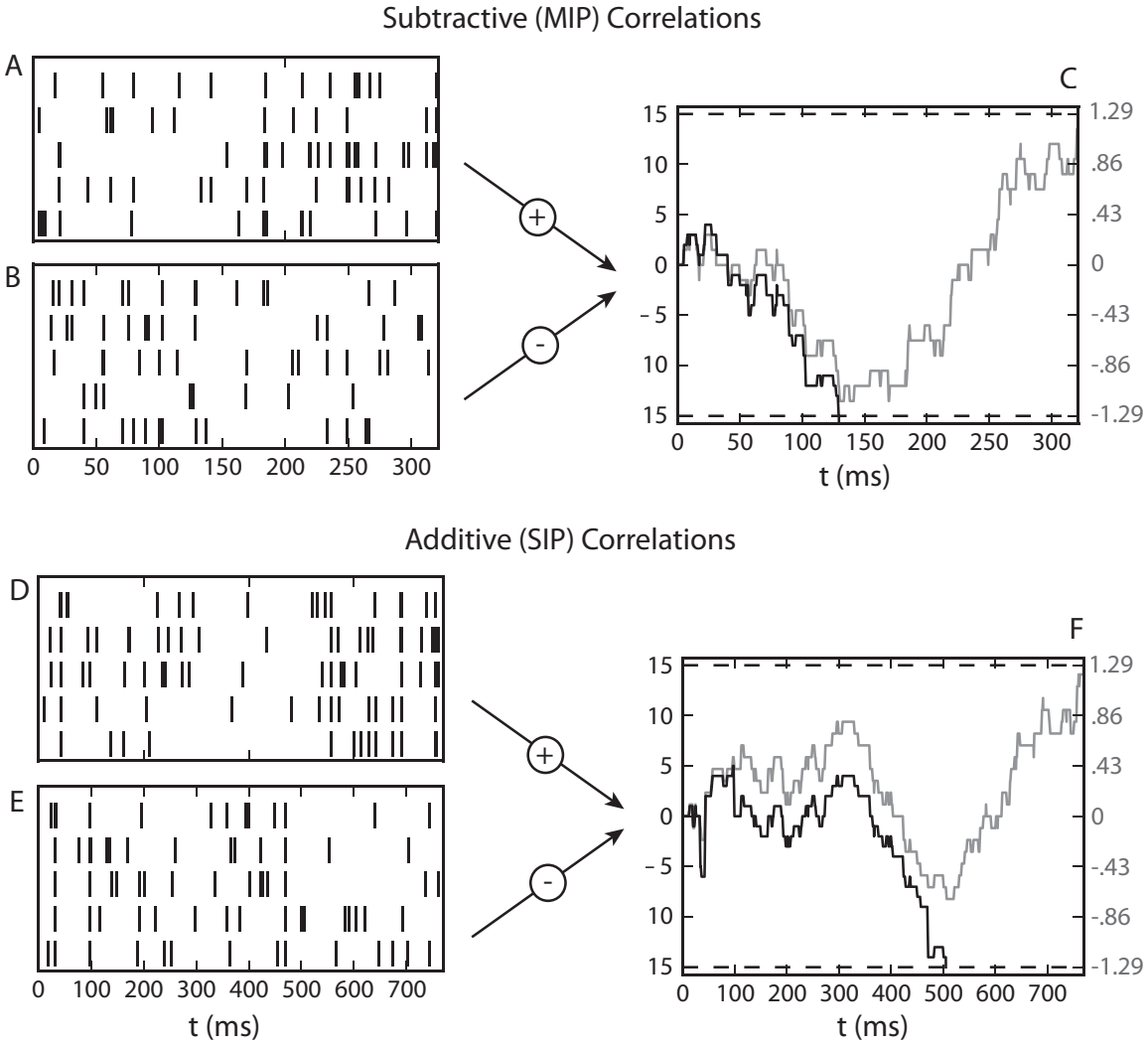


Figure 4.1: Spike integration (SI) and SPRT for a single trial, with subtractive (MIP) correlations (A,B,C) and additive (SIP) correlations (D,E,F). Rastergrams at C= 6.4 for preferred (A,D) and null (B,E) populations of 5 neurons, with spike count correlation within pools $\rho = .15$. In (C,F) these spikes are either integrated (black line) or provide input for the SPRT (gray line), until a decision threshold is reached. The decision threshold has been set so that all four cases will yield the same mean reaction time (In C, $\theta_{SI} = 15$ and $\theta_{SPRT} = 1.28$, and in F $\theta_{SI} = 14$ and $\theta_{SPRT} = 1.28$; in both cases the SPRT lines have been scaled for plotting purposes). On these trials, the SPRT accumulator crosses the "correct", upper, threshold, as opposed to the "incorrect", lower, threshold for the spike integrator. Unlike the independent case, the time evolution of the spike integration process is not simply a scaled version of the SPRT (though they are clearly similar) under either model of correlations.

4.3 Subtractive (MIP) correlations and decision making performance

4.3.1 The SPRT decision making model

We now study the impact of subtractive (MIP) correlations on decision making performance. As noted above, recall that within a time window ΔT, the spike counts from each neuron form a vector of random variables which are independent from window to window. These independent vectors provide the evidence for each of the two alternatives, which is then weighed via log-likelihood at each step in SPRT. In Sections 4.7.1 and 4.7.4, we compute the values h_0 and $E[W]$ that define the speed and accuracy of the SPRT (see Equations 5.3.1-5.3.2), for two pools with subtractive (MIP) correlations. As this computation is done in continuous time, it is natural to take $\Delta T \to 0$; doing so, we find:

$$h_0 = -1 \tag{4.3.1}$$

$$E[W] = \frac{1-(1-\rho)^N}{\rho}(\lambda_p - \lambda_n)\log\frac{\lambda_p}{\lambda_n}\Delta T + O(\Delta T^2) \tag{4.3.2}$$

Comparing these values against those of the independent SPRT given in Equations 4.2.16 and 4.2.17, we see that the only effect of correlations is a scaling of the expected increment via $\left(1-(1-\rho)^N\right)/\rho$. In the limit as $\rho \to 0$, this scale factor approaches N, which in turn reduces decision time (the scale factor is inversely proportional to DT via Equation 5.3.2). On the other hand, as $\rho \to 1$, the scale factor itself approaches 1; this agrees with the intuition that as all neurons become perfectly redundant, the performance should resemble that of a single neurons. In fact, the mechanism of the SPRT on a given sample can be seen as inferring the firing rate of the correlating train from a derived vector of noisy random variables. As N gets large, then, performance should be limited by performing an SPRT on the correlating "mother" trains themselves. This is precisely what happens when $N \to \infty$ in Equation 4.3.2: we obtain $E[W] \sim \frac{1}{\rho}(\lambda_p - \lambda_n)\log\frac{\lambda_p}{\lambda_n}\Delta T$, corresponding to decision making based on mother spikes of rate λ_p/ρ and λ_n/ρ.

One consequence of this interpretation is that the particular realization of a

spike vector (in a sufficiently small time-bin ΔT) carries no evidence about the decision of H_1 vs. H_0, beyond its identity as either the zero vector $\mathbf{0}$ or not. Of course, this is a consequence of the construction of the MIP model, as the spike deletions that create the realization of the spike vector have no dependence on the firing rate of the population. Concretely then, the increments (or decrements) are based solely on whether the vector of spikes in the preferred (or null) pool contains any spikes at all; the actual number of spikes is irrelevant in the SPRT.

It follows that the accumulation process E_n is a discrete-space random walk, with steps $\pm \log(\lambda_p/\lambda_n)$. To see this, note that for sufficiently small ΔT, there are only three possibilities for how spikes will be emitted from the two populations. First, both the preferred and null pools could produce no spikes. This event provides no information to distinguish the firing rates of the pools, so the increment is 0. Second, one of the pools could produce a vector of spikes caused by IID deletions from the "mother" spike train. If the spiking pool is the preferred one, each possible nonzero spike vector will increment the accumulator by the log of the ratio $(\lambda_p/\rho)/(\lambda_n/\rho)$; the opposite sign occurs if the null pool spikes. Events in which both pools spike are of higher order in ΔT, and thus become negligible for small time windows.

The discrete nature of the SPRT effect causes the *FC* curve in Figure 4.2(A) to take on only discrete values of accuracy; a small increase in θ above a multiple of $\log(\lambda_p/\lambda_n)$ will not improve accuracy because E_n on the final, threshold-crossing-step will overshoot the threshold. This also explains why some of the *FC* values at a given θ do not lie on the theoretical line defined by Equation 5.3.1; that equation is only exactly true in the case of zero overshoot past the threshold. We will return to this point later, and also in Section 4.9.

We next insert the values for h_0 and $E[W]$ computed above into Equations 5.3.1 and 5.3.2, and plot the resulting speed-accuracy curves relating *DT* and *FC* parametrically in the threshold θ (Figure 4.2(B)). (We plot the full *FC* and *RT* functions, although only discrete values of performance along each of the lines are achievable in practice, as indicated by the dots for the $\rho = .15$ case; see caption). By comparing speed-accuracy curves for different values of ρ ranging from 0 to 0.3, we see our first main result: *introducing MIP correlations within neural populations substan-*

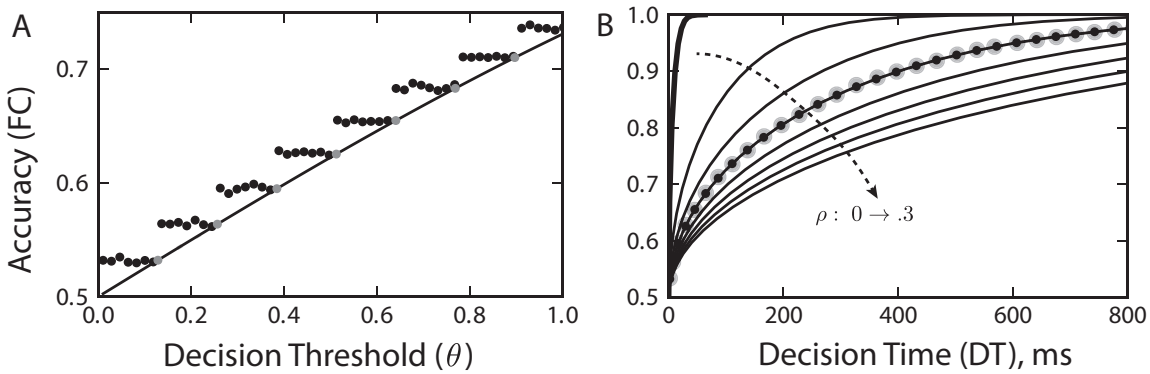


Figure 4.2: Subtractive (MIP) correlations significantly diminish decision performance under SPRT (C= 6.4, $N = 240$). (A) The discrete nature of the SPRT diffusion process implies that only discrete values of accuracy are possible. These occur at values of θ that are multiples of $\log(\lambda_p/\lambda_n) \approx .1.28$. (Similar results hold for decision time, not shown.) The solid dots are simulations of the SPRT, and gray dots are exact values taken at multiples of the log ratio; the interpolating line is Equation 5.3.1. (B) Accuracy (Equation 5.3.1) and decision time (Equation 5.3.2) are plotted parametrically as a function of threshold, for 8 different values of ρ (linearly spaced on [0,.35] with the a double-thickness line at $\rho = 0$). Performance of the simulation at multiples of the log-ratio of firing rates are plotted as solid dots, and theoretical values in gray (gray dots are enlarged to be distinguished).

tially diminishes the best-possible decision performance, that obtained via the SPRT. We will next derive the analogous results for the simpler spike integration model.

4.3.2 The spike integration decision making model

Next, we consider decision making performance for the simpler model in which spikes are simply integrated over time, as opposed to the likelihood ratio computation of the SPRT. In this case, the moment generating function of the difference in spike counts from the two pools is more straightforward (See Section 4.8.3), and provides an easy computation of $E[W]$:

$$E[W] = \Delta TN\left(\lambda_p - \lambda_n\right) \tag{4.3.3}$$

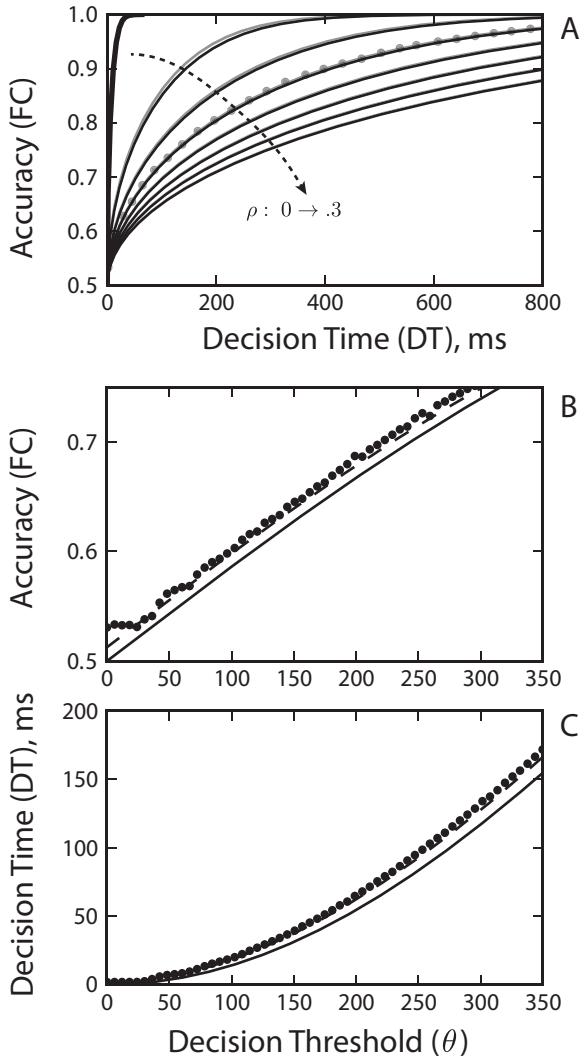


Figure 4.3: For the subtractive (MIP) model of spike correlations, decision making performance of the spike integration model is comparable to the SPRT, and is well described by Equations 5.3.1 and 5.3.2 despite overshoot past past the decision threshold. (A) Gray lines are reproductions of speed accuracy curves from the SPRT (Figure 4.2), and black lines are speed accuracy curves for spike integration. (B,C) Overshoot past the decision boundaries reduces the validity of Wald's approximations, but a constant shift in threshold can help mitigate the effect (See [40, 61] and Section 4.9). Such a shift is automatically accounted for when comparing curves that are parametric in θ (Panel A, for example).

The nontrivial root of the MGF h_0 is found to be the implicit solution of:

$$\left(((1+\rho(e^{t}-1))^{N}-1)\right)\lambda_p+\left(((1+\rho(e^{-t}-1))^{N}-1)\right)\lambda_n=0 \tag{4.3.4}$$

Here we see that correlations only impact the performance of the model through changing h_0, as the expected increment is the same is in the independent case (Equation 5.3.4). Moreover, performance under spike integration is diminished to a degree that is comparable to the performance loss of SPRT. To illustrate this, Figure **??**A plots the speed-accuracy tradeoff curves from both models of decision making under subtractive correlations, for the same values of ρ. As we must([125]), we see the optimal character of the SPRT in the fact that at a given level of accuracy, the SPRT requires, on average, fewer samples than spike integration. However, the difference is very slight. This yields our next main result, that *nearly optimal decisions are produced by the simple operation of linear integration over time for the MIP model of spike correlations across neural populations.*

Having established this, we pause to note a subtlety in our analysis. Figures **??**B and C show FC and DT as a function θ, for both simulated data and plots of Equations 5.3.1 and 5.3.2. The solid lines are the graphs of those equations as written (using the values for h_0 and $E[W]$ in Equations 4.3.4 and 4.3.3), and the mismatch between the lines and the data are a consequence of overshoot past the threshold. The broken line is a graph of the same formulas, with a shift in $\theta \rightarrow \theta + 14.5$, an offset computed as the sample mean of the overshoot distribution (See Figure 4.6 as well as the discussion in Section 4.9; also [40, 61]). This correction term helps the FC and DT equations better approximate the data when there is potential overshoot. Interestingly, however, parametric plots like Figure **??**A already take this effect into account.

4.4 Additive (SIP) correlations and decision making performance

4.4.1 The SPRT decision making model

As described in Section 4.2.4, the additive (SIP) model of spike train correlations also utilizes a common spike train to generate correlations, but does so in a manner that gives a distinct population-wide correlation structure. We now derive the consequences for decision making performance under the SPRT. In Appendices 4.7.1 and 4.7.3 we find the expressions for the parameters of the *FC* and *DT* curves, as the window size $\Delta T \to 0$:

$$h_0 = -1 \tag{4.4.1}$$

$$E[W] = (N(1-\rho)+\rho)(\lambda_p - \lambda_n)\log\frac{\lambda_p}{\lambda_n}\Delta T + O(\Delta T^2) \tag{4.4.2}$$

Comparing these with Equations 4.2.16 and 4.2.17, we see that, as in the subtractive (MIP) correlations model, the only difference with the independent case is a scaling factor on the average increment $E[W]$ in Equation 4.4.2. To explain the form of the scale factor, note that the spike vector from each pool is composed of N independent spike trains firing at rate $\lambda(1-\rho)$, and a single (highly redundant) spike train firing at a rate $\lambda\rho$.

As in the subtractive (MIP) model, E_n here also becomes a discrete random walk with increment $\pm\log(\lambda_p/\lambda_n)$. This can be seen by noting that for either pool, in a sufficiently small ΔT window, only one of two events is possible: (i) no spikes occur at all, or (ii) a single spike occurs in one neuron, in one of the two pools. The first case is uninformative about either H_1 or H_0. The second case occurs with probability $\lambda(1-\rho)$ under H_1 and $\lambda(1-\rho)$ under H_0 (Here $\lambda = \lambda_p$ if the spike occurred in the preferred pool, for example); taking the log ratio, we find our increment is independent of correlations. The resulting decision accuracy (FC) is plotted vs. threshold in Figure 4.4A, and is qualitatively similar to the subtractive (MIP) correlations case, with plateaus following from the discrete nature of E_n. However, the speed-accuracy tradeoff pictured in Figure 4.4B is very different from that found in the subtractive (MIP) model.

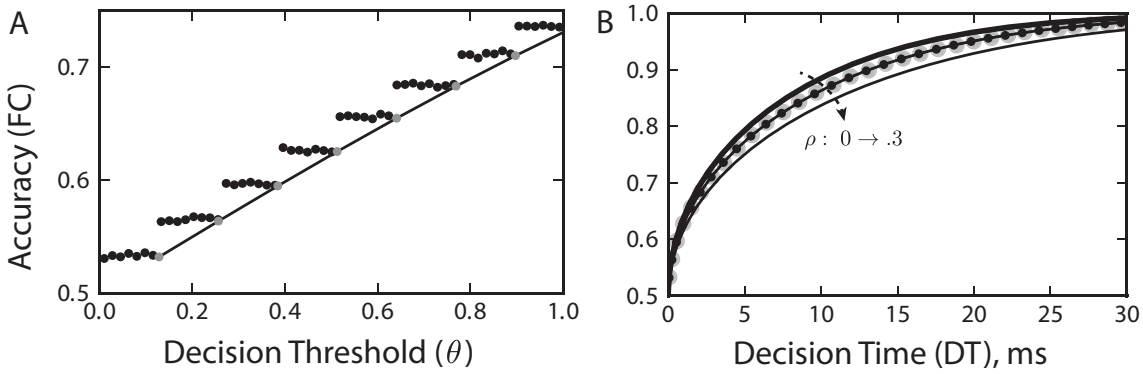


Figure 4.4: Additive (SIP) correlations do *not* significantly diminish decision performance under the SPRT. (A) The discrete diffusion with increment $\pm\log(\lambda_p/\lambda_n) \approx .128$ gives the same accuracy as the subtractive (MIP) correlations case (Figure 4.2A) *at each value of* θ. Because of the absence of overshoot, the FC and DT relationships can be applied exactly. (B) However, the resulting speed-accuracy curves are very different. In particular the impact of correlations on the speed-accuracy tradeoff is much smaller than for subtractive correlations (cf. Figure 4.2B, noting that here the abscissa ranges up to 30 ms, in contrast to 800 ms). Here only $\rho = 0, 0.15$, and 0.3 are plotted for clarity.

In particular, we see our third main result: *the impact of additive correlations on optimal (SPRT) decision performance is relatively minor.* For example, in the presence of pairwise correlations as strong as $\rho = .3$, the mean decision time required to reach a typical value of accuracy is increased by only a few milliseconds compared with the independent case, instead of by hundreds of milliseconds as for subtractive correlations. Equation 4.4.2 offers an intuitive explanation for this fact: $E[W]$ is inversely proportional to DT, and does not diminish nearly as fast for SIP correlations than MIP correlations (cf. Equation 4.3.2).

4.4.2 The spike integration decision making model

What about the ability of the simple spike integrator to perform decision making when confronted with additive correlations? Proceeding as in the subtractive-correlations case, we derive an implicit relationship for h_0, and the expected incre-

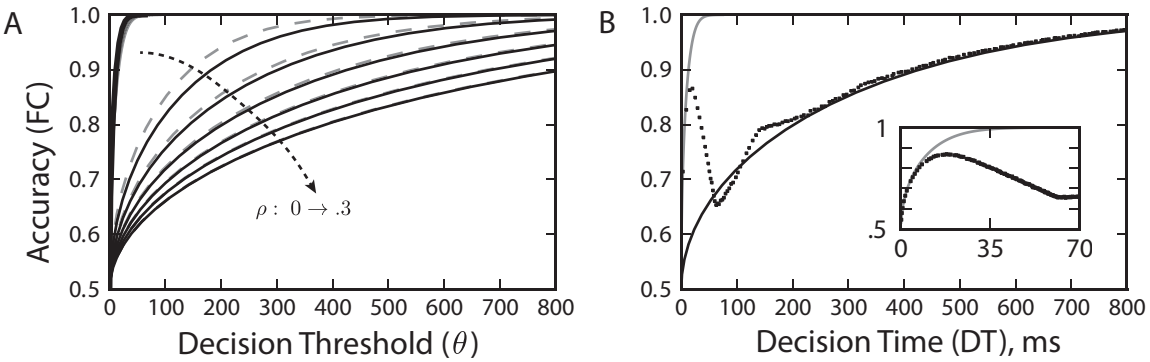


Figure 4.5: Decision making performance of the spiking integrator model with additive (SIP) correlations is comparable to subtractive correlations: correlations significantly decrease performance. (A) Black lines give the speed accuracy tradeoff predicted using h_0 and $E[W]$ from Equations 4.4.3 and 4.4.4 (and thereby assuming no overshoot of the decision threshold). Performance is similar to the subtractive-correlations case (broken gray lines), and significantly worse than performing SPRT on additive-correlated inputs (solid gray lines). (B) At $\rho = .15$, for example, major differences arise between this theory (again, solid black line, reproduced from A) and simulation of the model (dots), especially at short reaction times. This is a consequence of significant overshoot of E_n over the decision threshold, on the threshold crossing step. (Inset) At short reaction times, the simulations actually perform closer to the SPRT (gray line, reproduced from Figure 4.4A); see text.

ment $E[W]$:

$$\lambda_p(\rho(e^{Nt}-1)+(1-\rho)N(e^{t}-1))+\lambda_n(\rho(e^{-Nt}-1)+(1-\rho)N(e^{-t}-1))=0 \iff h_0 = t \tag{4.4.3}$$

$$E[W] = \Delta TN\left(\lambda_p - \lambda_n\right) \tag{4.4.4}$$

By comparing with (Equation 5.3.4), we see that, as for spike integration in the subtractive (MIP) case, correlation affects only the value of h_0 and not the expected increment. Substituting these values into Equations 5.3.1 and 5.3.2, we then plot the speed-accuracy tradeoff curves for this model *under the assumption of no overshoot* in Figure 4.5A. It appears that, when decisions are made via spike integration, correlations impact performance quite significantly (black lines), in contrast to the SPRT case (solid gray lines, reproduced from Figure 4.4B). Overall, the degree of performance loss is comparable to that under subtractive correlations (broken gray lines, reproduced from Figure **??**B). This is our fourth main result: *for additive correlations, if decisions are made via spike integration instead of the SPRT, correlations have a significant impact on reducing decision performance.*

However, the assumption that integrated spikes do not overshoot the decision threshold might seem suspect under the additive model of correlations, as there is a possibility that the threshold crossing step might occur as a result of every neuron in a pool simultaneously spiking at once. In fact, when the number of neurons in the pool is large (as in the cases we consider), additive correlations can indeed cause significant overshooting of thresholds; importantly, and unlike for subtractive (MIP) correlations, this effect cannot be compensated via a constant offset in the decision threshold.

Figure 4.5B demonstrates the consequences for the speed-accuracy tradeoff. Here, when the spike integration model is simulated directly, we see a surprising non-monotonic relationship between FC and DT in the presence of additive correlations of strength $\rho = .15$. This violates the usual intuition of that accuracy should increase at slower decision speeds. The explanation comes from the fact that, as the decision threshold is raised increases, DT correspondingly increases while accuracy suffers – a consequence of not finishing a trial before a (relatively rare) spike in a correlating spike train in one of the two pools causes the accumu-

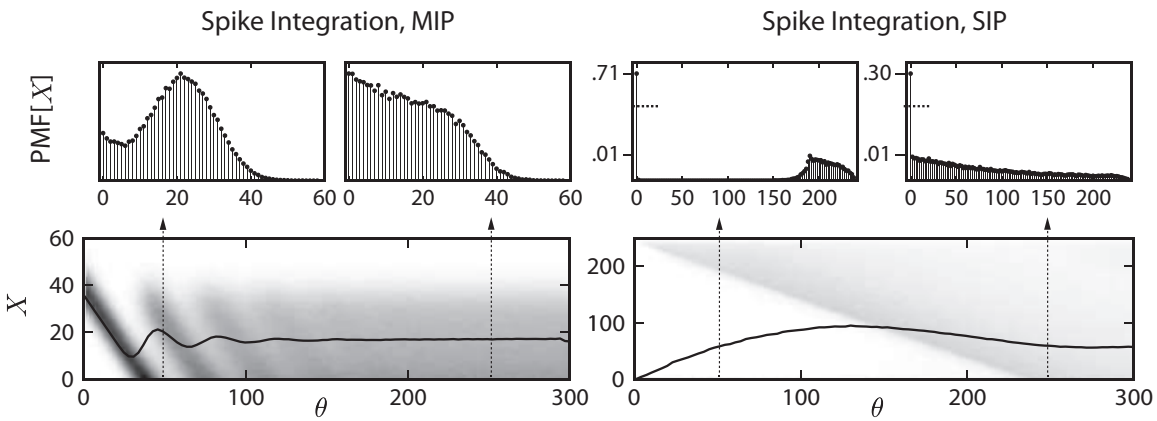


Figure 4.6: Overshoot distributions for spike integration under additive (SIP) and subtractive (MIP) correlations. The random variable X indicates the distribution of $E_n - \theta$ conditioned on crossing the upper threshold (similar results for the lower threshold are not shown). The probability mass function (PMF) of X varies as a function of θ, and two vertical slices through this density are shown at $\theta = 50$ and 250. Here the overshoot distributions are discrete, due to the integral nature of the increment distribution. For plotting purposes, the vertical axis has been split in the SIP case, to allow plotting of the outlier point at zero. The black line indicates $E[X]$ as θ varies; crucially, this quantity varies significantly and for higher values of θ under SIP correlations, resulting in the non-monotonic speed-accuracy tradeoff pictured in Figure 4.5.

lator to jump far beyond the threshold.

For large thresholds, the sequential sampling theory of Equations 5.3.1 and 5.3.2, which assume no overshoot, accurately approximates the simulated data; however for low values of θ the approximation is poor. In fact, the inset to Figure 4.5B shows that in this regime, the decision making performance of the spike integration model is far better described by the theory predicted by the SPRT. The intuition behind this observation is that for short reaction times, there is a small probability of a shared spike that will send the integrator significantly over the threshold. This allows accumulation to occur one spike at a time (for sufficiently small ΔT), where each spike arrives from an independent spike train. As we have seen, the process of integrating independent spikes is equivalent to the SPRT. It is only at longer decision times, when the chances of having integrated a large common spike event are larger, that a significant impact of correlations appears.

Figure 4.7 provides further evidence for this scenario. Density plots of the dis-

tribution of the overshoot $X = E_x - \theta$ (conditioned on crossing the upper threshold), for both additive (SIP) and subtractive (MIP) correlations are shown as a function of the decision threshold, with particular overshoot distributions plotted at $\theta = 50$ and 250. For the additive correlations model, a significant fraction of the trials terminate with zero overshoot at low values of θ (because, for example, large correlating events are relatively rare), implying that many trials underwent optimal accumulation of evidence, without experiencing a common, correlating spike event as discussed above.

Overall, the monotonic dependence of accuracy (FC) on decision time (DT) follows from the invariance of the moments of the overshoot distribution relative to changes in the threshold value θ; this is particularly true for the first moment (See Section 4.9). Figure 4.7(SIP) demonstrates that these moments continue to fluctuate over a larger range of θ, and with larger magnitude, for the additive correlations model. This serves to explain the strange shape of the speed-accuracy tradeoff curve pictured in Figure 4.5B that (unlike the subtractive correlations model) cannot be explained by a constant shift in θ.

4.5 Nonlinear computations and optimal performance via the SPRT

When the neurons in each pool spike independently, Zhang and Bogacz [133] demonstrated that linear summation of spikes across the two pools at each time step implements the SPRT. Because the SPRT is optimal in the sense of minimizing DT for a prescribed level of FC, the conclusion is that linear integration of spikes across pools, and then across time, provides an optimal decision making strategy. However, is this optimality of linear integration confined to the case of independent activity within the pool?

Above, we showed that when correlations are introduced into this model, it is no longer true that each spike should be given the same “weight", as in linear integration. Moreover, knowing only the pairwise correlations and firing rates alone does not allow one to write down a rule for the function that should be

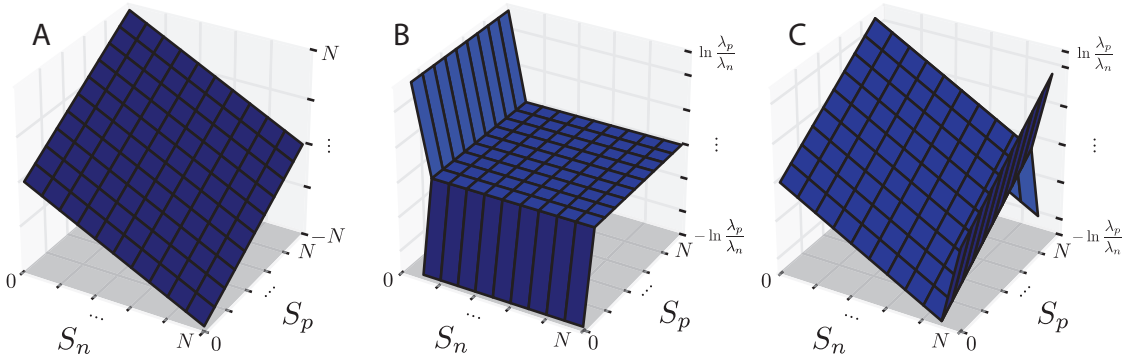


Figure 4.7: Increments for the SPRT are nonlinear when input spikes are correlated. (A) For both additive and subtractive correlations, the spike integration model of decision making implies a linear mapping between the number of spikes in the preferred and null populations, and the increment to the accumulator. (B) With subtractive correlations, a severe nonlinearity means that only increments of $\pm \log(\lambda_p/\lambda_n)$ occur. This stands in direct contrast to the optimality of linear summation in the zero-correlations case. (C) A nonlinear computation also appears as a consequence of the additive correlations model, however the nonlinearity is much less severe than in the subtractive model. (All results pictured hold in the case of vanishing ΔT.)

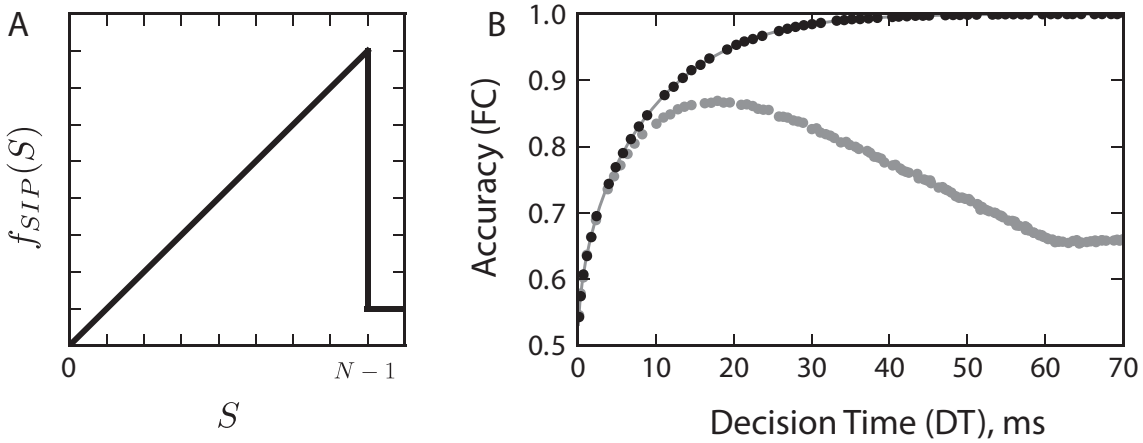


Figure 4.8: Optimal performance via spike integration under additive correlations can be realized with a simple nonlinearity. (A) A nonlinearity discounts the contribution to the accumulator of a shared spike event (See Equation 4.5.3). (B) Spike integration with this nonlinearity is suggested by Figure 4.7C, and recovers performance of the decision making model (black dots) to agreement with the results of SPRT (gray line). Without this nonlinearity to discount shared events, performance suffers (gray dots, reproduced from Figure 4.5B, Inset)

applied to incoming spikes in order to implement the SPRT, although in these cases this function takes the form of the difference between the result of a nonlinearity applied to both pools. This dependence on higher order statistics is demonstrated in Figure 4.7 by the fact that the nonlinearities for MIP correlations (Panel B) and SIP correlations (Panel C) take a significantly different form.

For MIP correlations, the nonlinearity pictured in Figure 4.7B that implements the SPRT (up to a change in threshold) takes the form:

$$W_i = f_{MIP}(\boldsymbol{S}_p^i) - f_{MIP}(\boldsymbol{S}_n^i) \tag{4.5.1}$$

$$f_{MIP}(\boldsymbol{S}) = \begin{cases} 1 & : \sum_{k=1}^{N} S_k^i \geq N \\ 0 & : \sum_{k=1}^{N} S_k^i = 0 \end{cases} \tag{4.5.2}$$

At first glance, it is surprising that such a severe nonlinearity, applied to two MIP-correlated spiking pools, results in nearly the same performance is simple spike integration (c.f. Figure **??**). The intuition here is that optimal inference requires essentially performing spike integration on the correlating spike train, as no information about the firing rate is added through spike deletions. This random walk on one of three cases (-1,0, or +1) is approximated by linear integration, in the limit as the size of the pool (N) increases.

Another perspective on the nonlinearities that enable optimal computation is that they leverage knowledge about the mechanism of correlations, to improve performance. In the SIP model, the nonlinear function depicted in Figure 4.7C is, as in the MIP case, a consequence of applying a nonlinearity to each pool, and then subtracting. However, in this case, the form is not as drastic—a shared spike event coming from the correlating train only registers as a single spike:

$$W_i = f_{SIP}(\boldsymbol{S}_p^i) - f_{SIP}(\boldsymbol{S}_n^i) \tag{4.5.3}$$

$$f_{SIP}(\boldsymbol{S}) = \begin{cases} \sum_{k=1}^{N} S_k^i & : \sum_{k=1}^{N} S_k^i < N \\ 1 & : \sum_{k=1}^{N} S_k^i = N \end{cases} \tag{4.5.4}$$

Intuitively, this strategy uses the fact the a simultaneous spike in every neuron in a pool only has one explanation for a sufficiently small window of integration,

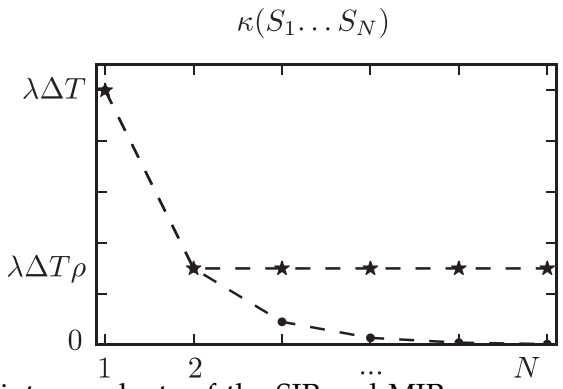


Figure 4.9: The joint cumulants of the SIP and MIP processes differ for pools of greater than two neurons. Under the additive (SIP) model, the joint cumulants of the spike counts from N neurons is constant for all $N > 2$. In contrast, the joint cumulants of the subtractive (MIP) model decay geometrically as the pool size increases, and this difference helps to characterized the differences in higher-order correlations between the two models. (See Section 4.10 for supplementary computations.)

and therefore uses the correlating spike train as an additional independent input in the likelihood ratio. At low values of θ, this does not confer much of an advantage; however as the threshold increases, higher accuracy is achievable at much shorter decision times. The nonlinearity is pictured Figure 4.8A, and also offers an intuition as to why, for low threshold values, spike integration performs almost optimally: when spikes from the correlating train are rare (or can be properly weighted), spike integration implements SPRT (Figure 4.8B).

4.6 Discussion

Correlated spiking among the neurons that encode sensory evidence appear ubiquitous. Such correlations might arise arise from any number of neuroanatomical features – the simplest being overlapping feedforward connectivity which can cause collective fluctuations across a population [6, 105, 26, 70]. They can also result from sensory events that impact an entire population, or from rapid modulatory effects. Moreover, for large neural populations it appears that accurate descriptions of population-wide activity can require more than the typically measured pairwise correlations, but higher order interactions as well [77, 38, 131].

The aim of our study is to improve our understanding of how correlated ac-

tivity in these populations can impact the speed and accuracy of decisions that require accumulating sensory information over time. Faced with the wide range of possible mechanisms and structures of correlations alluded to above, we choose to focus on two models for population-wide correlations that illustrate a key distinction in how correlations can occur. These models have identical first-order and pairwise statistics, but differ in how each common spiking events either involves a small subset of the neurons (the subtractive, MIP case) or each neuron in the pool (the additive, SIP case) [58, 116].

Figure 4.9 quantifies this difference: based on calculations in Section 4.10, we plot the joint cumulant across k neurons in a pool under both subtractive (MIP) and additive (SIP) correlations. While the additive model possesses a constant joint cumulant no matter how many neurons are included, the joint cumulant of k neurons falls off geometrically for the subtractive case. We conjecture that this is a statistical signature that could suggest when other, more general patterns of correlated activity – measured experimentally or arising in mechanistic models of neural circuits [70] – will produce similar effects on decisions. Exploring this conjecture via models and data is a target of our future research.

We summarize our main findings are as follows. For both models of correlated spiking, decisions produced by a simple, linear spike integration model (i.e., a neural integrator) become slower and less accurate as correlations increase. However, a strong difference appears for decisions made via the optimal decision strategy (SPRT). Here, additive correlations have only a minor impact on decision performance, while subtractive correlations continue to strongly diminish this performance. The conclusion is that decision making circuits, faced with subtractive (MIP) correlated sensory populations, will invariably produce diminished decision performance, and stand little to gain by implementing computations more complex that a simple integration of spikes over time and neurons. However, in the presence of additive (SIP) correlations, circuit mechanisms that implement or approximate the SPRT – perhaps via a nonlinearity such as that shown in Fig. 4.8 applied to the sum of incoming spikes – stand to produce substantially better decision performance than their linear counterparts.

In other contexts, nonlinear computations have also been shown to improve

discrimination between two alternatives. Field and Rieke [34, 35] demonstrated the importance of a thresholding nonlinearity in pooling the responses of rod cells, where this nonlinearity served to reject “background" noise. Closer to the present setting, gating inhibition that prevents accumulation of noise samples before the onset of evidence-encoding stimulus can account for visual search performance [85], and recent results suggest that related nonlinearities can improve performance for mistuned neural integrators ([16], see also [17]).

Our cases in which correlations decrease performance – in particular, when spikes are linearly integrated – are consistent with several prior studies of the role of correlated activity in decision making [134, 11, 23]. We note, however, two differences in our models. The first is the mechanism through which correlated spikes are generated; while we use additive and subtractive models based on Poisson processes, the authors of [11, 23] use a multivariate Gaussian description of spike counts. The second is that in [11, 23], decisions are rendered after a duration that is fixed before the trial begins (either a single duration, [11], or one that is drawn from a distribution of reaction times, in [23]). This is different from the setting here, where incoming signal on each trial determines the reaction time through a bound crossing.

Our result, in the case of subtractive (MIP) correlations, that linear integration of spikes closely approximates the optimal decision making strategy is similar to findings of Beck et al. [4]. Specifically, they model a dense range of differently tuned populations, and find that optimal Bayesian inference can be based on linear integration of inputs, for a wide set of correlation models. Our additive (SIP) case, however, behaves differently, as nonlinearities are needed to achieve the optimal strategy.

An aim of future work is extending the setting of our study to include tuning curves as in [23, 4, 133, 23]. This is more realistic for many decision tasks (including the direction discrimination task), and will also allow progress toward models with multiple decision alternatives. An important challenge will come from defining pairwise correlations that vary as a function of preferred tuning orientation (see [134, 22]), while also including the full structure of correlating events across multiple cells in a realistic way. For example, in the present paper, additive cor-

relating events occurred independently in the two populations; future work could take a more graded approach, in which some events impact the entire sensory population (i.e., as in an eyeblink or possibly an attentional shift during a visual task).

As long as each neuron remains modeled as a Poisson point process, the sequential accumulation theory utilized here will carry over directly. This points to another limitation of the present study and opportunity for future work. This is the lack of temporal correlations in the statistics of the inputs. A model of correlations that includes spikes from a correlating train that are temporally jittered [48, 116] could provide a starting place for a model of the input trains, however defining updates to the likelihood ratio for the two competing hypotheses will be more difficult. Nevertheless, it will be interesting to see how our results carry over; in particular, there will be many more different combinations of spike events that will contribute to increments for both spike integration and SPRT decision models.

While we therefore view the present study as a first step in exploring many possibilities, our findings demonstrate how the population-wide structure of correlations – beyond pairwise correlation coefficients – can strongly impact the speed and accuracy of decisions, and the circuit operations necessary to achieve optimal performance. This suggests that multi-electrode and imaging technologies, together with theoretical work on neural coding, will continue to play an exciting role in understanding the structure of basic computations like decision making over time.

4.7 Sequential Probability Ratio Test

4.7.1 Nontrivial root of the moment generating function (SPRT)

The nontrivial real root of the moment generating function (MGF) of a sampling distribution is critical to finding FC and DT of an independently sampled sequential hypothesis test (via Equations 5.3.1 and 5.3.2). For the SPRT, the increment distribution is given in Equation 4.2.15 as:

$$W_i = \log\left[\frac{P[\boldsymbol{S}_p^i, \boldsymbol{S}_n^i|H_1]}{P[\boldsymbol{S}_p^i, \boldsymbol{S}_n^i|H_0]}\right] \tag{4.7.1}$$

The "correct" hypothesis H_1 is in the numerator in order to orient a crossing of the positive decision threshold with a correct choice. Correspondingly, the probability of observing a *given* sample $\boldsymbol{S}_p^i$,$\boldsymbol{S}_n^i$ is known from assumption of this hypothesis, and by definition follows the distribution:

$$P[\boldsymbol{S}_p^i, \boldsymbol{S}_n^i|H_1] = P[\boldsymbol{S}_p^i|H_1]P[\boldsymbol{S}_n^i|H_1], \tag{4.7.2}$$

where the independence assumption of the spike count vectors from the two separate pools $\boldsymbol{S}_p^i$ and $\boldsymbol{S}_n^i$ have allowed the factoring of the distribution. Dropping the sampling index i for notational convenience, the MGF can then be computed as:

$$\phi_W(s) = E[e^{sW}] = \sum_{\boldsymbol{s}_p,\boldsymbol{s}_n} P[\boldsymbol{s}_p|H_1]P[\boldsymbol{s}_n|H_1] \left(\frac{P[\boldsymbol{s}_p|H_1]P[\boldsymbol{s}_n|H_1]}{P[\boldsymbol{s}_p|H_0]P[\boldsymbol{s}_n|H_0]}\right)^s \tag{4.7.3}$$

The nontrivial root ($h_0 \neq 0$) can then be seen by inspection (cf. Equation 4.2.16):

$$s = h_0 = -1 \quad \Longrightarrow \quad \phi_W(h_0) = 1 \tag{4.7.4}$$

We note that this computation is fully general, without any assumptions on the structure of correlations both within and across pools.

4.7.2 $E[w]$, Independent interactions (SPRT)

The other parameter of the sampling distribution critical to computing the FC and DT functions, $E[W]$, is computed for independent spike count distributions ($\rho = 0$, cf. 4.2.17) as follows (see also [133]):

$$E[W] = \sum_{\boldsymbol{s}_p,\boldsymbol{s}_n} P[\boldsymbol{s}_p|H_1]P[\boldsymbol{s}_n|H_1] \log\left[\frac{P[\boldsymbol{s}_p|H_1]P[\boldsymbol{s}_n|H_1]}{P[\boldsymbol{s}_p|H_0]P[\boldsymbol{s}_n|H_0]}\right] \tag{4.7.5}$$

$$= \sum_{\boldsymbol{s}_p,\boldsymbol{s}_n} P[\boldsymbol{s}_p|H_1]P[\boldsymbol{s}_n|H_1] \left(\log P[\boldsymbol{s}_p|H_1] + \log P[\boldsymbol{s}_n|H_1] - \log P[\boldsymbol{s}_p|H_0] - \log P[\boldsymbol{s}_n|H_0]\right) \tag{4.7.6}$$

$$= \sum_{\boldsymbol{s_p}} P[\boldsymbol{s_p}|H_1] \log P[\boldsymbol{s_p}|H_1] + \sum_{\boldsymbol{s_n}} P[\boldsymbol{s_n}|H_1] \log P[\boldsymbol{s_n}|H_1] \tag{4.7.7}$$

$$- \sum_{\boldsymbol{s_p}} P[\boldsymbol{s_p}|H_1] \log P[\boldsymbol{s_p}|H_0] - \sum_{\boldsymbol{s_n}} P[\boldsymbol{s_n}|H_1] \log P[\boldsymbol{s_n}|H_0] \tag{4.7.8}$$

$$= N \left(E\left[\log \frac{P[S_p|H_1]}{P[S_p|H_0]} | H_1 \right] + E\left[\log \frac{P[S_n|H_1]}{P[S_n|H_0]} | H_1 \right] \right) \tag{4.7.9}$$

$$= N\Delta T \left(\lambda_n - \lambda_p + \lambda_p \log \frac{\lambda_p}{\lambda_n} + \lambda_p - \lambda_n + \lambda_n \log \frac{\lambda_n}{\lambda_p} \right) \tag{4.7.10}$$

$$N\Delta T(\lambda_p - \lambda_n) \log \frac{\lambda_p}{\lambda_n} \tag{4.7.11}$$

When this quantity is substituted into Equation 5.3.2, the ΔT will cancel off, implying that DT is not a function of the sampling increment size. We compute this quantity for correlated spike count distributions next.

4.7.3 $E[W]$, additively (SIP) correlated interactions (SPRT)

When neurons within pools are correlated, the joint PDF of the spike count vector is no longer decomposable into the product of the marginal distributions (the critical step between Equations 4.7.8 and 4.7.9). However, an expression for $E[W]$ can be obtained in the limit as $\Delta T \to 0$, by repeatedly expanding via Taylor series about $\Delta T = 0$ throughout the computation.

First, we simplify the expression for the expected increment by using the independence of the two pools:

$$E[W] = E\left[\log \frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]} | H_1 \right] + E\left[\log \frac{P[\boldsymbol{S_n}|H_1]}{P[\boldsymbol{S_n}|H_0]} | H_1 \right] \tag{4.7.12}$$

Next we expand each term to first order in ΔT; below, we only demonstrate the expansion for the “preferred" population; the calculation for the null pool follows by exchanging λ_p and λ_n. In that case, by using the Law of Total Expectation conditioned on the number of spikes in the common spike train “shared" across

the pool $\hat{\boldsymbol{S}}_{\boldsymbol{p}}$ (which spikes at a rate $\rho\lambda_p\Delta T$), we have:

$$E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|H_1\right] = E\left[E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p, H_1\right]|H_1\right] \tag{4.7.13}$$

$$= \sum_{\hat{s}_p=0}^{\infty} P[\hat{S}_p = \hat{s}_p]E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = \hat{s}_p, H_1\right] \tag{4.7.14}$$

$$= (1-\rho\lambda_p\Delta T)E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 0, H_1\right] + \Delta T\lambda_p E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 1, H_1\right] + O(\Delta T^2) \tag{4.7.15}$$

Taking the case of $\hat{S}_p = 0$,

$$E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 0, H_1\right] = \sum_{\boldsymbol{s_p}} P[\boldsymbol{s_p}|\hat{S}_p = 0, H_1]\log\frac{P[\boldsymbol{s_p}|H_1]}{P[\boldsymbol{s_p}|H_0]} \tag{4.7.16}$$

The aim here is to take advantage of the conditioning; because the spike counts of neurons within the same pool are conditionally independent, given the number of spikes in the correlating spike train, the joint distribution across the vector $\boldsymbol{s_p}$ becomes the product of the conditioned marginal distributions. However, this is only true for the first factor in the summand of Equation 4.7.16. To continue, we must expand the log-ratio of the probability distributions, using the law of total probability, in ΔT:

$$P[\boldsymbol{s_p}|H_1] = \sum_{\hat{s}_p}^{\infty} P[\hat{S}_p|H_1]P[\boldsymbol{s_p}|\hat{S}_p = \hat{s}_p, H_1] \tag{4.7.17}$$

$$= (1-\rho\lambda_p\Delta T)P[\boldsymbol{s_p}|\hat{S}_p = 0, H_1] + \rho\lambda_p\Delta T P[\boldsymbol{s_p}|\hat{S}_p = 1, H_1] + O(\Delta T^2) \tag{4.7.18}$$

$$P[\boldsymbol{s_p}|H_0] = \sum_{\hat{s}_p}^{\infty} P[\hat{S}_p|H_0]P[\boldsymbol{s_p}|\hat{S}_p = \hat{s}_p, H_0] \tag{4.7.19}$$

$$= (1-\rho\lambda_n\Delta T)P[\boldsymbol{s_p}|\hat{S}_p = 0, H_0] + \rho\lambda_n\Delta T P[\boldsymbol{s_p}|\hat{S}_p = 1, H_0] + O(\Delta T^2) \tag{4.7.20}$$

Moreover, the $N-$term summation in Equation 4.7.16 need only be over $s_i \in \{0, 1\}$, as higher values will produce contributions of higher than first order in ΔT. Two cases emerge for the expansion: if $s_i = 0$ for any i, $P[\boldsymbol{s_p}|\hat{S}_p = 1, H_1] = P[\boldsymbol{s_p}|\hat{S}_p =$

$1, H_0] = 0$, and we have:

$$\log \frac{P[\boldsymbol{s_p}|H_1]}{P[\boldsymbol{s_p}|H_0]} = \log \frac{P[\boldsymbol{s_p}|\hat{S}_p = 0, H_1]}{P[\boldsymbol{s_p}|\hat{S}_p = 0, H_0]} \tag{4.7.21}$$

On the other hand, if $s_i = 1$ for all i, we can compute the expression directly via total probability, as there are only four possible ways for the event to originate; to first-order in ΔT, this is:

$$\begin{aligned} \log \frac{P[\boldsymbol{s_p} = \mathbf{1}|H_1]}{P[\boldsymbol{s_p} = \mathbf{1}|H_0]} &= \sum_{i=0}^{1} \sum_{\hat{s}_p=0}^{1} P[\boldsymbol{s_p} = \mathbf{1}|\hat{S}_p = \hat{s}_p, H_i] P[\hat{S}_p = \hat{s}_p, H_i] & (4.7.22) \\ &= \rho(\lambda_p + \lambda_n)\Delta T + O(\Delta T^N) & (4.7.23) \end{aligned}$$

Therefore, this single element of the sum offers no order one contribution (it is multiplied by $P[\boldsymbol{s_p} = \mathbf{1}|\hat{S}_p = 0, H_1]$ which is itself is $\mathrm{O}(\Delta T^N)$); thus,

$$\sum_{\boldsymbol{s_p}} P[\boldsymbol{s_p}|\hat{S}_p = 0, H_1] \log \frac{P[\boldsymbol{s_p}|\hat{S}_p = 0, H_1]}{P[\boldsymbol{s_p}|\hat{S}_p = 0, H_0]} = N(1-\rho)\Delta T \left(\lambda_n - \lambda_p + \lambda_p \log \frac{\lambda_p}{\lambda_n}\right) + O(\Delta T^2) \tag{4.7.24}$$

The case of $\hat{s}_p = 1$ is simpler, as only zero-order terms must be kept (due to the coefficient in Equation 4.7.15). Recycling the expansion from Equation 4.7.21, we have that to zero-order:

$$E\left[\log \frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 1, H_1\right] = \sum_{\boldsymbol{s_p}} P[\boldsymbol{s_p}|\hat{S}_p = 1, H_1] \log \frac{P[\boldsymbol{s_p}|H_1]}{P[\boldsymbol{s_p}|H_0]} = \log \frac{\lambda_p}{\lambda_n} + O(\Delta T) \tag{4.7.25}$$

Finally, combining Equations 4.7.15, 4.7.24, and 4.7.25, we have that:

$$\begin{aligned} E\left[\log \frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|H_1\right] &= (1 - \rho\lambda_p\Delta T)\left(N(1-\rho)\Delta T\left(\lambda_n - \lambda_p + \lambda_p \log \frac{\lambda_p}{\lambda_n}\right) + O(\Delta T^2)\right) & (4.7.26) \\ &\quad + \Delta T \rho \lambda_p \left(\log \frac{\lambda_p}{\lambda_n} + O(\Delta T^2)\right) & (4.7.27) \end{aligned}$$

Repeating the exercise for the other component of Equation 4.7.12 amounts to exchanging "p" for "n"; adding everything together gives the final result, to first-

order in ΔT:

$$E[W] = (N(1-\rho)+\rho)\,\Delta T(\lambda_p - \lambda_n)\log\frac{\lambda_p}{\lambda_n} + O(\Delta T^2) \tag{4.7.28}$$

We note here that as $\rho \to 0$ and $\rho \to 1$, we reproduce the results that would be expected from Equation 4.7.11. Also, a more intuitive and tractable computation can be done for an analogous additively-correlated Bernoulli process, resulting in the same solution.

4.7.4 $E[W]$, subtractive (MIP) correlations within pools (SPRT)

In the case of subtractive correlations within pools, the derivation of $E[W]$ is the same as the additive correlation case, up to Equation 4.7.14. In this case, however, we now have:

$$E[W] = (1-\frac{\lambda_p}{\rho}\Delta T)E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 0, H_1\right] + \frac{\lambda_p}{\rho}\Delta T E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 1, H_1\right] + O(\Delta T^2) \tag{4.7.29}$$

Taking the $\hat{S}_p = 0$ case first, we notice that it is impossible for any spikes to occur without a spike in the correlating spike train:

$$\boldsymbol{S_p} \neq \boldsymbol{0} \implies P[\boldsymbol{S_p}|\hat{S}_p = 0, H_1] = 0 \tag{4.7.30}$$

Because of this, we can simplify:

$$E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p = 0, H_1\right] = P[\boldsymbol{0}|\hat{S}_p = 0, H_1]\log\frac{P[\boldsymbol{0}|H_1]}{P[\boldsymbol{0}|H_0]} \tag{4.7.31}$$

$$= \log\frac{P[\boldsymbol{0}|H_1]}{P[\boldsymbol{0}|H_0]} \tag{4.7.32}$$

Interestingly, after conditioning on the number of correlating spikes, the probability of the zero vector (or any vector $\boldsymbol{s_p}$) is the same under both H_0 and H_1:

$$P[\boldsymbol{0}|\hat{S}_p = \hat{s}_p, H_0] = P[\boldsymbol{0}|\hat{S}_p = \hat{s}_p, H_1] \tag{4.7.33}$$

We then expand to first-order in ΔT:

$$\log \frac{P[\mathbf{0}|H_1]}{P[\mathbf{0}|H_0]} = \log \frac{\left(1-\frac{\lambda_p}{\rho}\Delta T\right)+\frac{\lambda_p}{\rho}\Delta T P[\mathbf{0}|\hat{S}_p=1]+O(\Delta T^2)}{\left(1-\frac{\lambda_n}{\rho}\Delta T\right)+\frac{\lambda_n}{\rho}\Delta T P[\mathbf{0}|\hat{S}_p=1]+O(\Delta T^2)} \quad (4.7.34)$$

$$= \frac{(\lambda_p-\lambda_n)\Delta T}{\rho}\left((1-\rho)^N-1\right)+O(\Delta T^2) \quad (4.7.35)$$

In the case of $\hat{S}_p = 1$, only zero-order terms must be computed. When computing

$$E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p=1,H_1\right]=\sum_{\boldsymbol{s_p}}P[\boldsymbol{s_p}|\hat{S}_p=1,H_1]\log\frac{P[\boldsymbol{s_p}|H_1]}{P[\boldsymbol{s_p}|H_0]}, \quad (4.7.36)$$

the summation only carries over $\{0,1\}$ for each element of $\boldsymbol{s_p}$. The case of $\boldsymbol{s_p}=\mathbf{0}$ provides no contribution at zero-order, as can be seen by Equation 4.7.35; for any other case, there will be a degeneracy in the expansion of the log, caused by an absence of order 0 terms:

$$\boldsymbol{s_p}\neq 0 \implies \log\frac{P[\boldsymbol{s_p}|H_1]}{P[\boldsymbol{s_p}|H_0]} = \log\frac{\left(\frac{\lambda_p}{\rho}-\frac{\lambda_p^2}{\rho^2}\Delta T\right)P[\boldsymbol{s_p}|\hat{S}_p=1]+\frac{\lambda_p^2}{2\rho^2}P[\boldsymbol{s_p}|\hat{S}_p=2]+\ldots}{\left(\frac{\lambda_n}{\rho}-\frac{\lambda_n^2}{\rho^2}\Delta T\right)P[\boldsymbol{s_p}|\hat{S}_p=1]+\frac{\lambda_n^2}{2\rho^2}P[\boldsymbol{s_p}|\hat{S}_p=2]+\ldots} \quad (4.7.37)$$

$$= \log\frac{\lambda_p}{\lambda_n}+O(\Delta T) \quad (4.7.38)$$

Therefore, to first-order in ΔT,

$$E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|\hat{S}_p=1,H_1\right] = \log\frac{\lambda_p}{\lambda_n}\sum_{\boldsymbol{s_p}\neq 0}P[\boldsymbol{s_p}|\hat{S}_p=1,H_1]+O(\Delta T) \quad (4.7.39)$$

$$= \log\frac{\lambda_p}{\lambda_n}(1-(1-\rho)^N)+O(\Delta T) \quad (4.7.40)$$

Combining Equations 4.7.29, 4.7.32, 4.7.35, and 4.7.40, we find that:

$$E\left[\log\frac{P[\boldsymbol{S_p}|H_1]}{P[\boldsymbol{S_p}|H_0]}|H_1\right]=\frac{\left(1-(1-\rho)^N\right)\left(\lambda_n-\lambda_p+\lambda_p\log\frac{\lambda_p}{\lambda_n}\right)}{\rho}\Delta T \quad (4.7.41)$$

As before, exchanging "p" for "n" takes care of the expression for the null pool, and adding together gives:

$$E[W] = \frac{\left(1-(1-\rho)^N\right)}{\rho}(\lambda_p - \lambda_n)\log\frac{\lambda_p}{\lambda_n}\Delta T + O(\Delta T^2) \tag{4.7.42}$$

Once again, as $\rho \to 0$ and $\rho \to 1$, we reproduce the results that would be expected from Equation 4.7.11.

4.8 Spike integration

4.8.1 Independent spiking (SI)

Computing FC and DT for the spike integration accumulation model relies on computation of the MGF for the sampling distribution. We begin with several identities that will be useful below. The MGF for the sum of N independent random variables is:

$$S = \sum_{i=1}^{N} S_i \quad \Longleftrightarrow \quad \phi_S(t) = \phi_{s_i}(t)^N \tag{4.8.1}$$

Given that the MGF for a random variable $S = \phi_S(t)$, it follows that

$$\phi_{-S}(t) = \phi_S(-t) \tag{4.8.2}$$

Finally, the MGF for a Poisson random variable is:

$$\phi(t) = e^{\lambda(e^t-1)} \tag{4.8.3}$$

Given the definition of the increment variable in Equation 4.2.14, and noting that each spike count random variable is independent, we can combine these observations to construct the MGF for the sampling random variable, over a time window ΔT:

$$\phi_W(t) = (e^{\lambda_p \Delta T(e^t-1)})^N (e^{\lambda_n \Delta T(e^{-t}-1)})^N \tag{4.8.4}$$

Now the nontrivial root can be calculated (cf. Equation 5.3.3):

$$h_0 = -\log\left(\frac{\lambda_p}{\lambda_n}\right) \quad \Longrightarrow \quad \phi_W(h_0) = 1 \tag{4.8.5}$$

Because the MGF is known explicitly, the computation of the expected increment is simple (cf. Equation 5.3.4):

$$E[W] = \phi'_W(0) = \Delta T N \left(\lambda_p - \lambda_n\right). \tag{4.8.6}$$

4.8.2 Additive (SIP) correlated interactions within pools (SI)

When additive correlations are introduced within pools, the spike count distribution MGF over a time period ΔT can still be broken into the product of two separate MGF's, one each for the preferred and null pools, which are identical in form but differ in their Poisson rate parameters (indicated by the semicolon):

$$\phi_W(t) = \phi_S(t; \lambda_p \Delta T)\phi_S(-t; \lambda_n \Delta T) \tag{4.8.7}$$

For the preferred pool, the spike count can be broken into two independent contributions—spikes $\hat{S}$ from the shared (i.e. correlating) spike train that get counted N times (firing at a rate $\rho\lambda$), and spikes from the N independent spike trains that get counted once (each firing at a rate $(1-\rho)\lambda$):

$$\phi_S(t; \lambda) = \phi_{\hat{S}}(t; \lambda)\phi_{S_i}(t; \lambda)^N \tag{4.8.8}$$

The MGF for the shared spike train can be computed directly from the definition, using its probability mass function (PMF):

$$P[\hat{S} = iN] = \begin{cases} \frac{e^{-\lambda}\lambda^i}{i!} & i \in \mathbb{N}_0 \\ 0 & \text{otherwise,} \end{cases} \tag{4.8.9}$$

and thus:

$$\phi_{\hat{S}}(t, \lambda) = \sum_{k=0}^{\infty} e^{tk} P[\hat{S} = k] = \sum_{k=0}^{\infty} e^{Ntk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{(e^{Nt}-1)\lambda} \tag{4.8.10}$$

The MGF for the independent spike trains $\phi_{S_i}(t,\lambda)$ follows from Section 4.8.1, giving the form of the MGF of the increment over a time ΔT as:

$$\phi_W(t) = e^{(e^{Nt}-1)\rho\lambda_p\Delta T} e^{(e^{-Nt}-1)\rho\lambda_n\Delta T} \left(e^{(e^t-1)(1-\rho)\lambda_p\Delta T}\right)^N \left(e^{(e^{-t}-1)(1-\rho)\lambda_n\Delta T}\right)^N \tag{4.8.11}$$

After rearranging, h_0 is implicitly defined as the nontrivial root of:

$$\lambda_p(\rho(e^{Nt}-1)+(1-\rho)N(e^t-1))+\lambda_n(\rho(e^{-Nt}-1)+(1-\rho)N(e^{-t}-1))=0 \implies h_0 = t \tag{4.8.12}$$

As $\rho \to 0$, we recover the solution from Section 4.8.1. The expected increment can be directly computed as:

$$E[W] = \Delta T N \left(\lambda_p - \lambda_n\right) \tag{4.8.13}$$

Note that this last expression is the same as in the independent case (Equation 5.3.4), as expected, and that unlike the SPRT, no limits in ΔT were necessary to compute the parameters for the FC and DT functions.

4.8.3 Subtractive (MIP) correlated interactions within pools (SI)

With subtractive correlations, we again derive an MGF for the spike count vector of an individual pool $\phi_{\hat{S}}(t;\lambda)$, and apply Equation 4.8.7. In this case, however, the number of spikes in a pool, conditioned on the number of spikes in that pools correlating train, is binomially distributed. Thus applying the Law of Total Probability:

$$P[S=s] = \sum_{\hat{s}=0}^{\infty} \text{Poiss}[\hat{s};\frac{\lambda}{\rho}]\text{Binom}[\hat{s}N, s;\rho] \tag{4.8.14}$$

using the definitions for the PMF's of the $\text{Poiss}[i;\lambda]$ and $\text{Binom}[N,k;p]$ distributions, we have:

$$\phi_S(t;\lambda) = E[e^{St}] = \sum_{s=0}^{\infty} P[S=s]e^{st} = e^{\left(((1+\rho(e^t-1))^N-1)\right)\frac{\lambda}{\rho}} \tag{4.8.15}$$

After applying Equation 4.8.7 with this MGF for both the preferred and null population, we find an implicit relationship for the non-trivial real root $t = h_0$ that does not depend on ΔT:

$$\left(((1+\rho(e^{t}-1))^{N}-1)\right)\lambda_p + \left(((1+\rho(e^{-t}-1))^{N}-1)\right)\lambda_n = 0 \implies h_0 = t \tag{4.8.16}$$

As before, the expected increment can be directly computed by differentiation, and we find the same expression as in the additive correlation case:

$$E[W] = \Delta T N\left(\lambda_p - \lambda_n\right) \tag{4.8.17}$$

4.9 Speed and accuracy functions with overshoot

The identities provided in Equations 5.3.1 and 5.3.2 are very useful, however are simplifications of the full formulas for FC and DT (assuming $E[W] \neq 0$) derived by Wald [123], which are:

$$FC = 1 - \frac{E[e^{h_0 E_n}|E_n \geq \theta] - 1}{E[e^{h_0 E_n}|E_n \geq \theta] - E[e^{h_0 E_n}|E_n \leq -\theta]} \tag{4.9.1}$$

$$DT = \frac{\Delta T}{E[W]}\left(E[E_n|E_n \geq \theta](FC) + E[E_n|E_n \leq -\theta](1-FC)\right) \tag{4.9.2}$$

Specifically, Equations 5.3.1 and 5.3.2 hold under the assumption that the value of the state variable on the decision step is exactly equal to the decision threshold. In practice, however, this “no-overshoot" assumption may not provide a particularly good approximation.

A correction term based on the mean of the overshoot distribution – that is, the distribution of the random variable defined by the excess distance over either the positive or negative threshold on the threshold crossing step – is suggested by Lee et al [61]. This correction is based on the Taylor expansion of the conditional expectations in Equation 4.9.1, and takes the form of a shift in the decision threshold. A correction of this form is relevant to our analysis, as the performance of two models are compared parametrically in the threshold to isolate the effects of the speed-accuracy tradeoff imparted by freely adjusting the threshold.

Denote the value of E_n conditioned on crossing the first threshold as $\hat{E}_n$, and let $X = \hat{E}_n - \theta$ overshoot random variable, with mean u_X. Expanding the conditional expectation (although dropping the conditional notation for convenience) via a Taylor series centered on this mean (the so-called delta method), we have

$$E[e^{h_0 \hat{E}_n}] = E[e^{h_0 r_0} + h_0 e^{h_0 r_0}(\hat{E}_n - r_0) + \frac{h_0^2 e^{h_0 r_0}(\hat{E}_n - r_0)^2}{2} + ...] \tag{4.9.3}$$

Choosing $r_0 = \theta$ yields an expression of Wald's truncation:

$$E[e^{h_0 \hat{E}_n}] = e^{h_0 \theta}\left(1 + h_0 E[X] + \frac{h_0^2 E[X^2]}{2} + ...\right) \tag{4.9.4}$$

$$\approx e^{h_0 \theta} \tag{4.9.5}$$

Here we see that if $E_n = \theta$, each term in the expansion becomes zero and Wald's approximation holds exactly. On the other hand, if E_n overshoots θ, error will accumulate at each term in the expansion, as a function of the moments of the overshoot distribution. If instead the expansion is performed about $r_0 = \theta + \mu_x$, a threshold-shifted approximation expresses the truncation error terms of the second and higher centered moments of the overshoot distribution:

$$E[e^{h_0 E_n}] = e^{h_0(\theta + \mu_X)}\left(1 + \frac{h_0^2 E[(X - \mu_X)^2]}{2} + ...\right) \tag{4.9.6}$$

$$\approx e^{h_0(\theta + E[X_n])} \tag{4.9.7}$$

In practice, the overshoot distribution is often nonzero; however, if its mean can be calculated and $h_0 < 0$, the truncation error associated with the latter approximation might provide a more favorable approximation as long as the higher-order moments do not grow too large. For the decision time, using this alternative approximation is exactly correct, and results in no additional error.

4.10 Joint cumulants for the SIP and MIP model

Staude et al. [116] suggest that cumulants provide a "natural and intuitive higher-order generalization of the covariance" for multineuron spiking. The two models of correlated activity examined here are indistinguishable when only examining first-order (i.e., mean firing rate) or second-order (i.e., pairwise correlations) statistics. Here, we derive the joint cumulants for each of these two models, to clarify how the spike count distributions produced by the two models differ at higher orders.

The derivation relies on the conditional independence of the spike counts for each neuron in a pool, conditioned upon the spike count in the common spike train. Let $S_1 ... S_N$ be the random variables giving spike counts in a windows of size ΔT from each of the $1...N$ neurons in a correlated pool, and $\hat{S}$ be the spike count in the common spike train. The law of total cumulance [10] allows a relatively simple expression of the joint cumulant on k members $S_1 ... S_N$ (Because of the homogeneity of the pool, we will express the k^{th} joint cumulant as calculated on $S_1 ... S_k$, but the same expression holds for any k-sized subset of $S_1 ... S_N$):

$$\kappa(S_1, ..., S_k) = \sum_{\pi \in \Pi} \kappa\left(\kappa(S_{B_1}|\hat{S}), ..., \kappa(S_{B_b}|\hat{S})\right) \tag{4.10.1}$$

Here Π is the set of all partitions of $\{1...k\}$, for example

$$\Pi[\{1,2,3\}] = \{\{\{1\},\{2\},\{3\}\},\{\{1,2\},\{3\}\},\{\{1\},\{2,3\}\},\{\{1,3\},\{2\}\},\{\{1,2,3\}\}\} \tag{4.10.2}$$

$$= \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\} \tag{4.10.3}$$

and $\kappa(S_{B_j}|\hat{S})$ is the conditional joint cumulant over the set of all spike counts indexed by an element of B_j—that is, the set $\{S_j : j \in B_j, B_j \in \pi_i\}$.

In our special case, $\kappa(S_{B_j}|\hat{S}) = 0$ whenever $|B_j| > 1$, owing to the conditional independence of each neuron given the common spike train. Moreover, from the definition of the cumulant, the term of (4.10.1) for the partition π_i that contains such a block B_j will also be zero. This implies that the only $\pi_i \in \Pi$ that contributes in Equation 4.10.1 is $\pi_i = \{\{1\}...\{k\}\}$ ($i = 1$ in the example of Equation 4.10.3);

thus

$$\kappa(S_1,...,S_k) = \kappa(E[S_1|\hat{S}],...,E[S_k|\hat{S}]) = \kappa_k(E[S_1|\hat{S}]), \tag{4.10.4}$$

where we have used the fact that the first cumulant is simply the expected value. Using the cumulant generating function, we then have a formula for the joint cumulant:

$$\kappa(S_1,...,S_k) = \frac{d^k}{(dt)^k}\left[\log E[e^{tE[S_1|\hat{S}]}]\right]\Big|_{t=0} \tag{4.10.5}$$

Thus, for the two models of correlations (assuming a firing rate λ), we have:

$$MIP: \quad E[S_1|\hat{S}=\hat{s}] = \sum_{s_1=0}^{\hat{s}} s_1 \binom{\hat{s}}{s_1} \rho^{s_1}(1-\rho)^{\hat{s}-s_1} = \rho\hat{s} \tag{4.10.6}$$

$$\implies \log E[e^{tE[S_1|\hat{S}=\hat{s}]}] = \log \sum_{\hat{s}=0}^{\infty} \frac{e^{-\Delta T\lambda/\rho}(\Delta T\lambda/\rho)^{\hat{s}}}{\hat{s}!} e^{t\rho\hat{s}} \tag{4.10.7}$$

$$= \frac{\lambda\Delta T(e^{\rho t}-1)}{\rho} \tag{4.10.8}$$

$$\implies \kappa(S_1...S_k) = \frac{d^k}{(dt)^k}\left[\frac{\lambda\Delta T(e^{\rho t}-1)}{\rho}\right]\Big|_{t=0} \tag{4.10.9}$$

$$= \boxed{\Delta T\lambda\rho^{k-1}} \tag{4.10.10}$$

$$SIP: \quad E[S_1|\hat{S}=\hat{s}] = \sum_{s_1=\hat{s}}^{\infty} s_1 \frac{e^{-(1-\rho)\Delta T\lambda}((1-\rho)\Delta T\lambda)^{s_1-\hat{s}}}{(s_1-\hat{s})!} = \hat{s} + \lambda\Delta T(1-\rho) \tag{4.10.11}$$

$$\implies \log E[e^{tE[S_1|\hat{S}=\hat{s}]}] = \log \sum_{\hat{s}=0}^{\infty} \frac{e^{-\Delta T\lambda\rho}(\Delta T\lambda\rho)^{\hat{s}}}{\hat{s}!} e^{t(\hat{s}+\lambda\Delta T(1-\rho))} \tag{4.10.12}$$

$$= \lambda\Delta T(\rho[e^t - t - 1] + t) \tag{4.10.13}$$

$$\implies \kappa(S_1...S_k) = \frac{d^k}{(dt)^k}\left[\lambda\Delta T(\rho[e^t - t - 1] + t)\right]\Big|_{t=0} \tag{4.10.14}$$

$$= \boxed{\begin{cases} \lambda\Delta T & : k = 1 \\ \lambda\Delta T\rho & : k > 1 \end{cases}} \tag{4.10.15}$$

Comparing Equations 4.10.10 and 4.10.15 (see also Figure 4.9), we see agreement for $k \leq 2$ as expected; these correspond the the intended firing rate and pairwise

covariance of neurons within the pool. However, for $k > 2$, we see the signature of the differences in the structure of the correlations. For the MIP model, the joint cumulant decays geometrically as more and more neurons are considered. In contrast, the joint cumulant remains constant for the SIP model.

CHAPTER 5

Role of Noise in a Spiking Integrator

5.1 Introduction

The noisy, collective activity of neurons in multiple cortical and subcortical brain regions underlie evidenced-based decisions. In the case of perceptual decisions, accumulation of this evidence over long timescales might be accomplished via neural components with relatively short memory properties. For example, in the dot motion discrimination task, evidence for the direction of a field of random dots moving with weak coherence in one direction is accumulated over several seconds [102, 13, 12, 79]. In contrast, the neurons that might be responsible for this computation exhibit firing variability at timescales on the order of tens of milliseconds (see, for example [114]). This suggests some type of circuit-based mechanism might underlie evidence accumulation.

Several models for the mechanism of evidence accumulation have been suggested. At one extreme, early models based on a random walk with a single state variable [59, 63, 117] have proven successful at modeling the psychophysical performance of subjects deciding between two alternatives. At the other end of the spectrum, a model by Wang [126] (hereafter called the spiking attractor model), governed by 9200 differential equations, can produce similar psychophysical performance, with detail down to the membrane voltage of individual neurons in the accumulation circuit. In between, reduced models that hypothesize leaky accumulation circuits [121], feedforward short-term memory [44], or reduced models for

population dynamics [129, 30, 95] all attempt to provide descriptive and predictive models of evidence accumulation over time.

How do assumptions about the entrance of signal and noise affect dynamical models of evidence accumulation? In models derived as reductions of other, more biophysically detailed, models, noise terms are integral to the model formulation, and can have significant impact on the decision making performance of the circuit. In this study we expand on the work of [129, 30, 31, 95], in examining the high-dimensional model described in [126]. We attempt to establish how a combination of internal and external noise, as well as nonlinear dynamics, affects decision making performance. To accomplish this, we begin by establishing an upper bound on the performance that the original 9200 dimensional circuit can produce, given the noise in the input signal it receives. We then compare this bound to the actual performance of the spiking attractor model, in order to assess its efficiency.

The upper bound on performance is computed via pure integration, and lacks any noise beyond the Poisson variability introduced by the population rate code. We conjecture that the performance difference between this upper bound and the true model performance is a consequence of two aspects of spiking attractor: internally generated noise sources, and nonlinear dynamics. Previous studies [129, 30, 95] have observed that the accumulation of evidence in the spiking attractor model is nonlinear, and can be well-described via reduced nonlinear models. Moreover, additional noise sources – in addition to those in the evidence signal – arise in the spiking network model. By fitting a reduced dimensional model to the dynamics, we hope to separate the effect of noise and nonlinear dynamics on performance.

5.2 Model Overview

Figure 5.1 provides a schematic overview of the spiking attractor model. According to the original model formulation [126], the input signal for the spiking attractor model is provided by a firing rate code, modeled as independent Poisson point processes (model spike trains) that drive two subpopulations of $N = 240$ model leaky integrate-and-fire neurons. During a trial, each of these subpopulations (S1 and S2) receives input from 240 independent spike trains (with one-to-one con-

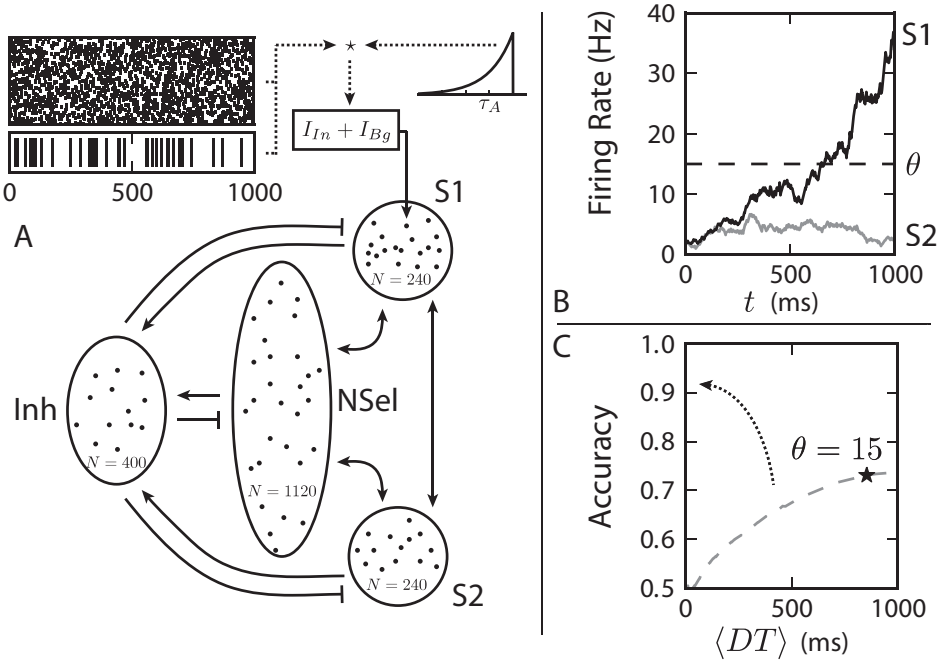


Figure 5.1: Background and input drive combine to influence each cell in the selective populations of the spiking attractor model; additional feedback then diminishes performance. (A) One neuron providing $\sim$ 40 Hz drive combines with 100 independent background neurons at 24 Hz at the level of current driving a model neuron in subpopulation S1. Each excitatory (S1, S2, NSel) and the inhibitory (Inh) subpopulation receives similar background drive, but only selective neurons receive input at a firing rate that depends on the task difficulty (Equations 5.2.1 for S1, and 5.2.2 for S2). (B) The resulting activity of the two selective populations determines the alternative that is selected, by a race-to-bound diffusion. (C) At coherence $C = 6.4$, the decision threshold θ determines the particular tradeoff between speed and accuracy; as this value is changes, a curve is plotted parametrically. A "*" shows the location of $\theta = 15$, from (B) on this parametric curve. A "rotation" of this curve towards the upper-left, indicated by the arrow, would correspond to improved performance.

nectivity), after a transitory period of activity that allows the network to reach a steady state (we use 2 seconds of simulation time). The firing rate λ of these model spike trains is dependent upon a parameter C (In the original study [126], C corresponds to the coherence of a field of randomly moving dots that are presented to a subject):

$$\lambda_1 = 40 + .4C \tag{5.2.1}$$

$$\lambda_2 = 40 - .4C \tag{5.2.2}$$

The problem of decision making is then framed as determining the pool with the higher firing rate.

In addition to the coherence-dependent input, each of the 1120 nonselective excitatory (NSel) and 400 inhibitory (Inh) cells receive 2400 Hz of excitatory independent Poisson background drive. This additional drive is also pictured in Figure 5.1 for a single cell in S1, next to an example evidence encoding input train, to emphasize the magnitude of this additional variability.

Over the course of a trial the firing rate of these subpopulations initially increases as a consequence of the additional external input. This dynamical phenomenon has been shown to result from an instability that is formed by the additional input spike trains [129, 30, 31]. Eventually, the firing rate of one selective population grows, while the alternative population recedes back to a low value; this is depicted in Figure 5.1B. The first population to reach a pre-specified firing rate determines the “decision" of the network.

As this pre-specified threshold, referred to as the decision threshold, is increased, trials take longer to terminate. On average, however, this results in better performance, as earlier variability is less likely to cause a spurious threshold crossing. The set of all accuracy and mean reaction-times attainable by a threshold adjustment defines a speed-accuracy tradeoff curve, (pictured in Figure 5.1C at $C = 6.4$), that provides the fundamental model comparison tool for this study. We note that an outward rotation of this curve is represents improved performance, i.e. for a given mean decision time, a higher accuracy.

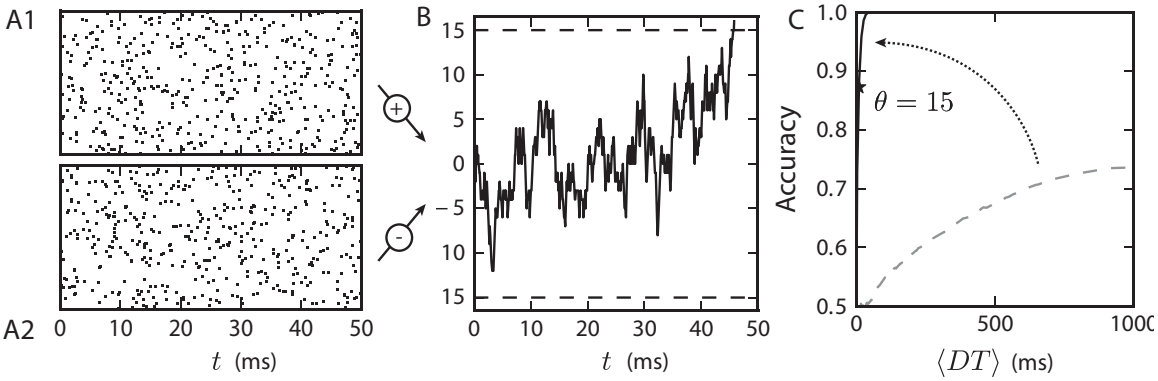


Figure 5.2: An integrate-to-bound model based on spike integration results in a speed-accuracy tradeoff curve. (A1,A2) Two rastergrams generated as independent Poisson point processes for $C = 6.4$ mark the spike for $N = 240$ neurons per pool. (B) A spike-integration model results from the spikes in these two pools being integrated to a decision bound at $\theta = 15$. (C) The speed accuracy tradeoff curve defined by Equations 5.3.1 and 5.3.2, is plotted as a solid line in the very top/left, with a "*" indicating $\theta = 15$, as in (B). Integrating spikes alone significantly outperforms the spiking attractor model, indicated with a broken line.

5.3 An upper bound on performance

When provided with independent spiking input, one decision method is summation (or integration) of the spikes from the pools, with opposite signs. This amounts to a random walk, and a decision is reached when the integration reaches one of two predefined bounds (See Figure 5.2, and also Chapter 4). As in the firing full spiking model the value of this boundary parameterizes the speed-accuracy tradeoff.

Zhang and Bogacz [133] derived the speed and accuracy functions resulting from integration from two populations of independent Poisson spike trains:

$$\text{Accuracy} = \frac{1}{1+e^{h_0\theta}} \tag{5.3.1}$$

$$DT = \frac{\theta}{\mu}\tanh\left(\frac{-h_0\theta}{2}\right) \tag{5.3.2}$$

where

$$h_0 = -\log\left(\frac{\lambda_1}{\lambda_2}\right) \tag{5.3.3}$$

$$\mu = N(\lambda_1 - \lambda_2). \tag{5.3.4}$$

For some drift-diffusion models, overshoot of the accumulator over the decision boundary on the final step can limit the extent to which these formulas can be applied [40]. However this does not pose a problem for this model as long as the boundary takes on integer values. Importantly, [133] also demonstrates that, up to a change in threshold, exact spike integration implements a sequential probability ratio test, which takes (on average) the minimum DT for a required accuracy [125] of any sequential test. In this sense the performance curve plotted in Figure 5.2C provides an upper bound for the mean decision time of the spiking attractor model, for a given level of accuracy. Clearly the performance of an integrator acting on the inputs alone performs much better than the spiking attractor model.

5.4 Additional variability via background inputs

Is it fair to compare the performance of a spike integrator that only receives evidence via the coherence-dependent input spikes again the spiking attractor model? In the complete spiking attractor model, spikes are not simply accumulated over time. Instead, each spike train independently contributes input to a model leaky integrate-and-fire (LIF) neuron within one of the selective subpopulations via:

$$C_m\frac{dV}{dt} = -g_L(V - V_L) - I_{In} - I_{Bg} - I_{Rec} \tag{5.4.1}$$

$$I_{In} = g_A(V - V_E)S_{In} \tag{5.4.2}$$

$$\frac{dS_{In}}{dt} = -\frac{S_{In}}{\tau_A} + X(t) \tag{5.4.3}$$

$$X(t) = \sum_k \delta(t - t_k) \tag{5.4.4}$$

The dynamics of the voltage for each of these neurons responds to input spikes from $X(t)$, at times t_k. As these equations indicate, the effect of an input spike at time t_k is not to increment a spike integrator, but rather to increase the conductance of a neuron in the network via Equation 5.4.2.

Two other terms contribute current to a neuron in the pool in Equation 5.4.1: I_{Bg} and I_{Rec}. On the one hand, I_{Rec} serves as the current source contributed by the internal dynamics of the circuit. On the other hand, I_{Bg} is a consequence of background drive that keeps the neuron in an excitable state, and responsive to other inputs. We regard this term as an external noise term, which (through direct summation with the input current) serves to add variability to the signal.

In the spiking attractor model, the contribution of the background current in terms of signal-to-noise is very large, as it results from summation of EPSC's from a 2400 Hz Poisson input source, relative to the $\sim$ 40 Hz drive from the evidence-encoding current. Because of this additional source of noise, it is unrealistic to expect the spiking attractor model to perform as well as a spike integrator—it has much more noise to contend with! Figure 5.1A describes the sources of external drive received by one neuron in the model, from both the input and background sources.

To what degree does this extra variability from background excitation worsen the decision making performance of the spiking attractor model? The current supplied by these two sources is summed at each individual neuron. In order to approximate the drive to the entire population, we summed the contributions from each neuron across the entire selective subpopulation:

$$I_{Ext} = \sum_{i=1}^{240} I_{In}^{i} + I_{Bg}^{i} \tag{5.4.5}$$

$$\approx \sum_{i=1}^{240} g_A(\bar{V} - V_E)\left(S_{In}^{i} + S_{Bg}^{i}\right), \tag{5.4.6}$$

where $\bar{V}$ is an average voltage taken across the population. The mean of this stochastic process varies linearly with the firing rate of the input neurons, yet is highly variable due to the 240×2400 Hz drive from the background neurons. When the two stochastic processes (one for each selective pool) are integrated to

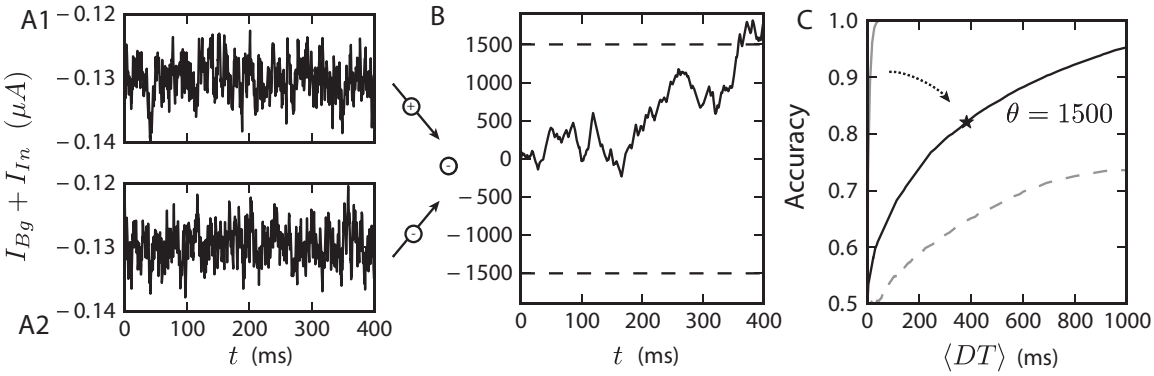


Figure 5.3: Including background, as well as input, current significantly diminishes performance. (A1,A2) Two traces of input current form the inputs to a drift diffusion model. (B) When these currents are differenced and integrated, a threshold scheme analogous to spike integration determines which input possessed a higher mean activity (Note that a multiplication by -1 has been added, to recast the upper bound as "correct"). (C) This scheme produces much worse performance, owing to the addition of background current; this is depicted with the solid black line (the dotted arrow indicates the magnitude of the performance loss from the speed-accuracy curve from Figure 5.2C, and a star indicates mean reaction time and accuracy at the level of decision bound depicted in B). The broken line indicates the performance curve of the full spiking attractor model, reproduced from Figure 5.1C.

bound, we have another drift-diffusion model.

For this model, decision making performance will likely be better than that of the entire network, by excluding the nonlinearity in the dynamics and additional variability contributed by the other network components. An example of this model dynamics at $C = 6.4$ is depicted in Figure 5.3:A1,A2,B. How does the performance of this model compare to the performance of a spike integration model, without the background input? Figure 5.3C plots the performance of this model, which we refer to as the current-integration model. Also replotted from Figure 5.1C is the performance of the spiking attractor model. Accounting for background input can be seen as a rotation of the speed-accuracy tradeoff curve from high values indicating optimal performance to lower ones the indicating significantly diminished performance (black line). However, the loss of performance of the spiking accumulator model cannot be fully explained from the additional variability contributed by background input alone.

5.5 Two factors that diminish spiking accumulator performance

In the spiking attractor model, the alternative associated with the first selective subpopulation to reach a pre-specified activity level is selected as the models choice on a given trial. This selection rule is similar to that in simple race-to-bound accumulation models [122, 120]. Importantly, several studies [133, 9] have shown that such models offer reduced performance relative to models that accumulate the *difference* of two quantities until it reaches a bound. The current-integration model introduced in Section 5.4, in contrast, makes a selection in exactly this way—by waiting for the difference between the value of two integrators to reach a threshold. Would the performance of the spiking integrator be significantly improved by selecting an alternative in the analogous way – when the difference of the firing rates of the selective subpopulations reaches a threshold?

A second distinction between the spiking attractor model and the current-integration model comes from the initial condition of the accumulation. In the former, a pre-specified amount of time (2 seconds in this study) is taken before starting a trial, to allow the network to reach a state of statistical equilibrium. This produces variability in the initial state of the circuit, a factor not present in the current-integration model (but which is a key component in other decision making models [89, 63]). Might the performance of the spiking integrator be significantly improved eliminating this source of “initial" noise?

Figure 5.4B answers both of the questions we have posed by illustrating the improvement in the performance of the spiking integrator that results from resolving both differences we have highlighted. By selecting an alternative based on the difference of two firing rates (Figure 5.4A), rather than a race to bound (Figure 5.1B), a modest performance improvement can be achieved Additionally, only trials where the initial (after the 2 seconds to settle into equilibrium) firing rate of the selective populations differed by less then .5 Hz were included, in an attempt to enforce a noise-free initial condition. As a result, the performance of the spiking attractor model is improved incrementally toward that produced by current integration, though the overall effect is small.

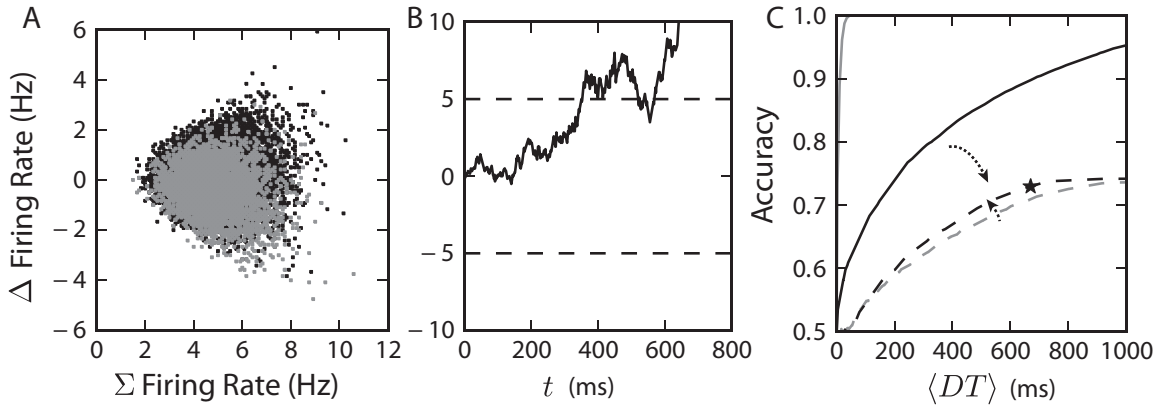


Figure 5.4: Reducing variability in the initial condition, and using a difference-to-bound technique, can improve the performance of the spiking accumulator model. (A) After two seconds of simulation time, the initial condition for the spiking attractor model can become quite variable, biasing which alternative is selected and reducing performance. (Gray dots indicate initial conditions that resulted in errors, and black dots successes.) (B) The same two firing rate traces from Figure 5.1B are subtracted. The threshold crossing of this difference determines which alternative is selected. (C) By only including trials that have an initial Δ Firing Rate less than .5 (in absolute value), and incorporating a difference-to-bound decision making rule,the performance of the spiking accumulator can be improved (broken gray line to broken black line).

5.6 Nonlinear accumulation does not limit performance

In the previous sections, we have attempted to explain the performance of the spiking attractor model by incrementally examining the consequences of various features of its dynamics and sources of variability. First, we established a theoretical upper bound on model performance, based on the limits imposed by the coding properties of a pool of independent Poisson neurons. We then recognized that, from the model's perspective, a large amount of additional variability is contributed by background excitation external to the circuit dynamics. This noise source can not be separated from the signal by any circuit dynamics, due to summation at the synapse. Therefore, a more reasonable bound on decision making followed from accumulating the summed current contributed by these two sources.

These observations resulted in a predicted performance curve depicted in Figure 5.4B as an unbroken dark line. Two features of this current-integration model differ from the mechanisms at work in the spiking attractor model. In the integration model, an noise-free, unbiased initial conditions and a difference-to-bound selection rule help to implement a drift diffusion model with nearly optimal performance. By incorporating these two features into the spiking attractor model, performance was improved. However, there remained a performance gap between the current-integration model and the spiking attractor model.

The current-integration model assumes that integration of the evidence-encoding input occurs linearly. However, several studies [129, 30, 31, 95] have found that the spiking attractor model produces firing rate nonlinear dynamics. Moreover, the additional external variability caused by background excitation, that contributes noise to the input signal received by the selective subpopulations, also drives the rest of the decision making circuit. We next attempt to disentangle these two effects, so that the performance of the spiking attractor model can be fully characterized.

To study the separate the effects of the background excitation and the nonlinear dynamics, we examined a simple one-dimensional dynamical model. The goal was to reproduce the attractor dynamics of the difference between the mean

firing rates of the selective pools. Several studies have shown how the spiking network dynamics can be reduced to a two dimensional system [129, 30, 31], in which the nonlinear firing rate dynamics are modeled by passing the inputs through a nonlinearity. The one-dimensional process we consider closely follows the work of [95], and is informed by an energy landscape characterization of the nonlinear dynamics. Two attracting fixed points serve as competing sinks for the global network activity:

$$\frac{dE}{dt} = -\beta E(\frac{E}{a}+1)(\frac{E}{a}-1)(E+id)(E-id) + \gamma \Delta S \tag{5.6.1}$$

here $\Delta S = S_{Bg1} + S_{In1} - S_{Bg2} - S_{In2}$, i is the imaginary unit, and E models the time-evolution of the difference in firing rate between the selective pools, hereafter called the firing rate difference. The constants β and γ describe the relative influence of the nonlinear dynamics and input ΔS respectively. When $\beta = 0$, this nonlinear diffusion equation reduces to a pure integration model, otherwise a corresponds to the location of the two attracting fixed points. The remaining roots of this 5^{th} order polynomial are constrained to be purely imaginary, to preserve the symmetry of the function. This form was informed by the fact that the network dynamics undergoes a subcritical pitchfork bifurcation, with a stability change near $E = 0$, upon the onset of the input stimulus. Here a five-fixed-point system changes into a three-fixed-point system, as demonstrated in [129].

We fixed the parameter $a = 34.63$ by observing the spiking integrator simulations over long trials, to identify the location of the attracting fixed points of the system. Next, we fit the parameters β, γ, and d by minimizing the difference between the firing rate traces generated from the model, and those that would be predicted by Equation 5.6.1 when integrating the same input streams. The error function we minimized took the form:

$$\sum_{i=1}^{k} \sqrt{\int_0^{\frac{T}{2}} \left(\hat{E}_i(t) - E_i(t)\right)^2 \mathrm{d}t} \tag{5.6.2}$$

here $\hat{E}$ is the attractor network's firing rate difference, k is the number of trials used for fitting, and T is the total trial length. This quantity corresponds to the total

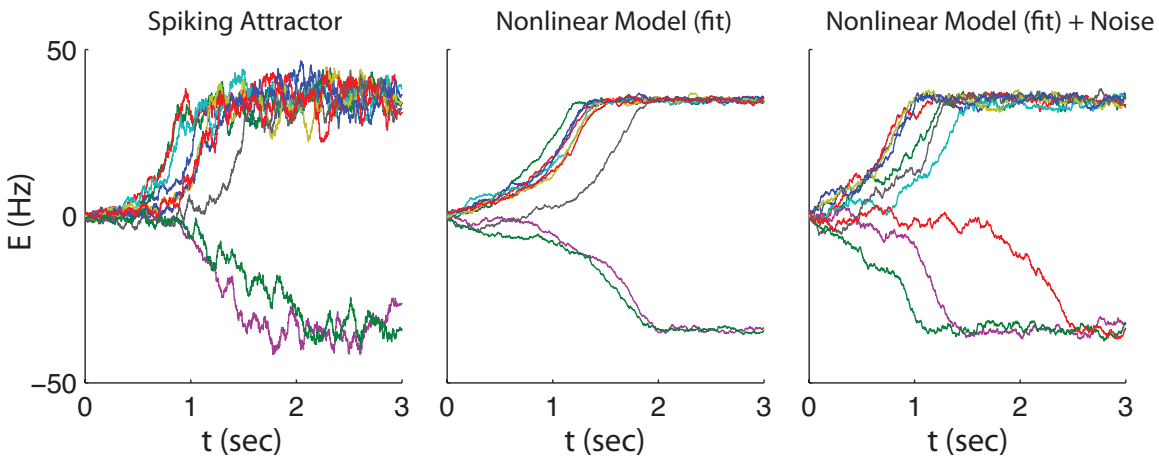


Figure 5.5: Firing rate traces generated from the same input data, for the spiking attractor model, and the nonlinear integrator, and the nonlinear integrator model with additive noise. Our goal is a qualitative, and as much as possible a quantitative, fit between our one-dimensional model and the stochastic dynamics of the spiking attractor model.

root squared error over the first half of a trial, which describes the period during which the network is accumulating information, until a stable point is reached. Thus, minimizing this term corresponds to matching the one-dimensional model dynamics to the spiking attractor dynamics. To minimize the error, we used the MATLAB implementation of the trust-region-reflective algorithm [24, 25].

Fitting resulted in the rate-of-change (dE/dt) function plotted in Figure 5.6A. Figure 5.5A and B compare several example traces of $E(t)$, for both the spiking attractor and the fit-nonlinear model. Qualitatively, both the form of the nonlinearity and the firing rate plots demonstrate the expected supra-linear expansion of trajectories away from the origin. Surprisingly, however, in our simulations this initial exponential growth does not appear to severely impact performance. This is illustrated in Figure 5.6B where the red-colored line shows that the nonlinearity of the dynamics results in very little performance loss.

This reduced, one dimensional model seems to suggest that nonlinear dynamics near the dynamical instability of the network are not enough to account for the performance of the spiking network. As suggested earlier, perhaps additional noise sources are to blame for this discrepancy. The variability introduced in the background excitatory drive to the selective pools directly contributes to the increased noise in the input signal, however the remaining background drive to

the nonselective excitatory neurons and inhibitory interneurons has yet to be addressed.

We quantify this additional variability, and introduce it as an additional additive noise source on the input signal. Following Smith [109], we first derive the variance of the signal into the selective pools, including input and background for only the selective subpopulations. The input signal takes the form of a shot-noise process. The steady-state mean and variance of the S (defined as the sum of output from a Poisson spike train where each spike has been convolved with an exponential modeling the resulting EPSC), can be computed as a function of the number of neurons in a pool, N, the timesacel of synaptic decay τ, and the firing each neuron in the pool λ:

$$E[S] = \lambda\tau N \tag{5.6.3}$$

$$Var[S] = \frac{\lambda\tau N}{2} \tag{5.6.4}$$

$$Var[\Delta S] = Var[S_{Bg1}] + Var[S_{In1}] + Var[S_{Bg2}] + Var[S_{In2}] \tag{5.6.5}$$

$$= 1171.2Hz^2 \tag{5.6.6}$$

where we have made a units conversion from seconds to ms. We add an additional noise signal that attempts to capture the variance contributed by the background external drive of the other two subpopulations, NSel and Inh. Here each neuron (a total of 1520) receives 2400 Hz background input, providing an additional noise source S_{Noise}

$$\Delta\hat{S} = \Delta S + S_{Noise}, \tag{5.6.7}$$

and resulting in a variance contribution similarly calculated from the variance of a shot-noise process:

$$\sigma^2_{noise} = Var[S_{NSel}] + Var[S_{Inh}] = 3648Hz^2 \tag{5.6.8}$$

In Figure 5.6(B) we took this noise source to be a zero mean Ornstein-Uhlenbeck

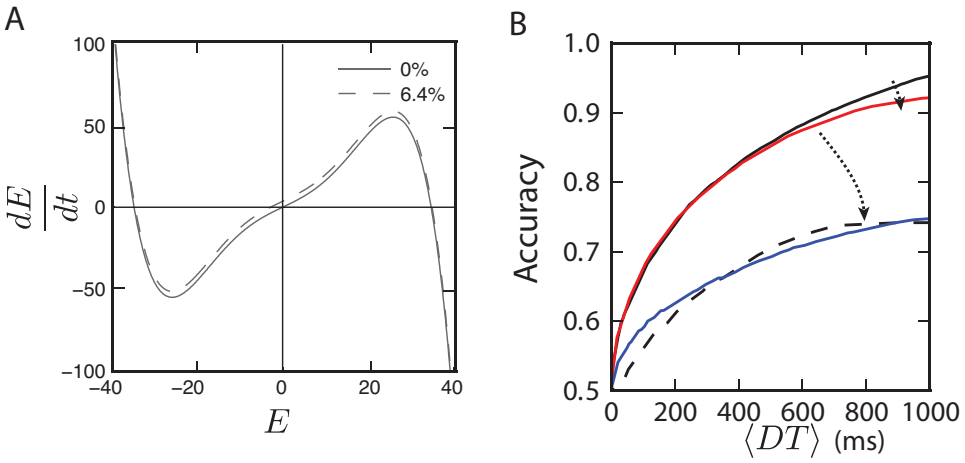


Figure 5.6: A nonlinear model of evidence integration appears to perform nearly optimal evidence accumulation. (A) The parameters $\beta = .0053$, $\gamma = 1.485$ and $d = 15.3480$ in Equation 5.6.1 were fit to the firing rate traces generated by the spiking network model. The resulting nonlinearity exhibits expansion about the origin. Here the vertical shift from $C = 0\%$ to $C = 6.4\%$ is computed as $\gamma E[\Delta S]$ [109]. (B) However, this nonlinear integrator, when applied to the same input integrated by the spiking accumulator, results in surprisingly little performance loss (red line). When additional noise with equal variance and timescale to the background drive into the non-selective and inhibitory neurons is added to the input signal, performance falls to a level comparable to the spiking attractor model (blue line).

process (following [129, 30]) with timescale and variance equal to that of the additional background drive.

$$dS_{Noise} = -\frac{S_{Noise}}{\tau_A}dt + \sqrt{\frac{2\sigma^2_{noise}}{\tau_A}}dW_t \tag{5.6.9}$$

Adding this extra noise results in significantly impacted performance, as shown in Figure 5.6(B) by the blue line. In fact, the performance is now at a level very comparable to the spiking attractor model, suggesting that the external noise is largely responsible for the reduced performance of the spiking accumulator model.

5.7 Summary

This chapter demonstrates that, for one value of dot-coherence (C=6.4), the performance of the spiking attractor model significantly underperforms that which could be obtained via optimal inference on its sensory inputs. Multiple aspects of the model could account for this fact, however our study strongly implicates the additional background variability exciting the network, on the scale of 2400 Hz independent Poisson input per neuron in the network. Far from being purely a “bug" of the model, this additional variability facilitates the rich attractor dynamics that make this model unique, and have been the focus of several detailed studies.

The performance of the spiking attractor model can be placed on a more level playing field, by comparison to the performance of models that receive similar amounts of variability. First, the strength of the input signal can be diluted by the background drive to the selective populations, that is summed at the level of the synapse. Second, by enforcing a noise-free, unbiased initial condition for the circuit, and an improved selection rule, the performance of the model itself can be enhanced. This leaves two potential factors that might contribute to the remaining performance discrepancy: additional circuit variability, ostensibly a consequence of the background drive to the nonselective excitatory and inhibitory neurons in the network, and nonlinear dynamics that prevent exact evidence integration.

To address the latter, we fit a one-dimensional reduced model to the dynamics of the difference in firing rates of the selective populations. Although performance at long mean decision times was reduced, the overall impact of the nonlinearity of best fit was surprisingly small. In contrast, when the noise obfuscating the input signal was roughly tripled, as would be predicted by the variability contributed by the background drive to the remaining neurons in the network, performance fell to a level qualitatively similar to the performance of the original spiking attractor model. This points to the conclusion that the background drive that enables the nonlinear dynamics of the system, and not the nonlinear dynamics themselves, is responsible for setting the performance of the spiking attractor model. Clearly, however, further, more rigorous study will be necessary to quantify the contributions to performance loss of nonlinearity and noise.

CHAPTER 6

Extensions: Decision making in the Retina

6.1 Introduction and Background

This chapter presents a theoretical study of a model of flash detection by the retina, that relies on the principles of decision making, and a novel estimator for the accuracy and decision time performance of the standard drift-diffusion model.

The first study is posed at the earliest stage of processing in the visual system, the retina. The three types of retinal interneurons situated between the input and output layers, the horizontal, bipolar, and amacrine cells, perform parallel computations (see [33] for a review) that are then communicated downstream. Clearly a component of this computation must include aggregation from the point-sources of the photoreceptors to the receptive fields of the ganglion cells. What is the nature of the underlying computation, and how can the principles of optimal decision making theory relate to this aggregation mechanism?

The retina is a complex tissue that converts light stimulus into neural activity, providing the first stage of neural visual processing. As a whole, despite considerable feedback, the visual system can be thought of as hierarchical, with staged computations being performed at the various levels of the hierarchy, upon input that describes an increasingly broad receptive field. At the photoreceptor cells, the first stage of processing is clearly defined as detection of photons in a small area

of visual space [115]. Retinal ganglion cells, the last layer of neurons in the retina, possess a center-surround receptive field first characterized by Kuffler [57]. These cells provide the input to the subsequent layers of the visual system, principally via the lateral geniculate nucleus and the superior colliculus.

We study the computation performed in scotopic vision, where light levels are low enough to shift visual processing to a specialized cone pathway that relies on rod → rod-bipolar → amacrime → cone-bipolar → ganglion cell transmission [34]. At this luminance level, the quantal nature of light [35] becomes relevant to the computation performed by the circuit. At the first stage of processing, rod-bipolar cells receive input from tens to hundreds of rod cells—a first example of the aggregation computation performed by the circuit. However, when only few photons reach one of these rod cells at a time, how does a rod-bipolar cell detect their presence amid the noisy input of the inactive rods that did not receive a photon?

We frame this question a two alternative decision making task for the cells in this circuit. On a given trial, a low-luminance flash activates a subset of the rod cells that possess feedforward connections to a rod-bipolar cell. The activity of the rod-bipolar must aggregate the signals from the rods, and detect the flash. What qualities of synaptic transmission from the rod pool to the bipolar cell facilitate optimal detection of flash events by the rod-bipolar cells?

Signal transmission across the dark-adapted rod→rod-bipolar cell synapse has been investigated theoretically [91], and characterized experimentally, as nonlinear [34, 113]. While the average response of rod responses (measured in pA) was found to vary linearly with the magnitude of a low intensity flash (measured in average number of rhodopsin isomerization events), rod-bipolar responses varied supralinearly [34]. This nonlinear signal transfer ignored many of the single-photon responses from upstream rods, but also removed much of the noise from rods that did not detect a photon. This tradeoff of ignoring portions of a signal to gain a performance enhancement via a thresholding operation is similar the robustness threshold discussed in Chapter 3. (See also Figure 1.2B.)

The implication is that the aggregation by the rod-bipolar cell is not linear summation, but rather is an operation that rejects some photon absorptions in order to

eliminate noise. Here we examine a model of this system and find that such a nonlinearity is a necessary attribute of an optimal flash detection circuit. Rather than constructing a mechanism that intentionally reduces noise, we instead construct a probabilistic model of the retinal circuitry that performs a statistically optimal computation. An outcome of this circuit, then, is a nonlinearity that agrees both with the intuition introduced above, as well as experimental observation.

6.2 Flash detection model

We formulate a statistical model of flash detection in Figure 6.1, that depicts the conditional dependence relationships that arise between random events that can occur in the circuit. On a given trial, the random variable F represents either a flash $F = 1$ or no-flash $F = 0$ stimulus presented to the aggregation circuit, with probability q. The outcome (decision) of the trial is represented by the random variable D; an exact, deterministic relationship between D and F would constitute a perfect flash detector.

However, the neural machinery of the retina make such a deterministic relationship impossible. In our model, this is due to three main factors. First, the weak flash implies that only a subset of the N intermediate photodetectors will even receive a photon. Even if a photon does arrive at a photodetector (here we assume that only one of two cases, one photon or none, are possible due to the low light level), that photon may not trigger a rhodopsin isomerization. These two possibilities are modeled by the random variable (RV) X_i, so that a flash triggers a single isomerization with probability p (a more elaborate model that separates these two effects would require separation of these cases into two separate, although not independent, probabilistic events).

Second, we account for the possibility of a false isomerization event with the RV Y_i, where a rhodopsin might trigger a photon-like event despite the absence of a flash. The outcomes of X_i and Y_i determine the probabilistic response in the i^{th} photodetector. Third, this response is modeled as a Gaussian RV, with a fixed standard deviation σ_a, and variable mean μ_a dependent upon the outcomes of X_i and Y_i. Downstream inference made by the aggregation circuit thus only has access

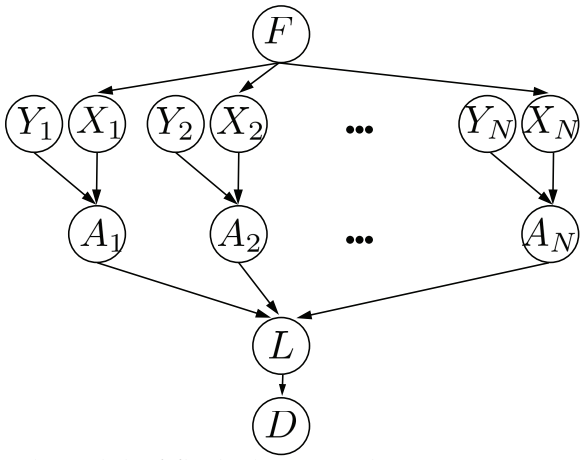


Figure 6.1: Statistical model of flash detection by response aggregation across array of noisy, unreliable photodetectors. Here F is a Bernoulli random variable (RV) representing the event of a flash occurring on the current trial (or failing to). On each trial, D (another Bernoulli RV) represents the decision made about F by the circuit. The intermediate RVs X_i, Y_i, A_i, and L model the attributes of noisy circuit components.

to the responses A_i, and thus an imperfect window into the presence or absence of a flash.

All three of these sources of variability factor into the probabilistic relationship between F and A_i. For convenience, we define the following notation for the cumulative density function (CDF) and probability density function (PDF) of a normally distributed RV $X \sim N\left(\mu, \sigma^2\right)$:

$$F_X(x; \mu, \sigma^2) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f_X(x; \mu, \sigma^2) = \frac{d}{dx} F_X(x; \mu, \sigma^2)$$

The following tables present the conditional probability distributions for the RVs in the model:

	$f = 1$	$f = 0$
$P[F = f]$	q	$1 - q$

	$y_i = 1$	$y_i = 0$
$P[Y_i = y_i]$	e	$1 - e$

	$x_i = 1$	$x_i = 0$
$P[X_i = x_i \mid F = 1]$	p	$1 - p$
$P[X_i = x_i \mid F = 0]$	0	1

	CDF
$P[A_i \leq a_i \mid X_i = 1, Y_i = 1]$	$F_{A_i}(a_i; \mu_p, \sigma_a^2)$
$P[A_i \leq a_i \mid X_i = 1, Y_i = 0]$	$F_{A_i}(a_i; \mu_p, \sigma_a^2)$
$P[A_i \leq a_i \mid X_i = 0, Y_i = 1]$	$F_{A_i}(a_i; \mu_p, \sigma_a^2)$
$P[A_i \leq a_i \mid X_i = 0, Y_i = 0]$	$F_{A_i}(a_i; 0, \sigma_0^2)$

In this section we construct the probability distribution across the responses of the photoreceptors, respecting the assumptions. We factor the joint probability distribution of the entire model according to conditional independence relationships, and find:

$$P[A_1 ... A_N, X_1 ... X_N, Y_1 ... Y_N, F] = P[F] \prod_{i=1}^{N} P[A_i, X_i, Y_i \mid F] \tag{6.2.1}$$

$$P[A_i, X_i, Y_i \mid F] = P[A_i \mid X_i, Y_i] P[Y_i] P[X_i \mid F]$$

These independence relationships are shown in the graphical model pictured in Figure 6.1, where any two random variables are conditionally independent when conditioned on the random variables that are “parents" in the directed, acyclic graph. Next, we marginalize this distribution across each X_i, Y_i to find the factored distribution of the A_i's.

$$P[A_1 ... A_N, F] = P[F] \prod_{i=1}^{N} P[A_i \mid F] \tag{6.2.2}$$

$$P[A_i \mid F] = \sum_{Y_i} P[Y_i] \sum_{X} P[A_i \mid X_i, Y_i] P[X_i \mid F] \tag{6.2.3}$$

Finally, we marginalize the distribution across F to arrive at the full probability distribution of the responses:

$$P[A_1 ... A_N] = q \prod_{i=1}^{N} P[A_i \mid F = 1] + (1 - q) \prod_{i=1}^{N} P[A_i \mid F = 0], \tag{6.2.4}$$

where the conditional distribution of a response given a flash is given by

	$P[A_i \mid F = f]$
$f = 1$	$(1-(1-e)(1-p))\, f_{A_i}(a_i; \mu_p, \sigma_a^2) + (1-e)(1-p) f_{A_i}(a_i; 0, \sigma_0^2)$
$f = 0$	$e f_{A_i}(a_i; \mu_p, \sigma_a^2) + (1-e) f_{A_i}(a_i; 0, \sigma_0^2)$

This conditional probability distribution describes the probabilistic responses of an array of photodetectors, that can be used for inference about the flash.

6.3 Inference based on the photodetector responses

Suppose we have two hypothesis about the the cause of the data $a_1 ... a_n$ that were observed on a trial, the null H_0 and alternative H_1. The null hypothesis states that there was no possibility of a flash on the given trial, and the alternative states that there was certainly a flash on the given trial.

$$H_0 : q = 0 \tag{6.3.1}$$

$$H_1 : q = 1$$

How does D, the decision variable, depend on these responses? We use a deterministic function L, the likelihood ratio for these two alternatives:

$$L(q; a_i ... a_N) = P[A_1 = a_1 ... A_N = a_i; q] \tag{6.3.2}$$

The likelihood ratio that compares H_0 against H_1 is (by definition):

$$LR(a_i ... a_N) = \frac{L(1; a_i ... a_N)}{L(0; a_i ... a_N)} \tag{6.3.3}$$

$$= \frac{\prod_{i=1}^{N} (1-(1-e)(1-p))\, f_{A_i}(a_i; \mu_p, \sigma_a^2) + (1-e)(1-p) f_{A_i}(a_i; 0, \sigma_0^2)}{\prod_{i=1}^{N} e f_{A_i}(a_i; \mu_p, \sigma_a^2) + (1-e) f_{A_i}(a_i; 0, \sigma_0^2)} \tag{6.3.4}$$

If the response of the circuit then, is to weigh the relative odds of a flash vs no flash, the response $LR(a_i ... a_N)$ from the circuit provides a random variable based on which to perform optimal inference.

6.4 Performance of the Retinal Circuit

The performance of a decision based upon this likelihood ratio can be assessed when multiple trials of the flash experiment are presented. By making several simplifying assumptions, a decision making model that linearly sums the responses of the photodetectors can implement this likelihood ratio. By assuming that both rhodopsin events and noise convey the same standard deviation in response ($\sigma_a = \sigma_0 = \sigma$), and that a flash elicits a response in each rod cell without errors ($p = 1$, $e = 0$) we can rewrite the likelihood ratio as:

$$LR(a_i...a_N) = \frac{\prod_{i=1}^{N} f_{A_i}(a_i; \mu, \sigma^2)}{\prod_{i=1}^{N} f_{A_i}(a_i; 0, \sigma^2)} = \prod_{i=1}^{N} e^{\frac{(2a_i - \mu)\mu}{2\sigma^2}} \tag{6.4.1}$$

Here we have replaced $\mu_p = \mu$ for ease of notation. Taking log on both sides of this equation, we can write the log-likelihood ratio (LLR) expressed as the sum of the responses from each of the individual rod cells:

$$L = LLR(a_i...a_N) = \sum_{i=1}^{N} \frac{(2a_i - \mu)\mu}{2\sigma^2} \tag{6.4.2}$$

This quantity can be interpreted as a test statistic for inference via the likelihood ratio test, that can be "computed" across the responses of the individual photoreceptors via linear summation. We call this test statistic L. Next, we define a decision D rendered as a likelihood ratio test; D is a Bernoulli random variable that takes the value 0 on each trial, unless $L > \theta$:

	$d = 1$	$d = 0$
$P[D = d \mid L \leq l]$	$1 - F_L(\theta)$	$F_L(\theta)$

Once f_L, the PDF of L, is computed (based on the PDF's of the various $A_i's$ given by Equation **??**), we can find the PDF of D by integrating:

$$P[D = 0] = \int_{-\infty}^{\theta} f_L(l)dl \tag{6.4.3}$$

$$P[D = 1] = 1 - P[D = 0]$$

However, the PDF of each A_i is conditional on whether or not a flash was pre-

sented. Because of this, we can only compute these probabilities conditioned on whether or not a flash was presented. Supposing that $F = 0$, it follows that each $A_i \sim N(0, \sigma^2)$. Because of this,

$$\frac{(2A_i - \mu)\mu}{2\sigma^2} \sim N(-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2, 4\left(\frac{\mu}{\sigma}\right)^2) \tag{6.4.4}$$

Because L is a sum of N IID variables with this distribution, L is similarly gaussian with distribution:

$$L \sim N\left(-\frac{N}{2}\left(\frac{\mu}{\sigma}\right)^2, 4N\left(\frac{\mu}{\sigma}\right)^2\right) \tag{6.4.5}$$

Once θ is given, the conditional distribution can then be specified:

$$P[D = 1|F = 0] = \frac{1}{2}\text{Erfc}\left(\frac{N\mu^2 + 2\theta\sigma^2}{4\sqrt{2N}\mu\sigma}\right) \tag{6.4.6}$$

Now we consider the opposite possibility, that $F = 1$. In this case, each $A_i \sim N(\mu, \sigma^2)$. In the same fashion as above, we find that:

$$P[D = 1|F = 1] = \frac{1 + \text{Erfc}\left(\frac{N\mu^2 - 2\theta\sigma^2}{4\sqrt{2N}\mu\sigma}\right)}{2} \tag{6.4.7}$$

These two quantities are actually the false-positive and true-positive probabilities, and together form the ROC curve of the decision making problem as θ is varied. By solving the false-positive curve using the Erfc inverse formula, we plot true-positive vs. false-positive in Figure 6.2.

6.4.1 Synaptic Transfer Functions

The value held at the i^{th} rod cell is denoted as A_i, and in general must be shaped by a transfer function $g(A_i)$ so that it can be coordinated with values from other rod cells after the synapse to define the random variable L. L in turn provides the test statistic used to decide the presence or absence of a flash, on a given trial. According to the simplifying model assumptions in the previous section, this transfer function is linear; this is analogous to a linear synapse by rod cell onto the rod-bipolar cell, and enables a likelihood ratio test statistic via direct summation by

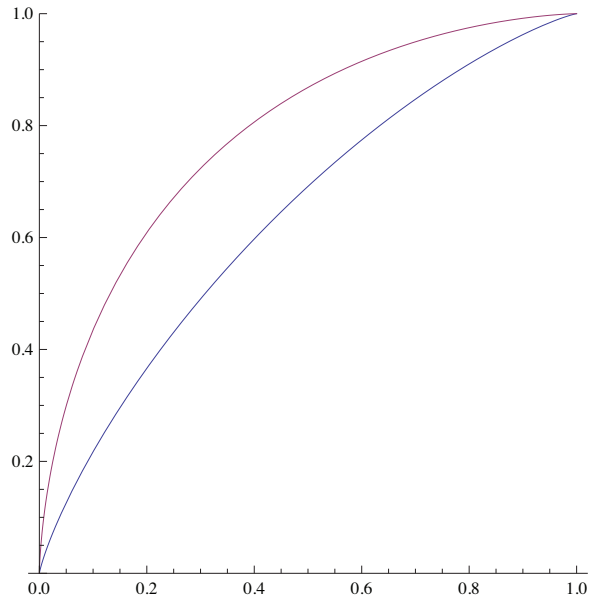


Figure 6.2: ROC curve, plotting P[D=1 | F=1] on the ordinate versus P[D=1 | F=0] on the abcissa. In blue, $N = 1$, in red $N = 5$; in both cases, $\mu = 1$ and $\sigma = 1$. As more photodetectors are included in the circuit, the response becomes increasingly reliable despite the noise inherent in the components.

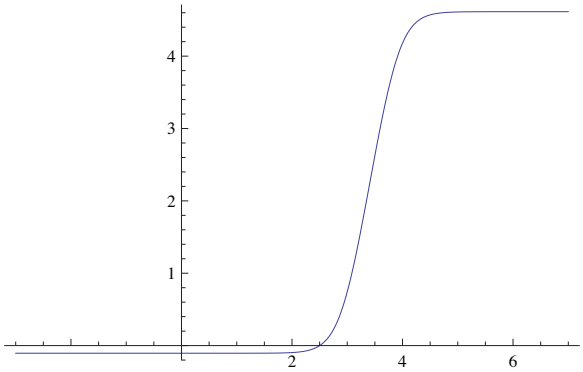


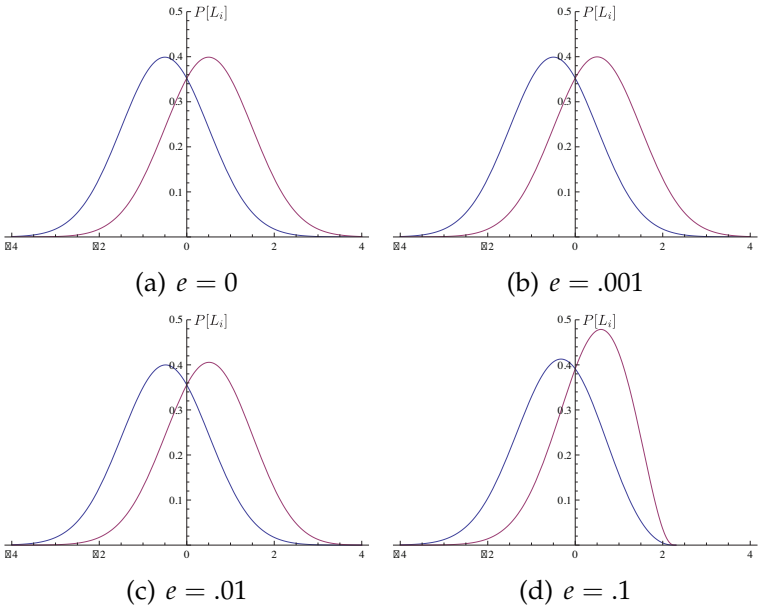
Figure 6.3: Sigmoidal transfer function $g(a_i)$, with $e = .001$, $p = .1$, $\mu = 5$, and $\sigma = 5$.

Equation **??**. We now ask the question of what happens when some of the modeling assumptions in the previous section are relaxed. We now allow discrete noise events ($e > 0$), and the more realistic assumption that not every rod cell receives a photon when a flash occurs. However, we maintain that the mean of A_i is the only change caused by a photon absorption.

The consequence of these two model changes is that the conditional probability distribution of A_i is no longer a gaussian, but rather a gaussian mixture. The likelihood ratio $LLR(a_1...a_N)$ can still be factored, however each individual factor does not have the cancellation that occurred in the equal-variance case:

$$LLR(a_i...a_N) = \sum_{i=1}^{N} g(a_i) \tag{6.4.8}$$

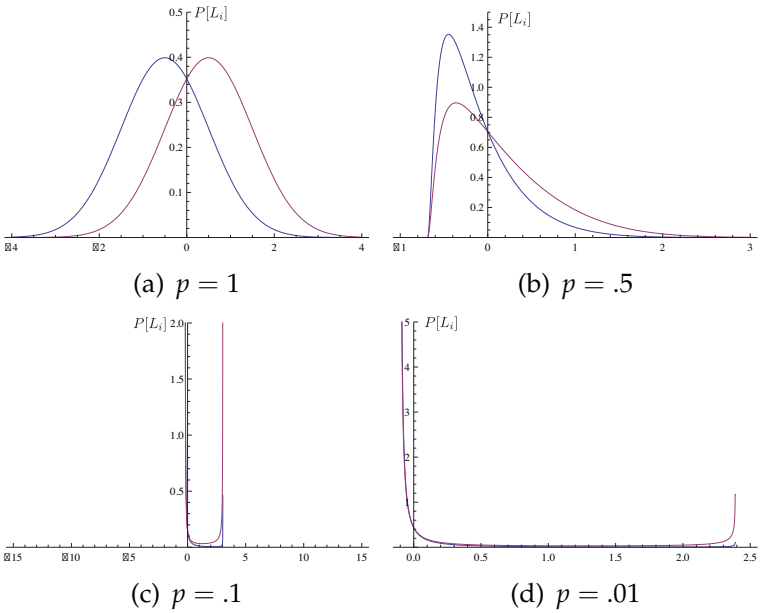
$$g(a_i) = \log\left(\frac{(1-(1-e)(1-p))\, f_{A_i}(a_i;\mu_p,\sigma^2) + (1-e)(1-p) f_{A_i}(a_i;0,\sigma^2)}{e f_{A_i}(a_i;\mu_p,\sigma^2) + (1-e) f_{A_i}(a_i;0,\sigma^2)}\right) \tag{6.4.9}$$



(a) $e = 0$ (b) $e = .001$

(c) $e = .01$ (d) $e = .1$

Figure 6.4: The effect o increasing e on the PDF of L_i, the random variable that is accumulated after being pushed through the transfer function $g(A_i)$. The two curves are the PDFs of L_i under the stimulus conditions (i.e., flash or no flash). A decision between these two alternatives can be made via a likelihood ratio test, which will establish how a given response should be classified. In all plots, $p = 1$, $\sigma = 1$, $\mu = 1$.



(a) $p = 1$ (b) $p = .5$

(c) $p = .1$ (d) $p = .01$

Figure 6.5: The effect of increasing p on the PDF of L_i, the random variable that is summed at the RBC after being pushed through the transfer function $g(A_i)$. In plots 6.5(a)-6.5(c), $e = 0$, $\sigma = 1$, $\mu = 1$; in 6.5(a) $\sigma = 1$, $\mu = 2$.

In order to compute the likelihood ratio test statistic via summation of the photoreceptor responses, the transfer function from the photoreceptors to the accumulator must be nonlinear. One can analytically compute the limits of this function as $a_i \to \pm\infty$:

$$\lim_{a_i \to \infty} g(a_i) = \log\left(1 + p\left(\frac{1}{e} - 1\right)\right) \tag{6.4.10}$$

$$\lim_{a_i \to -\infty} g(a_i) = \log\left(1 - p\right), \tag{6.4.11}$$

thereby inferring from this that $g(a_i)$ is sigmoidal. Figure 6.3 demonstrates that this is the case.

6.4.2 Extensions to the rod response model

In this final section, we briefly explore the effect of additional modifications to the assumptions on the parameters in the retinal response model. One can derive the PDF of a function of a random variable, given the inverse of the function. Given

$l = g(a_i)$ defined in the previous section, we find the inverse as:

$$a_i = g^{-1}(l) = \frac{\sigma^2 \log\left(\frac{(e-1)\left(e^l+p-1\right)e^{\frac{\mu^2}{2\sigma^2}}}{e(e^l-1)+p(e-1)}\right)}{\mu} \tag{6.4.12}$$

From this, we can define the pdf of the random variable produced by the photoreceptor synapse, that is eventually summed by the rod-bipolar cell :

$$f_{g(A_i)}(a_i) = |\frac{dg^{-1}}{da_i}(a_i)| f_{A_i}(g^{-1}(a_i)) \tag{6.4.13}$$

The distribution $f_{A_i}(a_i)$ is the PDF that describes the responses of the individual rod cells, after the transfer function, conditioned on whether or not a flash occurred ($F = 0$ or $F = 1$). The degree to which the distribution differs under these two conditions directly impacts the ability of the downstream circuit, receiving this signal, to make a decision about the presence or absence of a flash. We can plot this distribution for various parameter choices under each of the two possibilities, and get a sense of how changes to p, the probability that a photoreceptor is activated by a flash, and e, the probability of an photon-like error event, affect the likelihood random variable L.

Figure 6.4 demonstrates how changing e, the probability of a photon-like noise event, affects the distribution of the random variable summed by the rod-bipolar cell (blue). When $e = 0$, we see that the PDF's under the flash and no flash condition are Gaussian. Discrimination between these two possibilities via a likelihood ratio test continues to be a classification problem, although with more complicated alternative distributions. As e increases, the location of a line on the abscissa that classifies a flash event from a no-flash event must change. Figure 6.5 demonstrates how changing p , the probability that a flash triggers an isomerization event in a photoreceptor, affects the same distributions. These plots demonstrate that the expected response that must be summed by an accumulator computing a likelihood ratio test statistic can vary significantly as the functional properties of the upstream circuitry are varied. Although a strong nonlinearity is still necessary to classify flashes from no-flashes, the location and performance of this scheme is

greatly affected by the overlap of the distributions under the two possibilities, as the parameters in the model of retinal circuitry changes.

6.5 Summary

What is the role of the the retinal circuitry that accumulates rod responses at low-light levels? This modeling study operates under the *assumption* that the rod bipolar cell, which accumulates the signals of many inherently noisy photoreceptors (rod cells), computes a likelihood ratio test statistic. Assuming this computational role, what is the nature of the synaptic transmission that enables the computation? We found that under the assumption of perfect transmission (i.e, every photoreceptor is activated by a flash, and no false photo-isomerization events occur), a linear synapse combined with linear summation can reconstruct an likelihood ratio test statistic from which a decision about the presence or absence of a flash can be inferred.

However, when this idealized case is made more realistic, by assuming that few photodetectors are triggered by a flash event, a nonlinear transfer function is necessary to reconstruction the log-likelihood ratio via postsynaptic summation. This prediction agrees with observations made by Field and Rieke [34], where nonlinear rod→rod-bipolar signal transmission with a sigmoidal shape was observed in rod bipolar cell responses. One interpretation of this fact is that this nonlinearity "retains signals from those rods likely generating single-photon responses and rejects signals from those rods likely generating noise" [35], thereby optimizing the signal-to-noise ratio for transmission downstream. Our modeling study suggests another, complimentary interpretation to the mechanism: that the same thresholding nonlinearity enables the construction of likelihood ratio test statistic that overcomes the limitations of both the environment and photoreceptor machinery, and can be used as a signal to decide whether or not a flash has occurred.

6.6 Computation of speed and accuracy with overshoot

6.6.1 Introduction

Numerical computation of the accuracy and mean decision time of a drift diffusion model (DDM) can be framed as computation of the sample mean of two random variables. Accuracy is the mean of a Bernoulli RV that decides whether the "correct" threshold (i.e., the threshold favored by the mean of the sampling distribution). Similarly, mean decision time is the mean of distribution of the number of samples necessary to cross one of the two thresholds. Naively, these means can be estimated by repeatedly sampling these distributions N times via Monte Carlo simulation.

This direct approach will result in an estimator of FC with a standard deviation of:

$$\sigma_{FC} = \sqrt{\frac{FC(1-FC)}{N}} \tag{6.6.1}$$

after N IID trials (The variance of DT will rely on the distribution of decision times, which is model-dependent). This method utilizes no information about the sampling distribution, and as a consequence requires a rather large number of trials to achieve accurate estimates.

The problem of this slow convergence has been a topic of research in the past. Generally, the topic of variance reduction is concerned with increasing the precision of a Monte-Carlo estimate of a parameter, for a fixed number of trails. Sigmund [106] presents a particular technique for variance reduction known as importance sampling, whereby the variance of an estimator can be reduced by only sampling the distribution at places that impact the estimator, and then correcting for the bias introduced [49]. Using this technique,[106] formulates an algorithm that converges much more rapidly to the true value of accuracy than the naive estimation method.

One of the reasons such are necessary in the first place is that the accuracy and

decision time formulas derived by Wald [123]:

$$FC \approx \frac{1}{1+e^{h_0\theta}} \tag{6.6.2}$$

$$DT \approx \frac{\theta}{E[x_i]}\tanh\left(\frac{-h_0\theta}{2}\right) \tag{6.6.3}$$

are only approximate, the "overshoot problem" (Here h_0 is the nonzero root of the moment generating function of the sampling variable x_i set to unity). An exact formulation is expressed in terms of a random variable that expresses the amount that the accumulation exceeded the threshold:

$$FC = 1 - \frac{E[e^{h_0 E_n}|E_n \geq \theta] - 1}{E[e^{h_0 E_n}|E_n \geq \theta] - E[e^{h_0 E_n}|E_n \leq -\theta]} \tag{6.6.4}$$

$$DT = \frac{1}{E[x_i]}\left(E[E_n|E_n \geq \theta](FC) + E[E_n|E_n \leq -\theta](1-FC)\right) \tag{6.6.5}$$

The approximations are recovered when $E[e^{h_0 E_n}|E_n \geq \theta] \approx e^{h_0\theta}$, $E[e^{h_0 E_n}|E_n \leq \theta] \approx e^{-h_0\theta}$, $E[E_n|E_n \geq -\theta] \approx \theta$, and $E[E_n|E_n \leq -\theta] \approx -\theta$ are assumed, i.e. that the random variable E_n takes the value of θ or $-\theta$ on the threshold crossing step (and hence no overshoot over θ).

Consider repeated trials of a DDM with an continuous valued sampling distribution, with non-negligible overshoot over the decision boundary. In such a sequential sampling model, direct application of Equations 6.6.2 and 6.6.3 may result in a poor approximation to the true values of FC and RT. One alternative is estimation of the quantities via direct Monte-Carlo simulation, applying naive estimates for the sample mean. A second alternative is to apply Monte-Carlo importance sampling via [106], gaining faster convergence as more trials are performed (owing to a smaller variance on this estimator). The next section outlines a third alternative that exploits even more of the properties of the DDM, specifically Wald's exact formulas for FC and DT, that in many cases outperforms both of these previous options.

6.6.2 Faster Monte-Carlo convergence via Wald's Identities

An algorithm for an estimator of FC and DT that in many cases outperforms both the naive estimator and importance sampling, relying on knowledge of the increment distribution, is presented first. Numerical evidence for this assertion is provided in Figure 6.6, and in Table 6.1. Proof of the convergence rate of this algorithm in terms of the moments of overshoot distribution is left as future work.

On each Monte-Carlo trial of DDM, more information is available than simply noting which threshold (θ or $-\theta$) was crossed, and on which step this first crossing took place. Specifically, each trial provides a single sample from the overshoot distribution. (Were this not the case, a direct application of Equations 6.6.2 and 6.6.2 would obviate the need for Monte-Carlo simulation in the first place!) Using these samples, one can construct naive estimators for the expectations in Equations 6.6.4 and 6.6.5. For IID samples, let $E_t = \sum_{i=1}^{t} x_t$, and for $-\theta < 0 < \theta$ define:

$$T = \inf\{t : t \geq 1, E_t \notin (-\theta, \theta)\}. \tag{6.6.6}$$

Also define:

$$k = \sum_{i=1}^{n} I\left(E_N^i > \theta\right) \tag{6.6.7}$$

where $I(A)$ is the indicator function and i indexes one of n Monte-Carlo trials (The naive estimator for FC is then k/n). Now let:

$$O^+ = \frac{1}{k} \sum_{i=1}^{n} I\left(E_N^i > \theta\right) e^{h_0 E_N^i} \tag{6.6.8}$$

$$O^- = \frac{1}{n-k} \sum_{i=1}^{n} I\left(E_N^i < \theta\right) e^{h_0 E_N^i} \tag{6.6.9}$$

(Note that these quantities may be undefined if either $k = 0$ or $n - k = 0$.) We then have an estimator for accuracy $\hat{\alpha}$ that converges via Equation 6.6.4:

$$\hat{\alpha} = 1 - \frac{O^+ - 1}{O^+ - O^-} \tag{6.6.10}$$

$$\implies \lim_{n \to \infty} \hat{\alpha} = FC. \tag{6.6.11}$$

Similarly, defining:

$$M^+ = \frac{1}{k}\sum_{i=1}^{n} I\left(E_N^i > \theta\right) E_N^i \tag{6.6.12}$$

$$M^- = \frac{1}{n-k}\sum_{i=1}^{n} I\left(E_N^i < \theta\right) E_N^i \tag{6.6.13}$$

and using our estimator for accuracy, an estimator for DT follows naturally from Equation 6.6.5:

$$\hat{T} = \frac{1}{E[x_i]}\left(M^+\hat{\alpha} + M^-(1-\hat{\alpha})\right) \tag{6.6.14}$$

$$\implies \lim_{n\to\infty} \hat{T} = RT. \tag{6.6.15}$$

6.6.3 Numerical evidence

In order to concretely compare the performance of $\hat{\alpha}$ (as an estimator of accuracy) against the performance of the naive estimator, and the importance sampling estimator provided in [106], I used Monte-Carlo sampling of DDM with symmetric threshold θ, with a sequentially sampled IID from $x_i \sim N(\mu, 1)$. For n trials, the naive estimator,

$$\alpha_0 = \frac{1}{n}\sum_{i=1}^{n} I\left(E_N^i > \theta\right), \tag{6.6.16}$$

and the importance sampling estimator,

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} I\left(E_N^i > \theta\right) e^{-2\mu E_N^i}, \tag{6.6.17}$$

were both often outperformed by the "overshoot" estimator $\hat{\alpha}$ defined in Equation 6.6.10. The cases examined were chosen to correspond to those reported in [106] (Table 1) and are summarized in Table 6.1. To summarize, I found that as the accuracy of the DDM increased (either by an increase in threshold, or in drift rate μ), each estimator obtained a smaller standard deviation. However, the importance sampling and the overshoot estimators provided a significantly better method (a relative efficiency as high as ~ 50000). In the other limit, as accuracy is decreased, while both the importance sampling and overshoot estimators continued to be superior the naive estimator, the overshoot method begins to be superior

μ	θ	FC	$\bar{\sigma}_{\alpha_0}$	$\bar{\sigma}_{\alpha}$	$\bar{\sigma}_{\hat{\alpha}}$	$\bar{e}(\bar{\sigma}_{\alpha_0}, \bar{\sigma}_{\alpha})$	$\bar{e}(\bar{\sigma}_{\alpha}, \bar{\sigma}_{\hat{\alpha}})$	$\bar{e}(\bar{\sigma}_{\alpha_0}, \bar{\sigma}_{\hat{\alpha}})$
.5	9	.9999	1.17e-3	4.40e-6	5.16e-6	70600	.738	51400
.25	9	.992	1.27e-2	2.94e-4	2.68e-4	1880	1.20	2260
.125	9	.916	3.92e-2	3.88e-3	1.43e-3	102	7.26	743
.5	5	.996	8.61e-3	2.42e-4	2.41e-4	1260	1.01	1260
.25	5	.940	3.32e-2	2.84e-4	1.96e-3	136	2.10	286
.125	5	.799	5.64e-2	1.45e-2	3.74e-3	15.1	15.0	228

Table 6.1: Overshoot estimator ($\hat{\alpha}$) outperforms naive estimation (α_0) and an importance sampling (α) estimator of accuracy (FC). For different values of drift rate, μ and threshold, θ, the standard deviation of each estimator σ and numerically computed accuracy is provided after $n = 50$ Monte-Carlo trials ($\bar{\sigma}$ is the sample mean value over 100,000 repeats of this Monte-Carlo scheme). The relative efficiencies $e(\alpha_1, \alpha_2) = \sigma^2_{\alpha_1}/\sigma^2_{\alpha_2}$ comparing α_0 vs. α, α vs. $\hat{\alpha}$, and α_0 vs. $\hat{\alpha}$ suggest that, for these values of the DDM, the overshoot estimation technique provides a significant performance increase (by providing the ratio of the number of trials needed by each method to obtain a given accuracy level).

to the importance sampling method.

Figure 6.6A illustrates the $1/\sqrt{n}$ convergence of the standard deviation of each of the three estimators. At $\mu = .125$, $\theta = 5$, the condition with lowest FC, the overshoot estimator performs best. As suggested in Table 6.1, this estimator outperforms the importance sampling estimator by the greatest extent at low FC, but may not do so at higher FC. Figure 6.6B shows the convergence of each of the three estimators at an increasing number of Monte-Carlo trials; remarkably, the overshoot estimator provides a reasonable estimate for accuracy after only a few trials. Similar results are shown in Figure 6.6C, where the estimate of FC has been combined with estimates from Equations 6.6.12 and 6.6.13 via Equation 6.6.14. Counterintuitively, this technique produces a reasonable estimate of DT without ever actually recording the number of steps required to reach threshold on any of the n Monte-Carlo trials. With more work, these estimators may prove a valuable theoretical tool for investigating the properties of sequential sampling models.

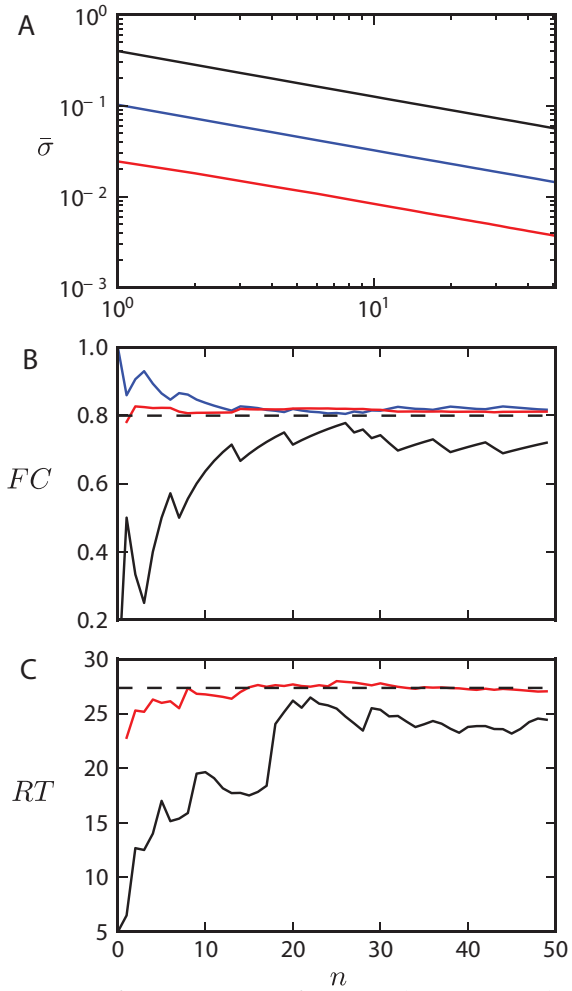


Figure 6.6: Convergence of estimators of FC and DT in a drift-diffusion model, $\mu = .125$, $\theta = 5$. (A) Numerically computed (50000 trials) standard deviation of FC estimators, α_0 (black), α (blue), $\hat{\alpha}$ (red). The overshoot estimator has the smallest standard deviation, and therefore provides the best estimate. (B) and (C) As the number of Monte-Carlo trials increases, the estimations of FC and RT become more reliable, converging the the (numerically computed) true values. The overshoot estimator provides a reasonable estimate of accuracy after remarkably few trials.

References

[1] AM Aertsen, GL Gerstein, MK Habib, and G. Palm. Dynamics of neuronal firing correlation: modulation of" effective connectivity". *Journal of Neurophysiology*, 61(5):900–917, 1989.

[2] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, May 2006.

[3] Wyeth Bair, Ehud Zohary, and William T Newsome. Correlated firing in macaque visual area MT: time scales and relationship to behavior. *Journal of Neuroscience*, 21(5):1676, 2001.

[4] Jeffrey M Beck, Wei Ji Ma, Roozbeh Kiani, Tim Hanks, Anne K Churchland, Jamie Roitman, Michael N Shadlen, Peter E Latham, and Alexandre Pouget. Probabilistic Population Codes for Bayesian Decision Making. *Neuron*, 60(6):1142–1152, January 2008.

[5] Patrick Billingsley. *Probability and measure*. Wiley-Interscience, 1986.

[6] M.D. Binder and R.K. Powers. Relationship between simulated common synaptic input and discharge synchrony in cat spinal motoneurons. *Journal of Neurophysiology*, 86(5):2266–2275, 2001.

[7] Martin Boerlin and Sophie Deneve. Spike-based population coding and working memory. *PLoS Computational Biology*, 7(2):e1001080, February 2011.

[8] R Bogacz. Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, 11(3):118–125, 2007.

[9] Rafal Bogacz, Eric Brown, Philip Holmes, and Jonathan D Cohen. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4):700–765, 2006.

[10] David R. Brillinger. The calculation of cumulants via conditioning. *Ann Inst Stat Math Annals of the Institute of Statistical Mathematics*, 21(1):215–218, 1969.

[11] K H Britten, W T Newsome, M N Shadlen, S. Celebrini, and J A Movshon. A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, 13:87–100, 1996.

[12] Kenneth H Britten, Michael N Shadlen, William T Newsome, and J Anthony Movshon. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, 12(12):4745–4765, 1992.

[13] Kenneth H Britten, Michael N Shadlen, William T Newsome, and J Anthony Movshon. Responses of neurons in macaque MT to stochastic motion signals. *Visual Neuroscience*, 10(06):1157–1169, 1993.

[14] Eric Brown, Juan Gao, Philip Holmes, Rafal Bogacz, Mark Gilzenrat, and Jonathan D Cohen. Simple neural networks that optimize decisions. *International Journal of Bifurcation Chaos in Applied Sciences and Engineering*, 15(3):803–826, 2005.

[15] Eric Brown and Philip Holmes. Modelling a simple choice task: Stochastic dynamics of mutually inhibitory neural groups. *Stochastics and Dynamics*, 1(2):159–191, 2001.

[16] Nicholas Cain, Andrea Barreiro, Mike Shadlen, and Eric Shea-Brown. A favorable tradeoff between robustness and performance in sequential decision tasks. In *COSYNE: 2011*, Salt Lake City, February 2011.

[17] Nicholas Cain and Eric Shea-Brown. Computational models of decision making: integration, stability, and noise. *Current Opinion In Neurobiology*, pages 1–7, May 2012.

[18] Stephen C Cannon and David A Robinson. A proposed neural network for the integrator of the oculomotor system. *Biological Cybernetics*, 49(2):127–136, December 1983.

[19] Anne K Churchland, R Kiani, R Chaudhuri, Xiao-Jing Wang, Alexandre Pouget, and M N Shadlen. Variance as a Signature of Neural Computations during Decision Making. *Neuron*, 69(4):818–831, February 2011.

[20] Anne K Churchland, Roozbeh Kiani, and Michael N Shadlen. Decision-making with multiple alternatives. *Nature Neuroscience*, 11(6):693–702, 2008.

[21] Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature Publishing Group*, 14(7):811–819, June 2011.

[22] Marlene R Cohen and William T Newsome. Context-Dependent Changes in Functional Circuitry in Visual Area MT. *Neuron*, 60(1):162–173, October 2008.

[23] Marlene R Cohen and William T Newsome. Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. *Journal of Neuroscience*, 29(20):6635–6648, 2009.

[24] T.F. Coleman and Y. Li. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Mathematical programming*, 67(1):189–224, 1994.

[25] Thomas F. Coleman and Yuying Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, 6(2):418–445, 1996.

[26] Jaime De La Rocha, Brent Doiron, Eric Shea-Brown, Krešimir Josić, and Alex Reyes. Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806, August 2007.

[27] Gustavo Deco, Edmund T Rolls, and Ranulfo Romo. Stochastic dynamics as a principle of brain function. *Progress in Neurobiology*, 88(1):1–16, May 2009.

[28] Jochen Ditterich. Evidence for time-variant decision making. *European Journal of Neuroscience*, 24(12):3628–3641, 2006.

[29] Kenji Doya, editor. *Bayesian brain: probabilistic approaches to neural codingâĂŐ*. The MIT Press, 2007.

[30] P Eckhoff, K F Wong-Lin, and P Holmes. Optimality and Robustness of a Biophysical Decision-Making Model under Norepinephrine Modulation. *Journal of Neuroscience*, 29(13):4301–4311, April 2009.

[31] P Eckhoff, KF Wong-Lin, and P Holmes. Dimension Reduction and Dynamics of a Spiking Neural Network Model for Decision Making under Neuromodulation. *Dimension*, 10(1):148–188, July 2011.

[32] Ward Edwards. Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2(2):312–329, 1965.

[33] G D Field and E J Chichilnisky. Information Processing in the Primate Retina: Circuitry and Coding. *Annual Review of Neuroscience*, 30(1):1–30, July 2007.

[34] GD Field and F Rieke. Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity. *Neuron*, 34(5):773–785, 2002.

[35] GD Field, AP Sampath, and F Rieke. Retinal processing near absolute threshold: from behavior to mechanism. *Annual Review of Physiology*, 67:491–514, 2005.

[36] S Ganguli, D Huh, and H Sompolinsky. Memory traces in dynamical systems. *Proceedings of the National Academy of Sciences*, 105(48):18970, 2008.

[37] Surya Ganguli and Peter Latham. Feedforward to the Past: The Relation between Neuronal Connectivity, Amplification, and Short-Term Memory. *Neuron*, 61(4):499–501, January 2009.

[38] E Ganmor, R Segev, and E Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proceedings of the National Academy of Sciences*, 108(23):9679, 2011.

[39] Crispin Gardiner. *Handbook of stochastic methods: for physics, chemistry and the natural sciences*. Springer, 2002.

[40] B. K. Ghosh and Pranab Kumar Sen. *Handbook of sequential analysis*. M. Dekker, New York, 1991.

[41] Daniel T Gillespie. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Physical Review E*, 54(2):2084–2091, 1996.

[42] Joshua I Gold and Michael N Shadlen. Banburismus and the Brain Decoding the Relationship between Sensory Stimuli, Decisions, and Reward. *Neuron*, 36(2):299–308, 2002.

[43] Joshual I Gold and Michael N Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30:535–574, 2007.

[44] Mark S Goldman. Memory without Feedback in a Neural Network. *Neuron*, 61(4):621–634, January 2009.

[45] Mark S Goldman, Joseph H Levine, Guy Major, David W Tank, and H S Seung. Robust Persistent Neural Activity in a Model Integrator with Multiple Hysteretic Dendrites per Neuron. *Cerebral Cortex*, 13(11):1185–1195, November 2003.

[46] MS Goldman, A Compte, and X.J. Wang. Neural integrator models. *In: Squire LR (ed.) Encyclopedia of Neuroscience*, 6:165–178, 2009.

[47] David Green and John Swets. *Signal detection theory and psychophysics*. Peninsula Pub, 1966.

[48] D A Gutnisky and K Josic. Generation of Spatiotemporally Correlated Spike Trains and Local Field Potentials Using a Multivariate Autoregressive Process. *Journal of Neurophysiology*, 103(5):2912–2930, May 2010.

[49] J. M. Hammersley and D. C. Handscomb. *Monte Carlo methods*. Methuen; Wiley /, London; New York, 1964.

[50] Desmond J Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[51] X Huang and S G Lisberger. Noise Correlations in Cortical Area MT and Their Potential Impact on Trial-by-Trial Variation in the Direction and Speed of Smooth-Pursuit Eye Movements. *Journal of Neurophysiology*, 101(6):3012–3030, May 2009.

[52] Alexander C Huk and Michael N Shadlen. Neural Activity in Macaque Parietal Cortex Reflects Temporal Integration of Visual Motion Signals during Perceptual Decision Making. *Journal of Neuroscience*, 25(45):10420–10436, November 2005.

[53] John Kemeny and J Snell. *Finite markov chains*. D. Van Nostrand, 1960.

[54] Roozbeh Kiani, Timothy D Hanks, and Michael N Shadlen. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, 28(12):3017–3029, March 2008.

[55] Soyoun Kim, Jaewon Hwang, and Daeyeol Lee. Prefrontal Coding of Temporally Discounted Values during Intertemporal Choice. *Neuron*, 59(1):161–172, July 2008.

[56] Alexei A Koulakov, Sridhar Raghavachari, Adam Kepecs, and John E Lisman. Model for a robust neural integrator. *Nature Neuroscience*, 5(8):775–782, August 2002.

[57] S.W. Kuffler. Discharge patterns and functional organization of mammalian retina. *J Neurophysiol*, 16(1):37–68, 1953.

[58] A Kuhn, A Aertsen, and S Rotter. Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation*, 15(1):67–101, 2003.

[59] D Laming. *Information theory of choice-reaction times* . Academic Press New York, 1968.

[60] Peter E Latham and Yasser Roudi. Role of correlations in population coding. *arXiv*, q-bio.NC, 2011.

[61] J. Lee, C. Park, and B. Kim. An Estimation Method for the Excess over the Boundaries in the SPRT and Its Applications. *Sequential Analysis*, 13(2):127–144, 1994.

[62] Sukbin Lim and Mark S Goldman. Noise tolerance of attractor and feedforward memory models. *Neural Computation*, 24(2):332–390, February 2012.

[63] S W Link and R A Heath. A sequential theory of psychological discrimination. *Psychometrika*, 40(1):77–105, March 1975.

[64] Stephen W Link. *The wave theory of difference and similarity*. Lawrence Erlbaum, 1992.

[65] JE Lisman, JM Fellous, and Xian-Jing Wang. A role for NMDA-receptor channels in working memory. *Nature Neuroscience*, 1998.

[66] R Luce. A Threshold Theory for Simple Detection Experiments. *Psychological Review*, 70(1):61–79, 1963.

[67] R Luce. *Response times: their role in inferring elementary mental organization*. Oxford University Press, Oxford psychology series, no. 8, 1986.

[68] Christian K Machens, Ranulfo Romo, and Carlos D Brody. Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science*, 307(5712):1121–1124, 2005.

[69] W T Maddox and C J Bohil. Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of experimental psychology. Learning, memory, and cognition*, 24(6):1459–1482, November 1998.

[70] Mark E Mazurek, Jamie D Roitman, Jochen Ditterich, and Michael N Shadlen. A Role for Neural Integrators in Perceptual Decision Making. *Cerebral Cortex*, 13(11):1257–1269, November 2003.

[71] Mark E Mazurek and Michael N Shadlen. Limits to the temporal fidelity of cortical spike rate signals. *Nature Neuroscience*, 5(5):463–471, April 2002.

[72] Gail McKoon and Roger Ratcliff. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4):873–922, 2008.

[73] Tyler McMillen and Philip Holmes. The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, 50(1):30–57, February 2006.

[74] P Miller. Power-Law Neuronal Fluctuations in a Recurrent Network Model of Parametric Working Memory. *Journal of Neurophysiology*, 95(2):1099–1114, October 2005.

[75] Paul Miller and Xiao-Jing Wang. Power-law neuronal fluctuations in a recurrent network model of parametric working memory. *Journal of Neurophysiology*, 95(2):1099–1114, February 2006.

[76] Andrew Miri, Kayvon Daie, Aristides B Arrenberg, Herwig Baier, Emre Aksay, and David W Tank. Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nature Publishing Group*, 14(9):1150–1159, August 2011.

[77] F Montani, R A A Ince, R Senatore, E Arabzadeh, M E Diamond, and S Panzeri. The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1901):3297–3310, July 2009.

[78] R Moreno-Bote. Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural Computation*, 22(7):1786–1811, 2010.

[79] William T Newsome, Kenneth H Britten, J Anthony Movshon, and Michael N Shadlen. Single neurons and the perception of motion. In Dominic Man-Kit Lam and Charles D. Gilbert, editors, *Neural mechanisms of visual perception*, pages 171–198. Portfolio Pub. Co., Woodlands, Tex., April 1989.

[80] E. Niebur. Generation of synthetic spike trains with defined pairwise correlations. *Neural Computation*, 19(7):1720–1738, 2007.

[81] Maxim Nikitchenko and Alexei Koulakov. Neural integrator: A sandpile model. *Neural Computation*, 20(10):2379–2417, 2008.

[82] John Palmer, Alexander C Huk, and Michael N Shadlen. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5(5):376–404, 2005.

[83] Amber Polk, Ashok Litwin-Kumar, and Brent Doiron. Correlated neural variability in persistent state networks. *Proceedings of the National Academy of Sciences*, 2012.

[84] Alexandre Pouget and Peter Latham. Digitized neural networks: long-term stability from forgetful neurons. *Nature Neuroscience*, 5(8):709–710, 2002.

[85] B A Purcell, J D Schall, G D Logan, and T J Palmeri. From Salience to Saccades: Multiple-Alternative Gated Stochastic Accumulator Model of Visual Search. *Journal of Neuroscience*, 32(10):3433–3446, March 2012.

[86] Braden A Purcell, Richard P Heitz, Jeremiah Y Cohen, Jeffrey D Schall, Gordon D Logan, and Thomas J Palmeri. Neurally constrained modeling of perceptual decision making. *Psychological Review*, 117(4):1113–1143, 2010.

[87] David Raposo, John P Sheppard, Paul R Schrater, and Anne K Churchland. Multisensory decision-making in rats and humans. *The Journal of Neuroscience*, 32(11):3726–3735, March 2012.

[88] R Ratcliff, T Van Zandt, and G McKoon. Connectionist and diffusion models of reaction time. *Psychological Review*, 106(2):261–300, April 1999.

[89] Roger Ratcliff. A Theory of Memory Retrieval. *Psychological Review*, 85(2):59–108, 1978.

[90] Roger Ratcliff and Jeffrey N Rounder. Modeling response times for two-choice decisions. *Psychological Science*, 9:347–356, 1998.

[91] F Rieke, WG Owen, and W Bialek. Optimal filtering in the salamander retina. *Proceedings of the 1990 conference on Advances in neural information processing systems 3*, pages 377–383, 1990.

[92] D.A. Robinson. Integrating with Neurons. *Annual Review of Neuroscience*, 12(1):33–45, 1989.

[93] Jamie D Roitman and Michael N Shadlen. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21):9475–9489, 2002.

[94] Ranulfo Romo, Adam Kepecs, and Carlos D Brody. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current Opinion In Neurobiology*, 13(2):204–211, 2003.

[95] A Roxin and A Ledberg. Neurobiological models of two-choice decision making can be reduced to a one-dimensional .... *PLoS Computational Biology*, 2008.

[96] E Salinas and TJ Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2(8):539–550, 2001.

[97] C Daniel Salzman, Chieko M Murasugi, Kenneth H Britten, and William T Newsome. Microstimulation in visual area MT: effects on direction discrimination performance. *The Journal of Neuroscience*, 12(6):2331–2355, 1992.

[98] Jeffrey D Schall. Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, 2(1):33–42, 2001.

[99] Jeffrey D Schall, Braden A Purcell, Richard P Heitz, Gordon D Logan, and Thomas J Palmeri. Neural mechanisms of saccade target selection: gated accumulator model of the visual-motor cascade. *The European journal of neuroscience*, 33(11):1991–2002, June 2011.

[100] H S Seung. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23):13339–13344, 1996.

[101] H Sebastian Seung, Daniel D Lee, Ben Y Reis, and David W Tank. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron*, 26(1):259–271, 2000.

[102] Michael N Shadlen, Kenneth H Britten, William T Newsome, and J Anthony Movshon. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *The Journal of Neuroscience*, 16(4):1486–1510, February 1996.

[103] Michael N Shadlen and William T Newsome. Motion perception: seeing and deciding. *Proceedings of the National Academy of Sciences*, 93:628–633, 1996.

[104] Michael N Shadlen and William T Newsome. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86(4):1916–1936, 2001.

[105] MN Shadlen and WT Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998.

[106] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics*, pages 673–684, 1976.

[107] Patrick Simen, Fuat Balci, Laura de Souza, Jonathan D Cohen, and Philip Holmes. A model of interval timing by neural integration. *The Journal of Neuroscience*, 31(25):9238–9253, June 2011.

[108] M A Smith and A Kohn. Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex. *Journal of Neuroscience*, 28(48):12591–12603, November 2008.

[109] P Smith. From Poisson shot noise to the integrated Ornstein-Uhlenbeck process: Neurally principled models of information accumulation in decision-making and response time. *Journal of Mathematical Psychology*, 54:266–283, 2010.

[110] Philip L Smith and Cameron R L McKenzie. Diffusive Information Accumulation by Minimal Recurrent Neural Models of Decision Making. *Neural Computation*, pages 1–33, April 2011.

[111] Philip L Smith and Roger Ratcliff. Psychology and neurobiology of simple decisions. *TRENDS in Neurosciences*, 27(3):161–168, 2004.

[112] Philip L Smith and Douglas Vickers. Modeling evidence accumulation with partial loss in expanded judgment. *Journal of Experimental Psychology: Human Perception and Performance*, 15(4):797, 1989.

[113] RG Smith. Noise removal at the rod synapse of mammalian retina. *Visual Neuroscience*, 15(5):809–821, 1998.

[114] WR Softky and C Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1):334–350, 1993.

[115] Larry R. Squire. *Fundamental neuroscience*. Elsevier / Academic Press, 3 edition, 2008.

[116] Benjamin Staude, Sonja Grün, and Stefan Rotter. Higher-Order Correlations and Cumulants. In Sonja Grün and Stefan Rotter, editors, *Analysis of Parallel Spike Trains*, pages 253–280. Springer US, 2010.

[117] Mervyn Stone. Models for choice-reaction time. *Psychometrika*, 25:251–260, 1960.

[118] J A Swets. Is there a sensory threshold. *Science*, 134(3473):168–177, 1961.

[119] Henry C. Tuckwell. *Stochastic processes in the neurosciences*. Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1989.

[120] M Usher. Hick's Law in a Stochastic Race Model with Speed–Accuracy Tradeoff. *Journal of Mathematical Psychology*, 46(6):704–715, December 2002.

[121] Marius Usher and James L McClelland. The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3):550–592, July 2001.

[122] D Vickers. Evidence for an Accumulator Model of Psychophysical Discrimination. *Ergonomics*, 13(1):37–58, 1970.

[123] A Wald. On cumulative sums of random variables. *Annals Of Mathematical Statistics*, 15:342–342, 1944.

[124] A Wald. Sequential Tests of Statistical Hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

[125] A Wald and J Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948.

[126] Xiao-Jing Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, 36(5):955–968, 2002.

[127] Xiao-Jing Wang. Decision Making in Recurrent Neuronal Circuits. *Neuron*, 60(2):215–234, October 2008.

[128] Tristan J. Webb, Edmund T Rolls, Gustavo Deco, and Jianfeng Feng. Noise in Attractor Networks in the Brain Produced by Graded Firing Rate Representations. *PLoS One*, 6(9):e23630, September 2011.

[129] KF Wong, AC Huk, MN Shadlen, and Xiao-Jing Wang. Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Frontiers in Computational Neuroscience*, 1, 2007.

[130] Kong-Fatt Wong and Xiao-Jing Wang. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.

[131] S. Yu, H. Yang, H. Nakahara, G.S. Santos, D. Nikolić, and D. Plenz. Higher-order interactions characterized in cortical activity. *The Journal of Neuroscience*, 31(48):17514–17526, 2011.

[132] M Zacksenhouse, R Bogacz, and P Holmes. Robust versus optimal strategies for two-alternative forced choice tasks. *Journal of Mathematical Psychology*, 54(2):230–246, March 2010.

[133] Jiaxiang Zhang and Rafal Bogacz. Optimal Decision Making on the Basis of Evidence Represented in Spike Trains. *Neural Computation*, 22(5):1113–1148, 2010.

[134] Ehud Zohary, Michael N Shadlen, and William T Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370:140–143, 1994.