

# **Neuromorphic Models of Response Time**

Andrew Heathcote

Department of Psychology, The University of Newcastle, Australia.

email: [heathcote@psychology.newcastle.edu.au](mailto:heathcote@psychology.newcastle.edu.au)

**Abstract**

A new approach to modeling response time is proposed, based on the activation dynamics of simplified neural units. The proposed neuromorphic decision units are dissipative and have bounded activation, as do real neurons. A decision is made either when the unit's activation exceeds a threshold or when it converges, using a criterion based on the derivative of activation. First, the relationship between varieties of neuromorphic and sequential sampling models is reviewed. Mathematical and simulation results are presented both for deterministic and random versions of the neuromorphic units. These results are used to highlight strengths and weaknesses of this new approach to dynamic models of decision making.

## Neuromorphic Models of Response Time

The models developed here were inspired by the similarity between sequential sampling models of choice and response time and first order approximations to neural dynamics, such as those used by Carpenter and Grossberg (1987) and Kohonen (1993). The term “neuromorphic” reflects the fact that the models are based on the (simplified) structure of biological nervous systems. The next section reviews sequential sampling models and a number of theoretical and empirical issues that have driven their development. The following section motivates the structure of the neuromorphic models with reference to empirical issues and constraints from physical instantiation. The next two sections develop the mathematical detail of the models and the final section discusses their strengths and weaknesses.

### Sequential Sampling Models

The following brief review can not do justice to the scope of sequential sampling models (c.f. Busemeyer & Townsend, 1993; Laming, 1968; Link, 1992; Link & Heath, 1975; Ratcliff, 1978; Smith, 1995). Instead, the review is selective with the aim of illustrating the many similarities of neuromorphic and sequential sampling models. In some cases, versions of the neuromorphic models are special cases of the sequential sampling models, and in other situations they are more general.

Sequential sampling is the dominant dynamic model of decision making (see Luce, 1986 for a review). Dynamic decision models predict not only which choice is made, but also response time ( $RT$ ). A decision is made when the sum of a series of samples from a signal exceeds a criterion. Use of the sum promotes accurate decisions by averaging over sample-to-sample noise in the signal. As more samples are taken, noise is suppressed but  $RT$  is increased. Hence, a tradeoff between speed and accuracy (Luce, 1986, Section 6.5) is a fundamental prediction of sequential sampling models. When the signal to noise ratio is small, slow responding promotes accuracy. Under time pressure fast decisions can be made, but the decision will be inaccurate unless the signal to noise ratio is large.

Noise, in the form of random sample-to-sample fluctuation in the signal, may be introduced from a number of sources. Link (1992) suggests that sample-to-sample noise is due to Poisson fluctuation in the transmission of signals in the nervous system. At a coarser temporal scale, Busemeyer and Townsend (1993) suggest that subjects switch attention between different consequences of the choice on each sample. Decisions based on a sum of samples allow the subject to weigh a large range of alternatives without having to hold them in short-term memory simultaneously.

Stone (1960) and Laming (1968) proposed early versions of sequential sampling models. Subjects sample a signal with mean  $m$  at regular time intervals of length  $dt$ . Hence, each sample has mean size  $m$  and is added in exactly  $dt$  time units. A response is made when the sum of samples exceeds a criterion,  $C$ . In the following I will use the term “activation” to refer to the sum of samples.  $RT$  equals the time per sample multiplied by the number of steps taken until the sum first equals or exceeds  $C$ . The model is often called a *random walk* because the sum follows a jagged path reminiscent of a particle undergoing Brownian motion.

The random walk is a discrete process. As the time unit,  $dt$ , goes to zero it becomes equivalent to a continuous Wiener diffusion process. Summation becomes integration with infinitesimal mean step size or drift,  $m$ , and infinitesimal variance or diffusion rate,  $\sigma^2$ . The diffusion process, with normally distributed sample-to-sample noise, was first proposed as a model of simple  $RT$  by Emerson (1970).

When drift is normally distributed and  $C \gg m$ , mean  $RT = C/m$  ( $C/m \times dt$  for the discrete case), and overall  $RT$  has an Inverse Gaussian distribution. This single criterion case can model simple  $RT$ , where the subject responds contingent only on the detection of a change in the stimulus. Multiple criteria or barriers model more complex choice tasks. In the two choice case, for example, an adaptation level can be subtracted from the signal so that samples from one class of stimuli have negative mean drift and the other class has positive mean drift (Link & Heath, 1975). The choice made depends on which barrier is crossed first. The  $RT$  distribution for the two-alternative Wiener diffusion model is given in Ratcliff (1978). It is more complex than the distribution for

the single barrier model because the *RT* distributions for each choice can be a mixture of correct and incorrect responses.

### Varieties of Sequential Sampling Models

In Link and Heath's (1975) Relative Judgement Theory, a random walk is driven by the difference between the signal and an adaptation level or referent. Subjects can control the distance between response barriers, the starting position of the random walk, and magnitude of the referent. This theory predicts not only *RT*, but is also able to account for variations of response probability modeled by static theories such as Signal Detection Theory (SDT) and finite state theories (see Green & Swets, 1966).

Link and Heath (1975) demonstrated that their model can produce similar Receiver Operating Characteristics (ROCs) to SDT by changing the random walk's starting point. Differences in the magnitude of mean drift rates (and/or their variances) from the two stimulus classes result in asymmetric ROC curves, assumed to be due to unequal variance in SDT. Increasing either the distance between the barriers or the magnitude of the drift rate causes increasingly curved ROCs associated with an increase in sensitivity in SDT. A more recent version of this model, Wave Theory, proposed by Link (1992), is consistent with an impressive list of psychophysical phenomena, including the psychometric function, the interval of uncertainty, Fechner's, Weber's, and Steven's Laws, as well as ROC results.

Despite the success of both Wave Theory and models based on the Wiener diffusion process, without elaboration both are unable to model an important finding related to speed-accuracy tradeoff, the conditional mean *RT* effect (Heath, 1992). Under speed instructions, mean error *RT* tends to be faster than mean correct *RT*, whereas, under accuracy instructions, errors tend to be slower than correct responses. Link and Heath (1975) showed that, when the moment generating function of a random walk's step size is symmetric, correct and error mean *RT*s at the same barrier are equivalent. As the Poisson moment generating function is symmetric (as are the moment generating functions of most commonly used noise distributions), Wave Theory cannot

account for the conditional mean *RT* effect. The Wiener diffusion model for two choices also predicts equal error and correct mean *RT*.

Two accounts of the conditional mean *RT* effect have been proposed. One account adds variability between trials to model parameters, the starting point and drift rate. Laming (1968) suggested that the starting point of a random walk might vary from trial-to-trial under speed instructions due to sampling before the onset of the stimulus. Fast and inaccurate responses will occur when the starting point is near the barrier for the wrong choice. In contrast, slow errors are produced when between trial variability is introduced to the drift rate of a two alternative Wiener diffusion model (Ratcliff, 1978). If accuracy stress causes an increase in the inter-trial variability of drift rates, slow errors for accurate responding can be explained. Ratcliff (1988) notes that fast errors can also be obtained with drift distributions that have heavier tails than the normal distribution, although he did not pursue this approach.

The second account introduces systematic changes in the signal strength over the course of a trial. Heath's (1981) tandem random walk model, and his later generalization for temporally and spatially nonstationary inputs (Heath, 1992), can account for both aspects of the conditional mean *RT* effect. If a response is initiated when the input is increasing, mean error *RT* is faster than mean correct *RT*. If the response is made when the input is stationary, correct and error mean *RTs* are equal. If a response is made when the input is decaying, mean error *RT* is slower than mean correct *RT*. Heath (1992) and Smith (1995) discuss the physiologically interesting case where a nonstationary input is the result of a deterministic cascade process.

Another variety of sequential sampling model assumes that the process of sample integration, rather than the signal, changes over the course of a trial. For example, some signal strength might be lost in the summation or integration process. Unlike a random walk or diffusion, a "leaky integrator" is dissipative; it loses some of the signal strength in the process of integration. Smith (1995) and Busemeyer and Townsend (1993) describe sequential sampling models in which the leakage is dependent on the activation of the unit. Consequently, the effective drift from a constant input signal

decreases to zero as activation grows, and so activation achieves equilibrium. Busemeyer and Townsend (1993) modulate state dependent leakage rates to model the finding that, in both animal and human decision making, avoidance-avoidance decisions take longer than approach-approach decisions. Busemeyer and Townsend's (1993) Decision Field Theory modulates state dependent leakage rates to model the finding that, in both animal and human decision making, avoidance-avoidance decisions take longer than approach-approach decisions. Decision Field Theory also accounts for a range of phenomena in risky decision making and matching.

Activation dependent leakage can answer Heath's (1992) "finite  $RT$  dilemma": why can subjects not become arbitrarily accurate by sampling for a sufficient time? Accuracy for a lossless random walk model increases indefinitely as more samples are taken. Hence, increased accuracy can always be obtained by placing the decision barriers further away from the starting point. For a state dependent leaky integration model, in contrast, activation tends to an asymptote, so a decision criterion cannot be placed arbitrarily far from the asymptote.

Mechanisms do exist within lossless random walk models that can address the finite  $RT$  dilemma, for example, Ratcliff's (1978) addition of trial-to-trial as well as sample-to-sample noise. While continued sampling can arbitrarily reduce sample-to-sample noise, it can do nothing to reduce trial-to-trial noise. When sufficient samples have been taken to reduce sample-to-sample noise below trial-to-trial noise, continued sampling has little effect on the overall accuracy of the decision. Ratcliff and Van Zandt (1995) discuss the roles of both sample-to-sample and trial-to-trial noise in a diffusion model of a signal detection task. As Luce (1986) notes: "... the original wrinkle introduced [by Ratcliff (1978) which] ... has given the [diffusion] model a great deal of flexibility, is that rates are in fact random variables over otherwise identical trials. ... This added freedom is adequate to permit the model to mimic a great deal of somewhat surprising data." (p. 439).

Clearly, experimental data may contain both sample-to-sample and trial-to-trial noise. Sequential sampling models have tended to emphasise sample-to-sample noise,

perhaps because of its role in explaining speed-accuracy tradeoff. In the next section we will explore why, in neuromorphic models, sample-to-sample noise is not necessary for this purpose.

### Motivating Neuromorphic Models of Choice

The first order behaviour of a neuron is similar to a random walk process in that a neuron integrates inputs over time. However, neurons are dissipative structures, so their integration is leaky. In contrast to a random walk, where a constant input causes activation to increase indefinitely, activation in a neuron converges to a value that is a (usually monotonic) function of a constant input. When input is removed the neuron returns to baseline firing. Both behaviours are characteristic of a unit with activation dependent leakage. Secondly, the activation of a neuron is bounded; the unit's activation cannot be driven above a maximum value no matter how large the input because its sensitivity decreases as activation approaches the bound. In the following we will discuss the implications of introducing these constraints to sequential sampling models of response time.

#### State Dependent Leakage

One important reason for introducing leakage into neuromorphic models is physical plausibility: state dependent leaky integration can attain only finite activation for finite inputs, whereas the activation of a lossless integrator diverges to infinity for a constant finite input. With leakage, activation also takes some time to fully reflect the input, a phenomenon commonly measured by a time constant in neural systems. Hence, an activation dependent leaky integrator has bounded sensitivity, both in time and activation.

Speed-accuracy tradeoff can be explained by the finite temporal sensitivity of activation dependent leaky integrators. A fast response is effectively based on a weak signal if it is made before activation has reached equilibrium. Consequently, sample-to-sample noise is not necessary to explain speed-accuracy tradeoff. While a complete lack of sample-to-sample noise is unlikely, it is an interesting question in many

experimental paradigms whether performance is dominated by sample-to-sample or trial-to-trial noise. In a recognition memory experiment, for example, variations in study-test lag, and word stimuli, may cause considerable extrinsic noise in inputs over trials within an experimental condition. The neuromorphic models developed here assume that trial-to-trial noise dominates. This was done mostly in the interest of mathematical tractability. Having gained some results with the assumption of trial-to-trial noise alone, we will return to the issue of sample-to-sample noise in the final section.

Leakage also results in decay of activation back to baseline when the input ceases. This is functionally useful for an autonomous decision system because decision units automatically reset between decisions. In contrast, a random walk requires reset by an external controller. Partial return to baseline between closely spaced trials may account for some patterns of sequential dependency in both accuracy and *RT*, although learning effects in adaptation levels and strategic probability matching are also likely to be important (e.g. Link, 1992, pp. 276-293).

### Representation

Activation dependent leakage is useful because the decision unit's asymptotic activation provides an estimate of the magnitude of the input. Hence, the unit's activation can be used to represent stimulus characteristics as well as to make decisions. In a lossless model, the unit's activation contains no information about the magnitude of the input, unless integration time is also known. In an activation dependent dissipative model, in contrast, activation is a function of the input magnitude independent of integration time (once the unit has converged).

Convergent activation provides a point of contact between dynamic and static models of choice. Due to the static nature of SDT, *RT* predictions do not emerge naturally, although additional mechanisms, such as *RT* proportional to the difference between signal and criterion, have been suggested (Murdock & Dufty, 1972). A state dependent leaky integrator model, in contrast, predicts both *RT* and accuracy. The static decision variable in SDT is identified with the static activation value of the

neuromorphic unit after convergence. Choice can be modeled in the manner of SDT, by comparing the converged activation value to a criterion. *RT* predictions follow naturally from the time to detect convergence.

### Response Criteria

The preceding discussion suggests a more radical departure from previous sequential sampling models: using detection of convergence as the criterion for response. Gregson (1992) suggests that a check on the stationarity of the Jacobian for a system of units can be used in this way. For a single unit, a simple way to detect convergence is to place a criterion on the derivative of activation with respect to time. In the discrete case, the criterion is placed on the size of each sample. A criterion on the derivative is convenient for the neuromorphic models described below as they are defined by first order differential equations. The derivative is not so convenient when the neuromorphic model contains sample-to-sample noise, as activation, while being almost everywhere continuous, is almost nowhere differentiable.

In the following development, both activation and derivative criteria will be investigated. The two criteria are not incompatible. For example, an activation criterion may be used when a fast response is required and a derivative criterion when an accurate magnitude estimate is required. Both criteria can be problematic for units with activation dependent leakage. If the input is small, a derivative criterion may be satisfied immediately, whereas an activation criterion may never be satisfied. Hence, there may be situations when both criteria are applied simultaneously.

### Activation Bounds

Many neural network models bound a unit's output activation with sigmoid transformation. The presence of bounds can be motivated by physical realisation: like any physical system, the unit can only take on states within a finite range. The particular choice of bounding function, a smooth sigmoid, is an idealisation for individual neurons, which often exhibit sensitivity thresholds. However, the sensitivity of a collection of neurons, where thresholds vary randomly between neurons, is described by a smooth sigmoid function. For the decision units discussed so far, activation was

unbounded. State dependent leakage ensures that activation does not diverge with integration time, but the dynamic range of the unit remains unlimited. Bounds of some type seem called for on a priori grounds, but, as will be shown in the following development, they also impart functionally useful properties to the neuromorphic units.

A second motivation for sigmoid bounds on activation comes from the suggestion that neuromorphic units can represent the input. Backpropagation neural networks use sigmoid hidden units as they provide universal function approximation abilities (Hornik, Stinchcombe, & White, 1989). One way, therefore, to provide general representational ability, is to make asymptotic activation a sigmoid function of input magnitude.

### Summary

The preceding sections suggest that a neuromorphic model should have both activation dependent leakage and bounded activation. They also suggest that asymptotic activation be used to represent the magnitude of the input, so neuromorphic units have the dual role of representation and decision. In order to make the representation general, sigmoid bounds are useful. In order to make the representation independent of integration time, a response criterion based on the derivative of activation was suggested, although an activation criterion may also be used in some situations. Finally, emphasis was placed on trial-to-trial noise in the model's parameters. While the likely occurrence of sample-to-sample noise was acknowledged, it was argued that it could be dominated by trial-to-trial noise, so that models with trial-to-trial noise alone are interesting, at least as a limiting case.

### Deterministic Models

In this section we will examine the deterministic behaviour of models with a minimal structure motivated by the considerations in the previous section. In the next section, we will examine the behaviour of these models with trial-to-trial noise in their inputs. We will consider two types of neuromorphic units based on deterministic

models of first order neural dynamics used by Carpenter and Grossberg (1987) and Kohonen (1993).

Kohonen Units

Kohonen (1993) suggests that important aspects of dissipative neural dynamics are captured by the differential equation:

$$\dot{x} = I - g(x) \tag{1}$$

where  $x$  is activation,  $\dot{x}$  its derivative,  $I$  is an input, and  $g(x)$  a loss function describing activation dependent leakage.

For example, consider the simple case where the loss function is linear:  $g(x) = \beta + \alpha x$ . By definition, the system described by Equation 1 converges when  $\dot{x} = 0$ . By solving (1) with the latter condition we find that asymptotic or equilibrium activation equals  $(I - \beta)/\alpha$ . Hence, the final state of the unit is a linear transformation of the input. A solution for activation,  $x$ , at all times,  $t$ , is derived by solving (1)

$$x(t) = \frac{I - \beta}{\alpha} (1 - e^{-at}) + x(0)e^{-at} \tag{2}$$

The activation of the unit at  $t = 0$  is  $x(0)$ . Unless otherwise stated, we will assume in the following that  $x(0) = 0$ . Figure 1a illustrates the solution, with activation converging at an exponentially decreasing rate toward its asymptotic value.

-----  
 Insert Figure 1 about here  
 -----

Interestingly, the activation update function used by most PDP models,  $x_j = \sum w_{ij} x_i$  can be viewed as the equilibrium solution to Equation 1 with  $I = \sum w_{ij} x_i$ , and  $g(x) = x$ . Similarly, Cohen, Dunbar, and McClelland's (1990) cascade update equation used to predict  $RT$  in the Stroop paradigm,  $x_{t+1} = (1 - \tau)x_t + \tau \sum w_{ij} x_i$  is an Euler's approximation to the same equation with a time constant  $\tau$ :

$$\dot{x} \approx \frac{x_{t+1} - x_t}{\tau} = I - x_t.$$

In Cohen et al.'s model,  $RT$  predictions were derived using a combination of the cascade update rule and a random walk decision unit attached to the output of the

network. For a neuromorphic model, in contrast, the dynamics of the neural unit alone determine  $RT$ .

$RT$  predictions for the linear leakage Kohonen unit can be derived by determining the time,  $t$ , at which  $x = a$  (Equation 3), the activation criterion, and at which  $\dot{x} = d$  (Equation 4), the derivative criterion.

$$t = \frac{1}{a} \ln\left(\frac{I-b-ax(0)}{I-b-aa}\right) = \frac{1}{a} \ln\left(\frac{x(\infty)-x(0)}{x(\infty)-a}\right) \quad (3)$$

$$t = \frac{1}{a} \ln\left(\frac{I-b-ax(0)}{d}\right) = \frac{1}{a} \ln\left(\frac{x(\infty)-x(0)}{x(\infty)-d}\right) \quad (4)$$

The second form of both equations specifies the solution in terms of the asymptotic value of activation  $x(\infty) = (I-b)/\alpha$ . For both types of criteria,  $t$  decreases as the criteria become less strict. Signal intensity, however, has opposite effects on  $t$ , with more intense signals decreasing  $t$  for the activation criterion but increasing  $t$  for the derivative criteria. Empirical  $RT$ s from simple and choice paradigms are usually a decreasing function of signal intensity and an increasing function of response caution. If we identify signal intensity with  $I-\beta$  and response caution with  $1/a$  or  $1/d$ , it is clear that the derivative criterion model is in error for signal intensity predictions: response time is predicted to be an increasing rather than decreasing function of signal intensity.

In order to restrict activation within a given range for all input values, the loss function must be nonlinear. Kohonen (1993) bounds activation in the range  $[-1,1]$  using:

$$g(x) = c \ln\left(\frac{1+x}{1-x}\right)$$

At equilibrium:

$$x(\infty) = \frac{1 - e^{-I/c}}{1 + e^{-I/c}} = \tanh(I/c) \quad (5)$$

Similarly,  $g(x) = c \ln(x/(1+x))$  produces  $[0,1]$  bounded asymptotic activation following a logistic function. Unfortunately, we cannot obtain an explicit expression for  $t$  for Kohonen's sigmoid equations as they are not soluble in closed form<sup>1</sup>. Instead, we proceed by considering a second mechanism for bounding activation, *shunting inputs*.

### Grossberg Units

Grossberg and co-workers (e.g. Carpenter & Grossberg, 1987) often approximate neural dynamics with the differential equation:

$$\dot{x} = -\mathbf{a}x + I^+(A - x) - I^-(B + x) = (AI^+ - BI^-) - x(\mathbf{a} + I^- + I^+) \quad (6)$$

Inputs are separated into excitatory ( $I^+$ ) and inhibitory ( $I^-$ ) components as is commonly observed in neural systems (note that all parameters and inputs in Equation 6 are positive). The form of Equation 6 given on the right can be usefully contrasted with Equation 1. The main difference is that the leakage term is a function of the inputs as well as the level of activation. This dependence is called a shunting interaction. The form of Equation 6 on the left makes it clear why activation is bounded in the range  $[-B, A]$ . The effect of excitation is scaled by the difference between the present level of activation and  $A$ , and the effect of inhibition is scaled by the difference between the present level of activation and  $-B$ . Even when an external perturbation leads to  $x > A$  or  $x < -B$  the system will be attracted back within the bounds.

The equilibrium solution to Equation 6 also shows that activation is bounded in the range  $[-B, A]$ :

$$x(\infty) = \frac{AI^+ - BI^-}{\mathbf{a} + I^- + I^+} \quad (7)$$

To compare Equation 7 with Kohonen's *tanh* sigmoids, where  $I$  can be positive and negative, Figure 1b illustrates the case where  $I^+ = I$ ,  $I^- = 0$  for  $I \geq 0$  and  $I^- = -I$ ,  $I^+ = 0$  for  $I < 0$  (i.e.  $x(\infty) = I/(\mathbf{a} + |I|)$ ). The slope of the shunting sigmoid is controlled by the rate constant  $\alpha$ . For comparison, three *tanh* sigmoids are also shown. Note that the shunting sigmoid (thick line) tends to have a shorter linear range and is steeper around the origin than the *tanh* sigmoids which cover approximately the same dynamic range.

Figure 1c illustrates the convergence behaviour of Equation 7. Unlike the linear loss form of Equation 1 (see Figure 1a), time to convergence (approximately where the graphs become flat in Figure 1a and 1c), and hence  $RT$  under a derivative criterion, is inversely proportional to input intensity.

To obtain an explicit expression for activation as a function of time, we begin by solving Equation 6:

$$x(t) = \frac{AI^+ - BI^-}{\mathbf{a} + I^- + I^+} \left( 1 - e^{-(\mathbf{a} + I^- + I^+)t} \right) + x(0)e^{-(\mathbf{a} + I^- + I^+)t} \quad (8)$$

As with the linear version of Kohonen's equation, activation converges at an exponentially decreasing rate toward its asymptotic value. Solving for  $x = a$ , the activation criterion, and  $\dot{x} = d$ , the derivative criterion:

$$t = \frac{1}{\mathbf{a} + I^+ + I^-} \ln \left( \frac{x(\infty) - x(0)}{x(\infty) - a} \right) \quad (9)$$

$$t = \frac{1}{\mathbf{a} + I^+ + I^-} \ln \left( \frac{x(\infty) - x(0)}{d} \right) \quad (10)$$

Equations 9 and 10 are similar to the solutions for the linear activation dependent leakage model (Equations 3 and 4). This occurs because they are both solutions to the general equation:  $x(t) = x(\infty) - (x(\infty) - x(0))e^{-Kt}$ , where  $K$  determines the rate of leakage.

As in the linear model, a stricter activation or derivative criterion causes a decrease in  $t$ . The models differ in that the scaling factor in the linear model ( $1/\alpha$ ) becomes dependent on the inputs ( $1/(\alpha + I^+ + I^-)$ ), due to the shunting interaction. For the activation criterion, the shunting interaction further enhances the decrease in  $t$  with increases in signal intensity.

The effect of increasing signal intensity on  $t$ , given the derivative criterion, depends on a balance between a hyperbolically decreasing tendency caused by the scaling factor and logarithmic increase identical to the linear model. Figure 1d illustrates the tradeoff, plotting  $t$  as a function of the derivative criterion,  $d$ , and input intensity,  $I^+$  (with  $I^- = 0$ ). Note that Equation 10 is valid only when  $d < (x(\infty) - x(0))$  as the difference between  $x$  at 0 and  $\infty$  is the maximum rate of change of activation. Hence, the negative solutions in the figure are not valid. For larger input values,  $t$  is a negatively accelerated, decreasing function of  $I^+$ , consistent with empirical evidence of a decrease in  $RT$  with increasing signal intensity. For smaller<sup>2</sup> values of  $I^+$ , however,

$t$  increases with signal intensity. The increasing region reflects approximately linear sensitivity and associated behaviour approximated by Equation 4.

The influence of the increasing region depends on the value of  $d$ . Considering Equation 10 with parameters specified as in Figure 1d, for example,

- i)  $d = 0.01$  (lax criterion, middle curve in Figure 1d):  $t$  at the point of inflection is 11.57, activation at response is 0.41, and activation at asymptote is 0.46,
- ii)  $d = 0.0001$  (strict criterion, upper curve in Figure 1d):  $t$  at the point of inflection is 44.2, activation at response is 0.1835, and activation at asymptote is 0.184.

For the strict criterion, the increasing region occurs for  $I < 0.023$  and in less than 20% of the domain of  $x$  ( $[0,1]$  for the parameters chosen). With the lax criterion, the increasing region occurs for  $I < 0.09$  and covers almost 50% of the domain of  $x$ . The difference between activation at response ( $\dot{x} = d$ ) and asymptote ( $\dot{x} = 0$ ) also changes with the criterion. For the strict criterion, the difference is only 0.3%, compared to almost 11% for the lax criterion.

Overall, the similarity between asymptotic sensitivities shown in Figure 1b suggests that Kohonen's and Grossberg's models will act in a similar manner. We will now turn our attention to the behaviour of the model when the inputs have a stochastic component. It turns out that in this context we can make some progress with Kohonen's bounded model as well as with Grossberg's shunting model.

#### Random Models

We will now explore the effect of trial-to-trial noise on the inputs to neuromorphic units. The distribution of the input,  $I$ , is assumed to be gaussian, a plausible assumption in a neural network context as  $I$  is the sum of a large number of weighted inputs. We will characterise the model by determining the density function and hazard function for response time.  $RT$  density is usually uni-modal and positively skewed, and increases in the mean are usually accompanied by increases in variance and skew (e. g. numerous examples in Luce, 1986, although distributions for very fast responses can be relatively

symmetric, see, for example, Luce, 1986, p.117 for simple  $RT$  to intense auditory stimuli). Examination of density functions will determine if the neuromorphic models share any of these characteristics. The hazard function is defined as probability density divided by the one minus cumulative density, and denotes the chance of an event occurring at time  $t$  given it has not already occurred. Luce (1986) suggests that hazard functions can help to differentiate models with similar density functions. For simple  $RT$ , empirical hazard functions rise to a peak and then descend, especially for intense stimuli. The neuromorphic model hazard functions will be compared to this pattern.

We will begin by obtaining analytic results for the lossless integrator with an activation criterion. Only the distribution of  $t$  is of interest as  $x$  is, by definition, a constant equal to the criterion. Analytic results were also obtained for asymptotic activation in the Kohonen and Grossberg models. However, the distribution of  $t$  for these models must be investigated by Monte Carlo techniques.

#### Lossless Integrator: Analytic Results for $t$

Before examining the dissipative models, we will explore the lossless integrator with normally distributed input,  $I$ , and a one-barrier activation criterion,  $B$ . For this model,  $t = B/I$  is inversely proportional to input intensity and proportional to response caution ( $B$ ). Distribution functions with only trial-to-trial noise are easily obtained because the relationship between barrier crossing time,  $t$ , and the random variable,  $I$  is monotonic and has a differentiable inverse (see DeGroot, 1986, p.152-154). Given that the density of  $I$ ,  $f(I)$ , is normal with mean  $\mu$  and standard deviation  $\sigma$ , the density of  $t$ ,  $g(t)$ , is:

$$g(t) = \frac{B}{\sqrt{2\pi}\sigma t^2} e^{-\frac{1}{2}\left(\frac{B-\mu}{\sigma t}\right)^2} \quad (11)$$

Figure 2 illustrates the effect of varying  $\mu$  and  $\sigma$  relative to a reference distribution with  $\mu = 10$ ,  $\sigma = 3$  ( $B = 3000$  in all cases). Note that the reference distribution is the curve with the second highest peak in all panels of Figure 2. A decrease in  $\mu$  causes an increase in the mode, mean, variance, and skewness and a decrease in the order and peakedness of the hazard function. An increase in  $\sigma$  causes a decrease in the mode, and

increase in mean variance, and skewness. The hazard function spreads at the same location with increases in  $\sigma$ , as does the density function.

-----  
 Insert Figure 2 about here  
 -----

Kohonen and Grossberg Units: Analytic Activation Distributions

Closed form results for the distribution of asymptotic activation in both Kohonen and Grossberg units can be obtained because the *tanh*, logistic, and shunting sigmoid functions are monotonic and have a differentiable inverse. For Kohonen units with a derivative ( $\dot{x} = d$ ) criterion, the activation distribution function is the same as for the asymptotic case, with the exception that the mean,  $\mu$ , is reduced by  $d$ . Hence, we will develop results only for the asymptotic case for Kohonen units.

For a Kohonen unit with *tanh* leakage, and given that the density of  $I$ ,  $f(I)$ , is normal with mean  $\mu$  and standard deviation  $\sigma$ , the density of  $x$ ,  $g(x)$ , is:

$$g(x) = \frac{c\sqrt{2}}{\sqrt{ps}(1-x^2)} \left( \frac{1+x}{1-x} \right)^{\frac{m}{s^2}} e^{-\frac{1}{2s^2} \left( m^2 + c^2 \left( \ln \left( \frac{1+x}{1-x} \right) \right)^2 \right)} \quad (12)$$

Figure 3 illustrates the properties of  $g(x)$  as a function of the mean and standard deviation of the input for  $c = 1$ . When  $\mu = 0$  and  $\sigma$  is small, the distribution is approximately normal, but as  $\sigma$  increases the distribution flattens and eventually becomes U shaped as probability mass from the tails builds up at either bound (see Figure 3a). When  $\mu$  is non-zero, the activation distribution becomes skewed, with the degree of skew strongly dependent on  $\sigma$  (see Figures 3b and 3c). A similar result holds for the logistic bound as the logistic function is a linear transformation of the *tanh* function.

-----  
 Insert Figure 3 about here  
 -----

The effect on of varying  $d$  can be seen in Figure 3c, because a stricter convergence criterion is equivalent to increased mean signal strength. If accuracy equals the

proportion of the activation distribution passing a criterion in  $x$  (e.g. below the origin in Figure 3c), then accuracy is inversely proportional to  $d$ .

The separation of excitatory (positive) and inhibitory (negative) inputs in Grossberg's shunting equation makes it difficult to determine the distribution of activation across the entire range  $[B,A]$ . We will instead consider the case where there is negligible probability of  $I \leq 0$ , and solve for the derivative criterion,  $\dot{x} = d$ .

Note that variation in the criterion,  $d$ , for a Grossberg unit is not equivalent to variation in the mean signal strength as it was for a Kohonen unit, due to the shunting interaction. Hence, we must derive an expression for the distribution of activation which includes  $d$ . Given that the density of the input signal  $I$ ,  $f(I)$ , is normal with mean  $\mu$  and standard deviation  $\sigma$ , the density of  $x$ ,  $g(x)$ , is:

$$g(x) = f(s(x)) \frac{ds(x)}{dx} = \frac{\mathbf{a} + d}{\sqrt{2\mathbf{p}\mathbf{s}}(1-x)^2} e^{-\frac{1}{2\mathbf{s}^2} \left( \frac{x(\mathbf{a}+\mathbf{m}) - (\mathbf{m}+d)}{1-x} \right)^2} \quad (13)$$

Figure 4 illustrates that the behaviour of asymptotic Grossberg unit activation is similar to that of asymptotic Kohonen unit activation shown in Figure 3, and that  $\mu$  and  $d$  act similarly (note that  $d$  is varied over a smaller range than  $\mu$ ). Once again, if accuracy equals the proportion of the activation distribution passing a criterion in  $x$  (e.g. above  $x = 0.5$  in Figure 4b), then accuracy is inversely proportional to  $d$ .

-----  
 Insert Figure 4 about here  
 -----

In summary, we have found closed form solutions for activation at a derivative criterion,  $d$ , for both Kohonen and Grossberg units. We have also shown that variation in  $d$  can be used to model speed-accuracy tradeoff as a decrease in  $d$  produces more accurate responding. The deterministic results for time to satisfy the derivative criterion indicate that the increase in accuracy will be accompanied by a decrease in  $RT$ , as found empirically.

### Kohonen and Grossberg Units: Monte Carlo $RT$ Distributions

In order to obtain results for  $t$  in Kohonen and Grossberg units we must resort to Monte Carlo techniques. For Grossberg units, the distribution of  $t$  can be sampled by calculating Equation 9 or 10 with  $I$  drawn from an appropriate random variable. For Kohonen units,  $t$  must be evaluated by numerical integration. We used Euler's method, which approximates the integral in the same way a random walk approximates a diffusion process, in small linear steps. Each step was set at  $\dot{x}/10$ , so the effective value of  $d$  in the following is an order of magnitude larger the actual value of  $d$  given. Note that noise was not added on each step of the solution process. Rather, inputs were constant within a trial but varied randomly between Monte Carlo replications. We performed simulations using only the derivative criterion.

In order to compare Grossberg and Kohonen units, we choose the parameters of their sigmoid bounds so they had a similar shape ( $\alpha = 0.1$  and  $c = 0.2$  respectively) and were sensitive to inputs in the range  $[-1,1]$ . Inputs varied normally with  $\sigma = 0.1$ , so that activation remains uni-modal, and  $\mu$  was chosen so that the input was usually greater than 0, avoiding the fast convergence region on the left of Figure 1d.

Figure 5a shows density estimates for the Kohonen unit. One thousand solutions for each of six values of  $\mu$  were calculated, and density estimated by adaptive time domain filtering with an Epanechnikov kernel. The distribution of  $t$  is positively skewed and uni-modal as desired for larger values of  $\mu$ . For smaller values, it become symmetric then negatively skewed. For the smallest value of  $\mu$  (0.2), the distribution becomes U shaped with  $t = 0$  for some solutions.

Figure 5b illustrates the results for Grossberg units. Unlike the Kohonen units,  $t$  is always positively skewed, with skew decreasing as mean input increases. Note that on the left of Figure 5b,  $t = 0$  solutions occur for  $\mu = 0.2$ .

-----  
Insert Figure 5 about here  
-----

Hazard functions were obtained for the simulated Grossberg  $t$  values by integrating densities smoothed by the Epanechnikov kernel (hazard functions were not determined for Kohonen unit due to computational cost). The estimated hazard functions are shown in Figure 5c. A clear peak occurs and is most evident for stronger inputs. Hence the hazard functions from both the lossless integrator and Grossberg units with a derivative criterion conform to the pattern seen in simple  $RT$  data, rising to a peak and then descending, especially for strong inputs.

The results of the Monte Carlo studies favor the Grossberg over Kohonen units as models of  $RT$ . The shallow sigmoid slope of Kohonen units for small inputs produces negatively skewed response time distributions that are not observed empirically. For Grossberg units, mean, variance, and skew decreases with increasing input intensity, a common empirical finding, and hazard functions are appropriately shaped.

#### Discussion

The work that reported here represents only a preliminary examination of neuromorphic models. Higher order differential equations can be used to provide greater fidelity to the neural substrate, although the first order characterizations used here may be adequate, especially when the “unit” represents the average behaviour of a tightly coupled system of neurons. The neuromorphic models also need to be extended to cover decisions with two or more choices. Such an extension might use one unit to represent each choice, with units linked in a winner-takes-all network to ensure that only one response is produced. While closed form solutions for such a model are unlikely, the extra complexity may be useful in modeling behaviour. Heathcote (1994) discusses the activation dynamics of such a unit and uses them to model signal detection results in recognition memory paradigms (see also Grossberg & Gutowski, 1987).

The neuromorphic models can also be extended to model response confidence by the asymptotic value of activation. Categorical response confidence can be modeled by

the same mechanism as SDT: placing multiple criteria on asymptotic activation (see Link and Heath, 1975; Link, 1992; Vickers 1979 for other approaches to response confidence in sequential sampling models). Clearly, nonlinear bounds will modify the usual SDT assumption of gaussian noise in the decision (activation) variable. However, such nonlinear effects may be useful in modeling cognition. For example, Heathcote (1993, 1994) discusses a case where bounded activation explains puzzling SDT results from recognition memory paradigms.

An important innovation in the present work is the use of a derivative criterion for response initiation. The intuitive motivation for a derivative criterion is that, in tasks emphasizing accuracy, there appears to be a natural scale for  $RT$ . It is possible that subjects respond when they feel that continued sampling will not result in accrual of further information. The latter rule approximately corresponds to a criterion on the derivative. However, mathematical analysis revealed an important shortcoming of the derivative criterion: weak signals may result in convergence after very few time steps (see Figure 1d). Clearly this is undesirable model behaviour, unless it is associated with paradigms where fast guesses occur (Ollman, 1966). Competitive dynamics, and perhaps a derivative criterion on the activation of the entire winner-takes-all network, may be necessary to remove this shortcoming. Another solution might be to not allow a decision before a lower time limit.

One important issue is the way in which noise enters the decision. As modeled here, random variation in the input occurs over trials rather than being added to the change of activation on each step. As one reviewer noted, why is there a need for repeated samples at all if there is no sample-to-sample noise? It seems likely that given the noisy nature of neurons, and fluctuations in attention and input selection within a trial, that sample-to-sample noise is necessary (however see Dzhafarov, 1992, for another approach to simple  $RT$  using only trial-to-trial variability). Sample-to-sample noise will force a change in the derivative criterion because a single sample criterion is too vulnerable to the effects of sample-to-sample noise. The criterion may have to be

placed on a moving average of activation changes, or the change may have to be consistently less than a criterion for several time steps.

### References

- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, *100*, 432-459.
- Cohen, J. D., Dunbar, K., & McClelland, J. M. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. Psychological Review, *97*, 332-361
- Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing, *37*, 54-115.
- DeGroot, M. H. (1986). Probability and Statistics, 2nd Edn., Addison-Wesley, Reading.
- Dzhafarov, E. N. (1992). The structure of simple reaction time to step-function signals. Journal of Mathematical Psychology, *36*, 235-268.
- Emerson, P. L. (1970). Simple reaction time with markovian evolution of gaussian discriminial processes. Psychometrika, *35*, 99-109.
- Green, D. M., & Swets, J. A. (1966). Signal Detection Theory and Psychophysics, New York: Wiley.
- Gregson, R. A. M. (1992). n-Dimensional Nonlinear Psychophysics. Earlbaum Associates, Hillsdale, NJ.
- Grossberg, S., & Gutowski, W. E. (1987). Neural dynamics of decision makin under risk: Affective balance and cognitive-emotional interactions. Psychological Review, *94*, 300-318.
- Heath, R. A. (1981). A tandem random walk model for psychological discrimination. British Journal of Mathematical and Statistical Psychology, *34*, 93-99.
- Heath, R. A. (1982). A model for signal detection based on an adaptive filter. Biological Cybernetics, *45*, 95-100.
- Heath, R. A. (1992). A general nonstationary diffusion model for two-choice decision-making. Mathematical Social Sciences, *23*, 283-309.
- Heathcote, A. (1993). Episodic ART: A Model of Episodic Recognition Memory, In A. C. Tsoi & T. Downs (eds.) Proceedings of the Fifth Australian Conference on Neural Networks. University of Queensland Electrical and Computer Engineering.
- Heathcote, A. (1994) Episodic ART: A model of episodic recognition memory. In A. C. Tsoi & T. Downs (Eds.) Proceedings of the World Congress on Neural Networks, *2*, 194-197.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, *2*, 359-366.
- Kohonen, T. (1993). Physiological interpretation of the Self-Organizing Map algorithm, Neural Networks, *6*, 895-906.
- Laming, D. R. J. (1968). Information theor of chice-reaction times. Academic Press, London.
- Link, S. W. (1992). The Wave Theory of Difference and Similarity, Lawrence Earlbaum Associates, Hillsdale, New Jersey.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination, Psychometrika, *40*, 70-105.
- Luce, R. D. (1986). Response Times, Oxford University Press, New York.
- Murdock, B. B., & Dufty, P. O. (1972). Strength theory of recognition memory. Journal of Experimental Psychology, *94*, 284-290.

- Ollman, R. T. (1966). Fast guesses in choice-reaction time. Psychonomic Science, 6, 155-156.
- Ratcliff, R. (1978). A theory of memory retrieval, Psychological Review, 85, 59-108.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. Psychological Review, 95, 238-255.
- Ratcliff, R., & Van Zandt, T. (1995). The diffusion model and signal detection, Paper presented at the 36th Annual Meeting of The Psychonomic Society, Los Angeles, California.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. Psychological Review, 102, 567-593.
- Stone, M. (1960). Models for choice-reaction time. Psychometrika, 25, 251-260.
- Vickcler, D. (1979). Decision Processes in Visual Perception. Academic Press, New York.

## Footnotes

<sup>1</sup>Some progress can be made using a first order Taylor series approximation to the leakage function,  $-\ln\left(\frac{1+x}{1-x}\right) \approx 2\left(x + x^3/3\right)$  for  $-1 < x < 1$ , which gives a differential equation in Bernoulli's form. However, the approximation is inaccurate near the bounds.

<sup>2</sup>The point of inflection dividing increasing and decreasing relationships can be determined from the roots of the derivative of Equation 10. In the simple case where inhibitory inputs are zero,  $x(0) = 0$ , and  $A = 1$ , the equation to be solved is  $1 + \frac{a}{I} + \ln\left(\frac{k}{I}\right) = 0$ . No closed form solution exists, but, for the five functions depicted in Figure 1d, numerical methods give the points of inflection as  $I = 0.0226, 0.0379, 0.086, 0.216,$  and  $0.359$  for  $k = 0.0001, 0.001, 0.01, 0.05,$  and  $0.1$  respectively.

Author Notes

Thanks to Richard Heath, Doug Mewhort, Mike Humphreys, and Jerry Busemeyer, Scott Brown, and Stephan Lewendowsky for comment on drafts of this paper.

## Figure Captions

Figure 1. (a) The growth in activation ( $x$ ) as a function of time ( $t$ ) for Equation 1 with linear leakage,  $\alpha=0.1$ ,  $\beta=0$ ,  $x(0)=0$ , and a range of inputs ( $I$ ). Note the convergence to asymptotic values ( $I/\alpha = 1, 0.5$ , and  $0.1$ ). (b) Asymptotic activation as a function of input intensity for a shunting sigmoid (thick line  $\alpha=0.1, A=B=1$ ) and  $\tanh$  sigmoids,  $c = 0.08, 0.2$ , and  $0.333$  (from steeper to flatter curves respectively). (c) Activation growth over time with positive inputs,  $I$ , for Equation 7,  $\alpha=0.01, A=B=1, I=0$ . (d)  $t$  (plotted as the ordinate) from Equation 10 as a function of  $I^+$  (the abscissa),  $\alpha=0.1, x(0)=0, A=B=1$ , and  $I=0$ , for five values of  $d$ .

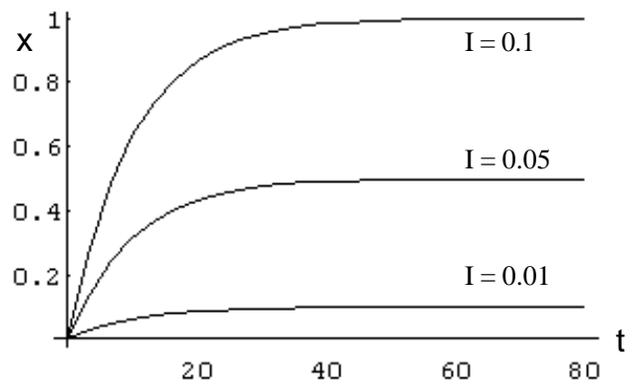
Figure 2. (a, b) Probability density functions and (c, d) Hazard functions (with support on  $t = 200-700$ , except  $t = 150-700$  for  $\mu = 10, \sigma = 5$  in (d)) for normal input to a deterministic single barrier random walk,  $B = 3000$ , (a, c)  $\sigma = 3, \mu = 12, 10, 8$  and (b, d)  $\mu = 10, \sigma = 2, 3, 5$  (iterating over curves from upper to lower).

Figure 3. Asymptotic activation distributions for a Kohonen unit with  $c = 1, x = \tanh(N(\mu, \sigma^2))$  and  $\sigma = 0.5, 1, 1.5, 2, 3$ , for (a)  $\mu = 0$ , and (b)  $\mu = -1$ , and (c)  $\sigma = 0.5$  for  $\mu = 0, -0.5, -1, -1.5$ .

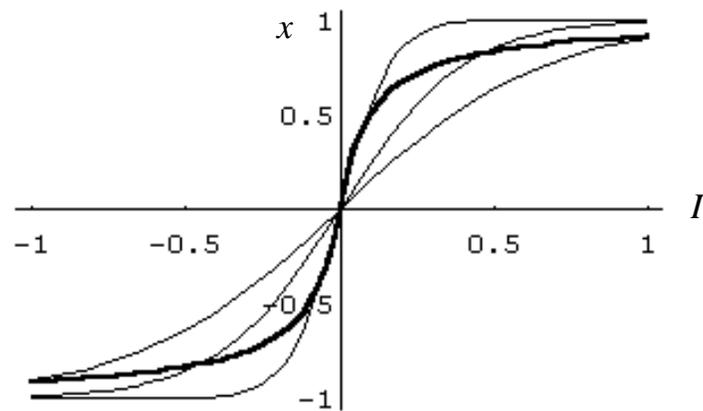
Figure 4. Asymptotic activation distributions for a Grossberg unit with  $x = ((I(\mu, \sigma^2) - k)/(\alpha + I(\mu, \sigma^2)))$ ,  $\alpha = 0.1, \sigma = 0.1$ , and (a)  $k = 0, \mu = 0.1, 0.2, 0.3, 0.4$  (curves from left to right) and (b)  $\mu = 0.3, k = 0.05, 0.025, 0.01, 0.0$  (curves from left to right).

Figure 5. (a) Probability density of  $t$  for Kohonen units with  $\tanh$  leakage ( $c=0.2$ ) and a derivative criterion ( $d = 0.0001$ ). Inputs were normal with,  $\sigma = 0.1, \mu = 0.2, 0.3, 0.4, 0.5, 0.6$  and  $0.7$  (for curves from right to left in the figure). (b) Probability density of  $t$  for Grossberg units with shunting bounds ( $\alpha=0.1$ ) and a derivative criterion,  $d = 0.0001$ . Inputs were normal with,  $\sigma = 0.1, \mu = 0.5, 0.4, 0.3$ , and  $0.2$  (for curves from left to right in the figure). (c) Grossberg unit hazard functions with  $\mu = 0.5, 0.4, 0.3$ , and  $0.2$  (for peaks from left to right respectively).

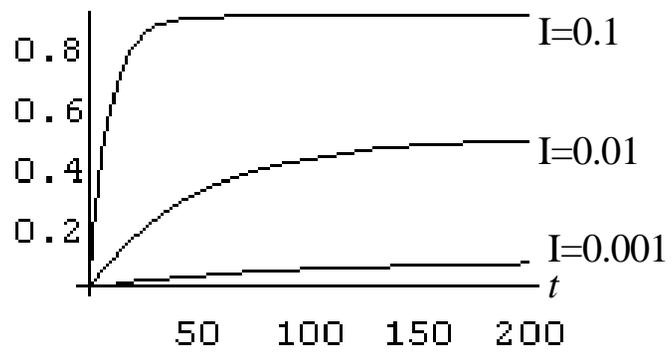
Figure 1



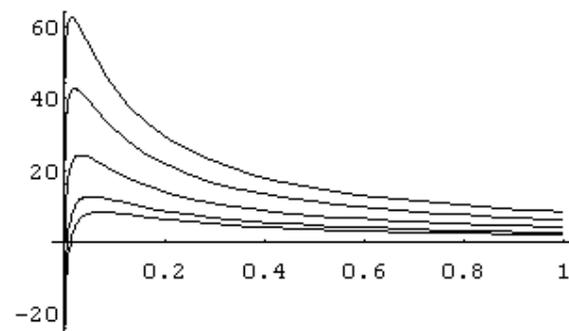
(a)



(b)

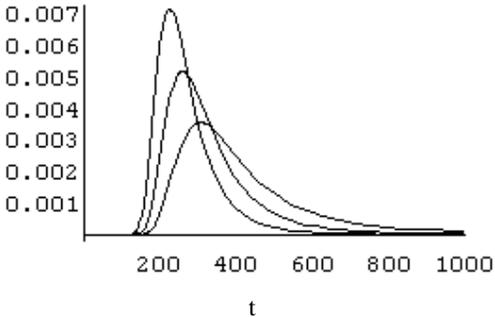


(c)

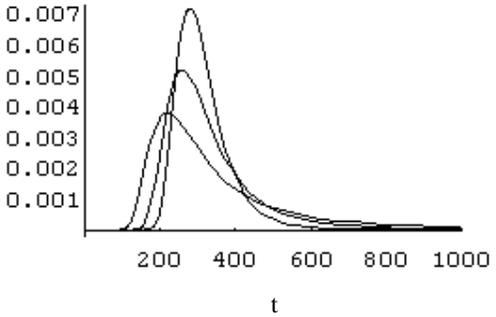


(d)

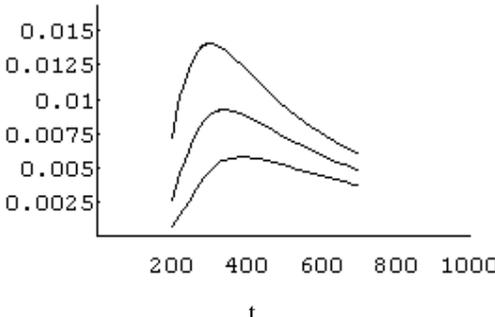
Figure 2



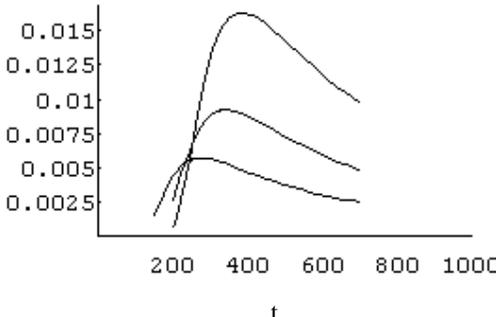
(a)



(b)



(c)



(d)

Figure 3

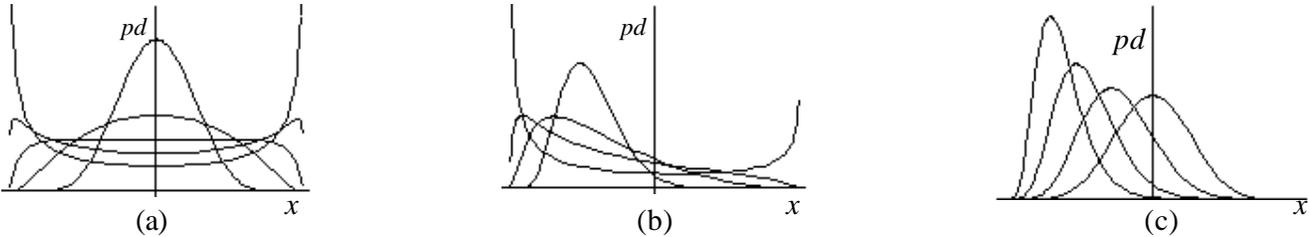
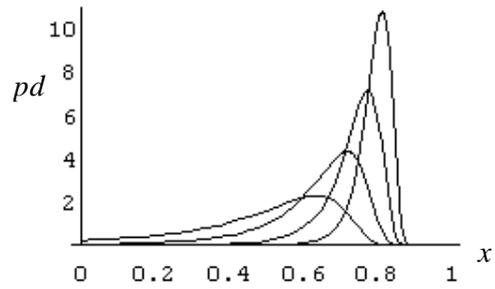
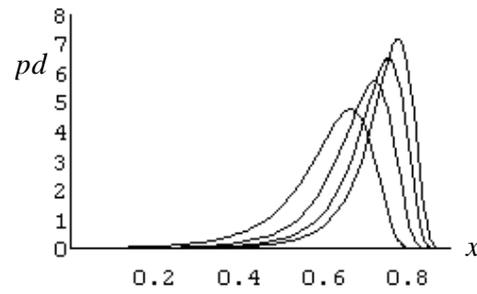


Figure 4



(a)



(b)

Figure 5

