LETTER Communicated by Peter Latham

Population Coding with Correlation and an Unfaithful Model

Si Wu*
Hiroyuki Nakahara
Shun-ichi Amari
RIKEN Brain Science Institute, Hirosawa 2-1, Wako-shi, Saitama, Japan

This study investigates a population decoding paradigm in which the maximum likelihood inference is based on an unfaithful decoding model (UMLI). This is usually the case for neural population decoding because the encoding process of the brain is not exactly known or because a simplified decoding model is preferred for saving computational cost. We consider an unfaithful decoding model that neglects the pair-wise correlation between neuronal activities and prove that UMLI is asymptotically efficient when the neuronal correlation is uniform or of limited range. The performance of UMLI is compared with that of the maximum likelihood inference based on the faithful model and that of the center-of-mass decoding method. It turns out that UMLI has advantages of decreasing the computational complexity remarkably and maintaining high-level decoding accuracy. Moreover, it can be implemented by a biologically feasible recurrent network (Pouget, Zhang, Deneve, & Latham, 1998). The effect of correlation on the decoding accuracy is also discussed.

1 Introduction

Population coding, a method to encode and decode stimuli in a distributed way by using the joint activities of a number of neurons, has the advantage of suppressing the fluctuation in a single neuron's signal (Georgopoulos, Schwartz, & Kettner, 1986; Paradiso, 1988; Seung & Sompolinsky, 1993). Recently there has been an expanded interest in understanding the population decoding methods, which particularly include the maximum likelihood inference (MLI), the center of mass (COM), the complex estimator (CE), and the optimal linear estimator (OLE) (see Pouget, Zhang, Deneve, & Latham, 1998; Salinas & Abbott, 1994). Among them, MLI has an advantage of having small decoding error (asymptotic efficiency) but may suffer from the expense of computational complexity, especially when the decoding model is complex.

Let us consider a population of N neurons coding a variable x. The encoding process of the population code is described by a conditional prob-

* Present address: Dept. of Computer Science, Sheffield University, U. K.

ability $q(\mathbf{r}|x)$ (Anderson, 1994; Zemel, Dayan, & Pouget, 1998), where the components of the vector $\mathbf{r} = \{r_i\}$ for $i = 1, \ldots, N$ are the firing rates of neurons. We study the following MLI estimator given by the value of x that maximizes the log-likelihood $\ln p(\mathbf{r}|x)$, where $p(\mathbf{r}|x)$ is the decoding model, which might be different from the encoding model, $q(\mathbf{r}|x)$. Thus far, when people have studied MLI in a population code, it normally (or implicitly) assumes that $p(\mathbf{r}|x)$ is equal to the encoding model $q(\mathbf{r}|x)$. This requires that the estimator has full knowledge of the encoding process. Taking account of the complexity of the information process in the brain, it is more natural to assume $p(\mathbf{r}|x) \neq q(\mathbf{r}|x)$. Another reason for choosing this is to save computational cost. Therefore, a decoding paradigm in which the assumed decoding model is different from the encoding one needs to be studied. In the context of statistical theory, this is called estimation based on an unfaithful or a misspecified model. This study is an extension of Wu, Nakahara, Murata, and Amari (2000). Hereafter, we refer to the decoding paradigm of using MLI based on an unfaithful model as UMLI, to distinguish it from that of MLI based on the faithful model, which is called FMLI.

Cross-correlation in neuronal activities is observed in both primary sensory and motor areas (Fetz, Yoyama, & Smith, 1991; Gawne & Richmond, 1993; Zohary, Shadlen, & Newsome, 1994; Lee, Port, Kruse, & Georgopoulos, 1998), where population coding is believed to be used. The neural correlation is often complex and may have various sources. For example, it can be inherited from the stimulus noise, generated by the multilayer structure of the information transformation process, or due to the feedback control procedure. It is very hard, even if it is possible, for an estimator to store and utilize all this information. This leads us to investigate an unfaithful decoding model without using the information of correlation. This is different from that of assuming an uncorrelated encoding model. In fact, it has been shown that assuming an uncorrelated encoding model for the cortex will lead to pathological conclusions (Pouget, Deneve, Ducom, & Latham, 1999).

By treating neural signals as if uncorrelated in decoding, UMLI decreases the computational cost of FMLI remarkably. More attractive, it maintains high-level decoding accuracy, which is better than that of COM and comparable to that of FMLI in many cases. Thus, UMLI can be a good compromise between decoding accuracy and computational complexity.

Recently, Pouget et al. (1998) and Deneve, Latham, and Pouget (1999) proposed a biologically feasible recurrent network model (RNM), which performs decoding in a coarse code format. We note that for the correlation structures in this study, the decoding of UMLI can be achieved by RNM.

An important issue discussed in this article concerns the asymptotic efficiency of MLI. It is known that when neuronal activities are uncorrelated, MLI is asymptotically efficient in the sense that it can asymptotically achieve the optimal decoding accuracy, that is, the Cramér-Rao bound (Paradiso, 1988; Seung & Sompolinsky, 1993). For correlated neuronal signals, MLI

may not be asymptotically efficient, depending on the correlation structure. We study this problem and prove the asymptotic efficiency of FMLI and the quasi-asymptotic efficiency of UMLI for the uniform and limited-range correlations.

As a by-product of the calculation, we also illustrate the effect of correlation on the decoding accuracy of a population code. It shows that the correlation, depending on its form, can either improve or degrade the decoding accuracy, which agrees with the general belief (see Snippe & Koenderink, 1992; Shadlen & Newsome, 1994; Yoon & Sompolinsky, 1999; Abbott & Dayan, 1999).

In section 2.1, an unfaithful decoding model is introduced, which neglects the pair-wise correlation between neuronal activities. Two different kinds of correlation structures are considered. In section 2.2, we calculate the decoding error of UMLI and FMLI and investigate their asymptotic efficiency for the models we consider. In section 2.3, the implementation of UMLI by RNM is discussed. In section 3, the performances of UMLI, FMLI, and COM are compared for two tuning functions. The effect of correlation on the decoding accuracy is also discussed. In section 4, some overall conclusions and discussions are given.

2 The Population Decoding Paradigm of UMLI

2.1 An Unfaithful Decoding Model of Neglecting the Neuronal Correlation. Let us consider a pair-wise correlated neural response model in which the neuron activities are assumed to be multivariate gaussian,

$$q(\mathbf{r}|x) = \frac{1}{\sqrt{(2\pi\sigma^2)^N \det(\mathbf{A})}} \times \exp\left[-\frac{1}{2\sigma^2}\sum_{i,j} A_{ij}^{-1}(r_i - f_i(x))(r_j - f_j(x))\right], \tag{2.1}$$

where $f_i(x)$ is the mean value of the response of the ith neuron representing its tuning function, that is,

$$\langle r_i \rangle = f_i(x), \tag{2.2}$$

$\langle\cdot\rangle$ denotes averaging over many trials, and N is the number of neurons.

The tuning function $f_i(x)$ usually has a radial symmetric shape,

$$f_i(x) = \phi[(x - c_i)^2], \tag{2.3}$$

where c_i is the preferred stimulus of the ith neuron. For example, the function $\phi(x)$ can be of the gaussian function form.

The matrix $\mathbf{A} = \{A_{ij}\}$ is the covariance matrix, which is defined by

$$\langle (r_i - f_i(x))(r_j - f_j(x)) \rangle = \sigma^2 A_{ij}, \tag{2.4}$$

and $\mathbf{A}^{-1}$ is its inverse.

Two correlation structures are considered here. One is the uniform correlation model (Johnson, 1980; Abbott & Dayan, 1999) with the covariance matrix

$$A_{ij} = \delta_{ij} + c(1 - \delta_{ij}), \tag{2.5}$$

where the parameter c (with $-1 < c < 1$) determines the strength of correlation. The inverse of $\mathbf{A}$ is

$$A_{ij}^{-1} = \frac{\delta_{ij}(Nc + 1 - c) - c}{(1 - c)(Nc + 1 - c)}. \tag{2.6}$$

The other correlation structure is of limited range (Johnson, 1980; Snippe & Koenderink, 1992; Abbott & Dayan, 1999) with the covariance matrix

$$A_{ij} = b^{|i-j|}, \tag{2.7}$$

where the parameter b (with $0 < b < 1$) determines the range of correlation. The model captures the fact that the correlation strength between neurons decreases with dissimilarity in their preferred stimuli, a property often observed in cortical areas. The limited-range correlation is translational invariant in the sense that $A_{ij} = A_{kl}$, if $|i - j| = |k - l|$. The inverse of $\mathbf{A}$ is

$$A_{ij}^{-1} = \frac{1 + b^2}{1 - b^2}\left[\delta_{ij} - \frac{b}{1 + b^2}(\delta_{i+1,j} + \delta_{i-1,j})\right]. \tag{2.8}$$

There are many possible ways to choose an unfaithful decoding model, depending on the trade-off between computational complexity and decoding accuracy. For the above correlation structures, a natural choice is

$$p(\mathbf{r}|x) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \exp\left[-\frac{1}{2\sigma^2}\sum_i (r_i - f_i(x))^2\right], \tag{2.9}$$

which neglects the neural correlation in the encoding process but keeps the tuning functions unchanged.

2.2 The Decoding Error of UMLI and FMLI. The decoding error of UMLI has been studied in the statistical theory (Akahira & Takeuchi, 1981;

Murata, Yoshizawa, & Amari, 1994). Here we generalize it to the population coding. The decoding error of FMLI can be solved similarly.

For convenience, some notations are introduced. $\nabla f(\mathbf{r}, x)$ denotes $df(\mathbf{r}, x)/dx$. $E_q[f(\mathbf{r}, x)]$ and $V_q[f(\mathbf{r}, x)]$ denote, respectively, the mean value and the variance of $f(\mathbf{r}, x)$ with respect to the distribution $q(\mathbf{r}|x)$.

Given an observation of the population activity $\mathbf{r}^\star$, the UMLI estimate $\hat{x}$ is the value of x that maximizes the log-likelihood $L_p(\mathbf{r}^\star, x) = \ln p(\mathbf{r}^\star|x)$.

Denote by x_{opt} the value of x satisfying $E_q[\nabla L_p(\mathbf{r}, x_{\text{opt}})] = 0$. For the faithful model where $p = q$, $x_{\text{opt}} = x$. Hence, $(x_{\text{opt}} - x)$ is the error due to the unfaithful setting, whereas $(\hat{x} - x_{\text{opt}})$ is the error due to sampling fluctuations. For the unfaithful model, equation 2.9, since $E_q[\nabla L_p(\mathbf{r}, x_{\text{opt}})] = 0$,

$$\sum_i [f_i(x) - f_i(x_{\text{opt}})] f_i'(x_{\text{opt}}) = 0. \tag{2.10}$$

Hence, $x_{\text{opt}} = x$, and UMLI gives an unbiased estimator in this cases.

Let us consider the expansion of $\nabla L_p(\mathbf{r}^\star, \hat{x})$ at x,

$$\nabla L_p(\mathbf{r}^\star, \hat{x}) \simeq \nabla L_p(\mathbf{r}^\star, x) + \nabla\nabla L_p(\mathbf{r}^\star, x)\ (\hat{x} - x). \tag{2.11}$$

Since $\nabla L_p(\mathbf{r}^\star, \hat{x}) = 0$,

$$\frac{1}{N}\nabla\nabla L_p(\mathbf{r}^\star, x)\ (\hat{x} - x) \simeq -\frac{1}{N}\nabla L_p(\mathbf{r}^\star, x), \tag{2.12}$$

where N is the number of neurons. Only the large N limit is considered in this study.

Let us analyze the properties of the two random variables $\frac{1}{N}\nabla\nabla L_p(\mathbf{r}^\star, x)$ and $\frac{1}{N}\nabla L_p(\mathbf{r}^\star, x)$. By using equation 2.9, we get

$$\frac{1}{N}\nabla L_p(\mathbf{r}^\star, x) = \frac{1}{N\sigma^2}\sum_i [r_i^\star - f_i(x)] f_i'(x), \tag{2.13}$$

$$\frac{1}{N}\nabla\nabla L_p(\mathbf{r}^\star, x) = \frac{1}{N\sigma^2}\sum_i [r_i^\star - f_i(x)] f_i''(x) - \frac{1}{N\sigma^2}\sum_i f_i'(x)^2, \tag{2.14}$$

where $f_i'(x) = df_i(x)/dx$ and $f_i''(x) = d^2 f_i(x)/dx^2$.

2.2.1 The Uniform Correlation Case. We consider first the uniform correlation model. In this case, we can write

$$r_i^\star = f_i(x) + \sigma(\epsilon_i + \eta), \tag{2.15}$$

where η and $\{\epsilon_i\}$, for $i = 1, \ldots, N$, are independent random variables having zero mean and variance c and $1 - c$, respectively. η is the common noise for

all neurons, representing the uniform character of the correlation. It is easy to check that the above formulated $\mathbf{r}^\star$ has the required correlation structure of equation 2.5.

Substituting equation 2.15 into 2.13 and 2.14, we get

$$\frac{1}{N}\nabla L_p(\mathbf{r}^\star, x) = \frac{1}{N\sigma}\sum_i \epsilon_i f_i'(x) + \frac{\eta}{N\sigma}\sum_i f_i'(x), \tag{2.16}$$

$$\frac{1}{N}\nabla\nabla L_p(\mathbf{r}^\star, x) = \frac{1}{N\sigma}\sum_i \epsilon_i f_i''(x) - \frac{1}{N\sigma^2}\sum_i f_i'(x)^2 + \frac{\eta}{N\sigma}\sum_i f_i''(x). \tag{2.17}$$

Without loss of generality, we assume that the distribution of the preferred stimuli is uniform. For the tuning functions of the form 2.3, hence, $\sum_i f_i'(x)/N$ and $\sum_i f_i''(x)/N$ approach zero when N is large. Therefore, the correlation contributions (the terms of η) in equations 2.16 and 2.17 can be neglected. UMLI performs in this case as if the neuronal signals are uncorrelated.

Thus, by the weak law of large numbers,

$$\frac{1}{N}\nabla\nabla L_p(\mathbf{r}^\star, x) \simeq -\frac{1}{N\sigma^2}\sum_i f_i'(x)^2 = \frac{Q_p}{N}, \tag{2.18}$$

where

$$Q_p \equiv E_q[\nabla\nabla L_p(\mathbf{r}, x)]. \tag{2.19}$$

According to the central limit theorem, $\nabla L_p(\mathbf{r}^\star, x)/N$ converges to a gaussian distribution,

$$\frac{1}{N}\nabla L_p(\mathbf{r}^\star, x) \sim \mathcal{N}\left(\prime, \frac{\infty - \rfloor}{\mathcal{N}^\in\sigma^\in}\sum_{\rangle}\{_\rangle'(\S)^\in\right) = \mathcal{N}\left(\prime, \frac{\mathcal{G}}{\mathcal{N}^\in}\right), \tag{2.20}$$

where $N(0, t^2)$ denoting the gaussian distribution having zero mean and variance t^2, and

$$G_p \equiv V_q[\nabla L_p(\mathbf{r}, x)]. \tag{2.21}$$

Combining the results of equations 2.12, 2.18, and 2.20, we obtain the decoding error of UMLI,

$$(\hat{x} - x)_{\mathrm{UMLI}} \sim \mathcal{N}(\prime, \mathcal{Q}^{-\in}_{\surd}\mathcal{G}_{\surd}),$$
$$= \mathcal{N}(\prime, \frac{(\infty - \rfloor)\sigma^{\in}}{\mathcal{N}\mathcal{F}_{\infty}(\S)}), \tag{2.22}$$

where

$$F_1(x) = \frac{1}{N}\sum_i f_i'(x)^2, \tag{2.23}$$

is scaled to be of order one. The decoding error of UMLI decreases as a function of c and N.

In a similar way, the decoding error of FMLI is obtained,

$$(\hat{x} - x)_{\mathrm{FMLI}} \sim \mathcal{N}(\prime, \mathcal{Q}_{\mathrm{II}}^{-\in}\mathcal{G}_{\mathrm{II}}),$$
$$= \mathcal{N}\left(\prime, \frac{(\infty - \rfloor)\sigma^{\in}}{\mathcal{N}\mathcal{F}_{\infty}(\S)}\right), \tag{2.24}$$

which has the same form as that of UMLI except that Q_q and G_q are now defined with respect to the faithful decoding model, that is, $p(\mathbf{r}|x) = q(\mathbf{r}|x)$. To get equation 2.24, the condition $\sum_i f_i'(x) = 0$ is used. Interestingly, UMLI and FMLI have the same decoding error because the uniform correlation effect is actually neglected in both UMLI and FMLI.

Note that in FMLI, $G_q = -Q_q = V_q[\nabla L_q(\mathbf{r}|x)]$ is the Fisher information. $Q_q^{-2}G_q$ is therefore the Cramér-Rao bound, which is the optimal accuracy for an unbiased estimator to achieve. The above derivation of equation 2.24 indicates that FMLI is asymptotically efficient in the uniform correlation model. For an unfaithful decoding model, Q_p and G_p are usually different from the Fisher information. We call $Q_p^{-2}G_p$ the generalized Cramér-Rao bound and UMLI quasi-asymptotically efficient if its decoding error approaches $Q_p^{-2}G_p$ asymptotically. Equation 2.22 shows that UMLI is quasi-asymptotic efficient.

We calculate the Cramér-Rao bound and the generalized Cramér-Rao bound for the gaussian response models, equations 2.1 and 2.9 (valid for general correlation structures).

From the definitions,

$$Q_p = -\frac{1}{\sigma^2}\sum_i f_i'(x)^2, \tag{2.25}$$

$$G_p = \frac{1}{\sigma^2}\sum_{i,j} A_{ij} f_i'(x) f_j'(x), \tag{2.26}$$

$$Q_q = -\frac{1}{\sigma^2}\sum_{ij} A_{ij}^{-1} f_i'(x) f_j'(x), \tag{2.27}$$

$$G_q = \frac{1}{\sigma^2}\sum_{ij} A_{ij}^{-1} f_i'(x) f_j'(x). \tag{2.28}$$

Thus,

$$Q_q^{-2} G_q = \frac{\sigma^2}{\sum_{ij} A_{ij}^{-1} f_i'(x) f_j'(x)}, \tag{2.29}$$

$$Q_p^{-2} G_p = \frac{\sigma^2 \sum_{ij} A_{ij} f_i'(x) f_j'(x)}{[\sum_i f_i'(x)^2]^2}. \tag{2.30}$$

2.2.2 The Limited-Range Correlation Case. In section 2.2.1, the asymptotic efficiency of FMLI and UMLI is proved when the neuronal correlation is uniform. The result relies on the radial symmetry of the tuning functions and the uniform character of the correlation, which make it possible to cancel the correlation contributions from different neurons. For general tuning functions and correlation structures, the asymptotic efficiency of UMLI and FMLI may not hold. This is because the law of large numbers (see equation 2.18) and the central limit theorem (see equation 2.20) are not in general applicable.

We note that for the limited-range correlation model, since the correlation is translationally invariant and its strength decreases quickly with the dissimilarity in the neurons' preferred stimuli, the correlation effect in the decoding of FMLI and UMLI becomes negligible when N is large. This ensures that the law of large numbers and the central limit theorem hold in the large N limit. Therefore, FMLI and UMLI are asymptotically and quasi-asymptotically efficient, respectively, in this case. This is confirmed in the simulation in section 3.2.

When FMLI and UMLI are asymptotically and quasi-asymptotically efficient, their decoding errors in the large N limit can be described by the Cramér-Rao bound (see equation 2.29) and the generalized Cramér-Rao bound (see equation 2.30), respectively.

From equations 2.7 and 2.30, we get

$$\langle(\hat{x} - x)^2\rangle_{\text{UMLI}} \sim \frac{\sigma^2 F_2(x)}{N F_1(x)^2}, \tag{2.31}$$

where

$$F_2(x) = \frac{1}{N}\sum_{i,j} b^{|i-j|} f_i'(x) f_j'(x). \tag{2.32}$$

For a fixed value of N, $F_2(x)$ is a nonmonotonic function of the parameter b; so is the decoding error of UMLI. Note that the contribution of

$f_i'(x)f_j'(x)$ to the summation is weighted by $b^{|i-j|}$, which decreases exponentially with $|i-j|$. In the case of small b, those $f_i'(x)f_j'(x)$ having small values of $|i-j|$ contribute to the major part of $F_2(x)$. They are mainly positive, excluding the region around the stimulus x. Therefore, $F_2(x)$ increases with b when b is small. When b is large enough, more negative $f_i'(x)f_j'(x)$ (having large values of $|i-j|$ with $c_i < x < c_j$, or $c_i > x > c_j$) start to contribute, so that $F_2(x)$ decreases with b. This property is confirmed in section 3.

For a fixed value of b, in the large N limit, the difference between f_i' and $f_j'(x)$ is negligible when $|i-j|$ is small.[1] Thus,

$$F_2(x) \sim \frac{1}{N}\sum_i f_i'(x)^2(1+2b+2b^2+\cdots)$$
$$\sim F_1(x)\frac{1+b}{1-b}, \tag{2.33}$$

and

$$\langle(\hat{x}-x)^2\rangle_{\mathrm{UMLI}} \sim \frac{\sigma^2(1+b)}{N(1-b)F_1(x)}, \tag{2.34}$$

which is a decreasing function of N.

The decoding error of FMLI has the same asymptotic behavior, which has been analyzed by Abbott and Dayan (1999). Comparing equation 2.34 with their result (equation 4.12 in Abbott & Dayan, 1999), we see that UMLI and FMLI have the same decoding error in the large N limit.

2.3 Implementation of UMLI by a Recurrent Network Model. Recently Pouget et al. (1998) and Deneve et al. (1999) proposed a biologically feasible RNM, which performs population decoding in a coarse code format. For the correlation models we consider in this article, RNM can implement UMLI.

The essential idea of RNM may be understood as constructing a dynamic system that is triggered by a stimulus x and evolves into a stationary state satisfying

$$\sum_i [r_i - f_i(\hat{x})]v_i(\hat{x}) = 0, \tag{2.35}$$

where the functions $v_i(x)$ for $i = 1, \ldots, N$ are determined by the network parameters, and $\hat{x}$ is the network estimation.

It is easy to check that for the encoding model (see equation 2.1) and the unfaithful decoding model (see equation 2.9), the estimates of FMLI and

[1] For the gaussian tuning function, the difference between f_i' and $f_j'(x)$ is the order of $1/N$ when $|i-j|$ is small. For the triangular tuning function, the difference is zero in most cases, but can be ± 2 in the region where the preferred stimuli are close to the stimulus. In the large N limit, the contribution of this region is negligible.

UMLI are determined, respectively, by

$$\sum_i [r_i - f_i(\hat{x})] A_{ij}^{-1} f_j'(\hat{x}) = 0, \tag{2.36}$$

$$\sum_i [r_i - f_i(\hat{x})] f_i'(\hat{x}) = 0. \tag{2.37}$$

Hence, if RNM is designed such that, $v_i(x) \propto A_{ij}^{-1} f_j'(x)$, or $v_i(x) \propto f_i'(x)$, it implements FMLI, or UMLI, respectively. An example of using RNM to implement UMLI is given in Pouget et al. (1998, 1999).

3 Performance Comparison

The performance of UMLI is compared with that of FMLI and the COM decoding method.

The neural population model we consider is a regular array of N neurons (Baldi & Heiligenberg, 1988; Snippe, 1996) with the preferred stimuli uniformly distributed in the range $[-D, D]$, that is, $c_i = -D + 2iD/(N+1)$, for $i = 1, \ldots, N$. The comparison is done at the stimulus $x = 0$. Two different tuning functions are studied: the triangular function and the gaussian one.

Compared with FMLI, UMLI has an equivalent or larger decoding error, since it uses less information of the encoding process. By simplifying the covariance matrix, UMLI largely reduces the computational cost.

COM is a simple decoding method without using any information of the encoding process, whose estimate is the averaged value of the neurons' preferred stimuli weighted by the responses (Georgopoulos, Kalaska, Caminiti, & Massey, 1982; Snippe, 1996), that is,

$$\hat{x} = \frac{\sum_i r_i c_i}{\sum_i r_i}. \tag{3.1}$$

The shortcoming of COM is a large decoding error.

The decoding errors of COM for the two correlation structures we study are calculated to be

$$\langle (\hat{x} - x)^2 \rangle_{\text{COM}} \sim \frac{\sigma^2 \sum_{ij} A_{ij} c_i c_j}{[\sum_i f_i(x)]^2}, \tag{3.2}$$

where the condition $\sum_i f_i(x) c_i = 0$ is used, due to the regularity of the distribution of the preferred stimuli.

3.1 The Case for the Triangular Tuning Function. The triangular tuning function has the form (Brunel & Nadal, 1998)

$$f_i(x) = \begin{cases} 1 - |x - c_i|/a & |x - c_i| \le a \\ 0 & |x - c_i| > a \end{cases} \tag{3.3}$$

where the parameter a is the tuning width.

We note that the gaussian response model does not give zero probability for negative firing rates. To make it more reliable, we set $r_i = 0$ when $f_i(x) = 0$ (i.e., $|c_i - x| > a$), which means that only active enough neurons contribute to the decoding. This seems to be biologically plausible by considering the time limitation of the decoding process. The mathematical effect of the cutoff will be discussed later. For the tuning width a, there are effectively $N = \text{Int}[2a/d - 1]$ neurons involved in the decoding process, where d is the difference in the preferred stimulus between two consecutive neurons and the function Int[·] denotes the integer part of the argument.

Since the triangular tuning function is piece-wise linear, the decoding errors of UMLI and FMLI can be calculated exactly (see the appendix).

Figure 1 compares the decoding errors of the three methods for the uniform correlation model. It shows that UMLI has the same decoding error as that of FMLI and a lower error than that of COM.

Figure 2 compares the decoding errors of the three methods for the limited-range correlation model. We see that UMLI has a lower decoding error than that of COM and a larger error than that of FMLI (finite N effect). For a fixed value of b, the discrepancy between FMLI and UMLI decreases with the number of neurons (see Figure 2a). For a fixed value of N, the discrepancy increases with the correlation strength and decreases subsequently when the correlation is very strong (see Figure 2b).

Figures 1b and 2b also illustrate the effect of correlation on the decoding accuracies. We see that the uniform correlation improves the decoding accuracies in the three methods; the limited-range correlation degrades the decoding accuracies when the correlation strength is small and improves the accuracies when the strength is large enough.

3.2 The Case for the Gaussian Tuning Function. The gaussian tuning function has the form

$$f_i(x) = \exp\left(-\frac{(x - c_i)^2}{2a^2}\right), \tag{3.4}$$

where the parameter a is the tuning width.

Again, to make the gaussian response model more reliable, we set $r_i = 0$ when $f_i(x) < 0.11$ ($|x - c_i| > 3a$). For the tuning width a, there are $N = \text{Int}[6a/d - 1]$ neurons involved in the decoding process, where d is the difference in the preferred stimuli between two consecutive neurons.

Figure 3 compares the decoding errors of the three methods for the uniform correlation model. It shows that UMLI has the same decoding error as that of FMLI[2] and a lower error than that of COM.

[2] We prove that UMLI and FMLI have the same decoding error in the large N limit. Here, they have the same error in the finite N due to the regular distribution of preferred stimuli.

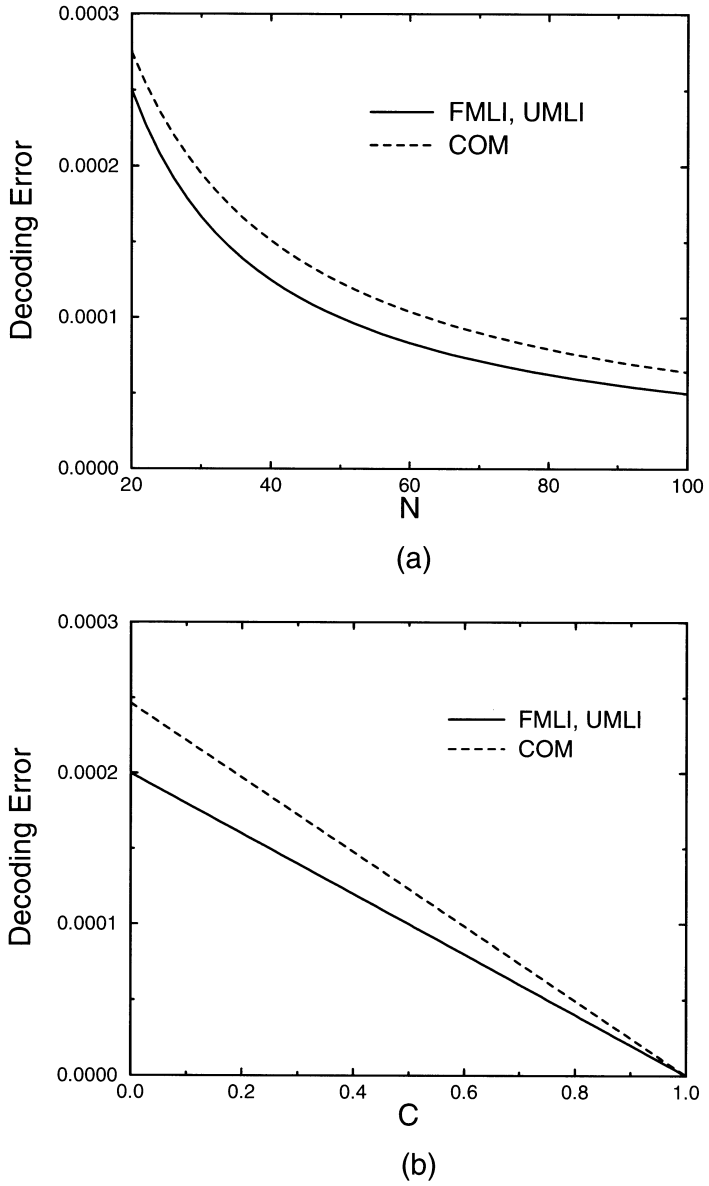


Figure 1: Comparison of the decoding errors of UMLI, FMLI, and COM for the uniform correlation model and the triangular tuning function. The parameters are chosen as $a = 1$ and $\sigma = 0.1$. (a) The decoding errors change with the number of neurons, N. The correlation strength c is 0.5. (b) The decoding errors change with the correlation strength c. The number of neurons, N, is 50.

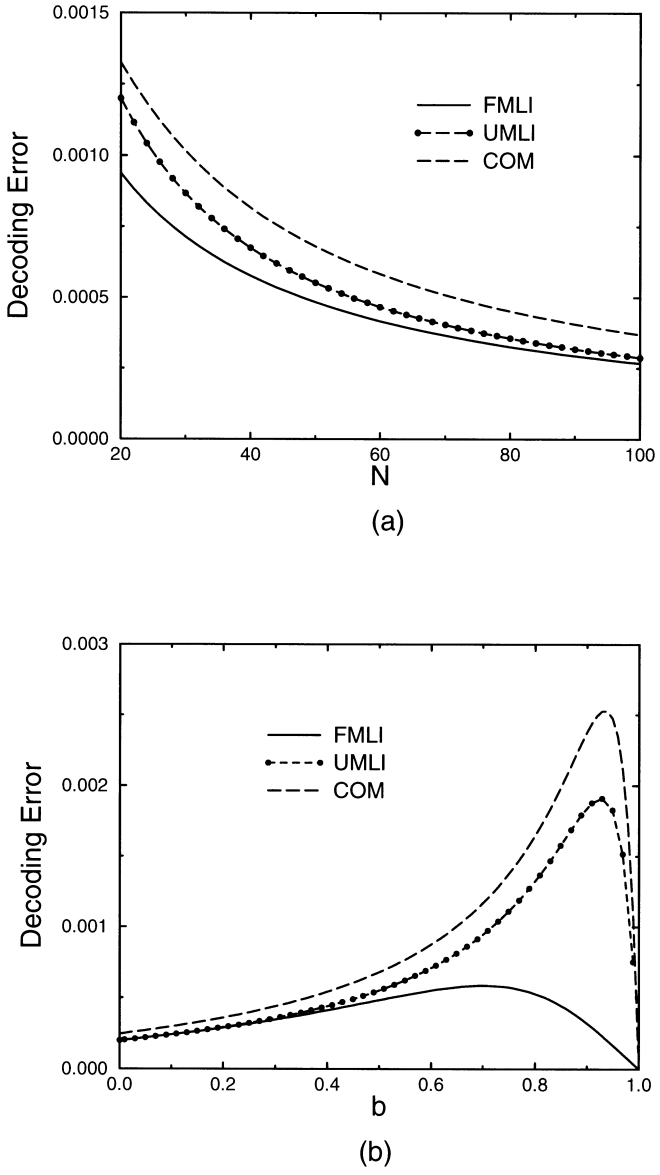


Figure 2: Comparison of the decoding errors of UMLI, FMLI, and COM for the limited-range correlation model and the triangular tuning function. The parameters are chosen as $a = 1$ and $\sigma = 0.1$. (a) The decoding errors change with the number of neurons, N. The correlation strength b is 0.5. (b) The decoding errors change with the correlation strength b. The number of neurons, N, is 50.

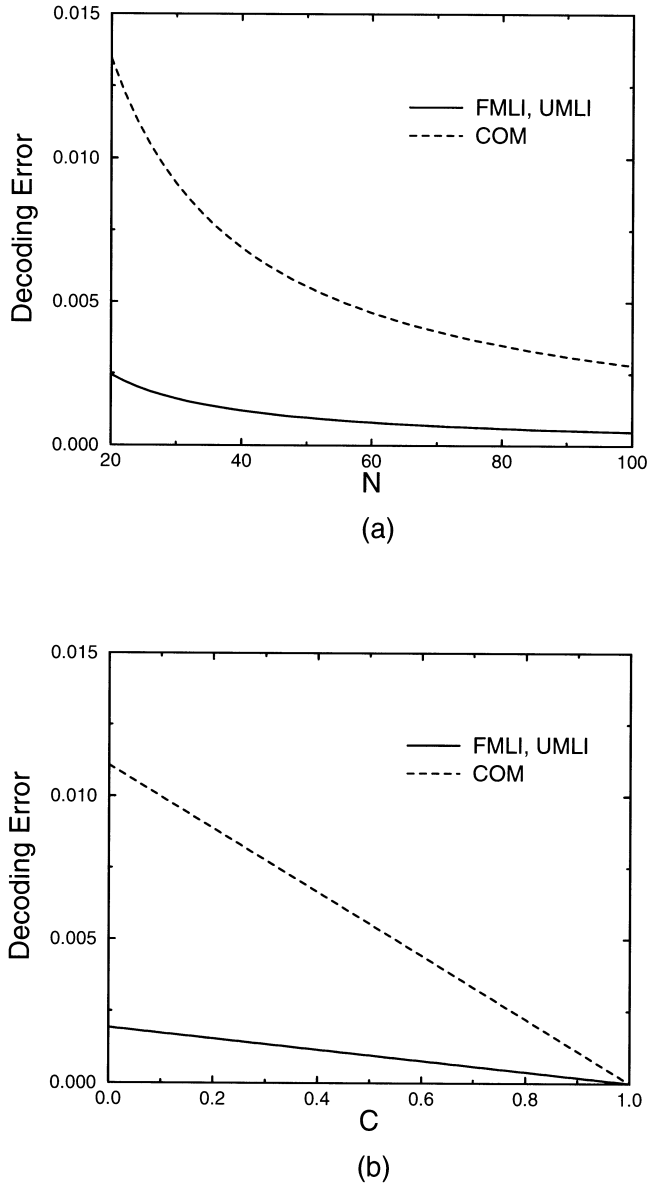


Figure 3: Comparison of the decoding errors of UMLI, FMLI, and COM for the uniform correlation model and the gaussian tuning function. The parameters are chosen as $a = 1$ and $\sigma = 0.1$. (a) The decoding errors change with the number of neurons, N. The correlation strength c is 0.5. (b) The decoding errors change with the correlation strength c. The number of neurons, N, is 50.

In Figure 4, the simulation results for the decoding errors of FMLI and UMLI in the limited-range correlation model are compared with those obtained by using the Cramér-Rao bound and the generalized Cramér-Rao bound, respectively. It shows that the two results agree very well when the number of neurons, N, is large, which means that FMLI and UMLI are asymptotically and quasi-asymptotically efficient as we analyzed. We also see that a stronger correlation leads to a slower convergence for the decoding error of FMLI to the Cramér-Rao bound. For a finite value of N, their discrepancy can be quite large in the strong correlation case.

In Figure 4a, we observe an interesting behavior for the Cramér-Rao bound (the case of $b = 0.8$). It increases first when N is small and decreases subsequently when N is large. Asymptotically, it is inversely proportional to N, or, equivalently, the Fisher information is proportional to N. The reason can be understood as follows. When the number of neurons, N, increases, on one hand, the number of signals becomes larger, which tends to increase the Fisher information. On the other hand, the total correlation in the population is also enlarged, which tends to decrease the Fisher information. These two factors compete and generate the above behavior. Moreover, since the correlation strength between two consecutive neurons, b, is constant, the former factor finally dominates. Thus, asymptotically, the Fisher information increases proportionally with N.

In the simulation, the standard gradient-descent method is used to maximize the log-likelihood, and the initial guess for the stimulus is chosen as the preferred stimulus of the most active neuron. The CPU time of UMLI is around one-fifth that of FMLI when the number of neurons $N = 50$. UMLI reduces the computational cost of FMLI significantly.

Figure 5 compares the decoding errors of the three methods for the limited-range correlation model. The simulation results for FMLI and UMLI are used. The figure shows that UMLI has a lower decoding error than that of COM and a comparable performance with that of FMLI.

Again, we observe that the uniform correlation improves the decoding accuracies of the three methods (see Figure 3b). The limited-range one degrades the decoding accuracies when the correlation strength is small and improves the accuracies when the strength is large enough (see Figure 5b).

To make the gaussian response model more plausible, we have forced the neuronal activity to be zero when the value of the tuning function is under a threshold (0 for the triangular tuning function and 0.11 for the gaussian one). This partly compensates the defect of the gaussian model of having nonzero probability for negative firing rates. It is easy to see that this cutoff does not affect much the results of UMLI and FMLI, due to their nature of decoding by using the derivative of the tuning functions, whereas the decoding error of COM will be greatly enlarged without cutoff. Without cutoff, the range of the preferred stimuli should be finite to avoid divergence of the decoding error of COM. In this case, the model will become more similar to the one used by Snippe (1996), whose calculation shows that COM is extremely

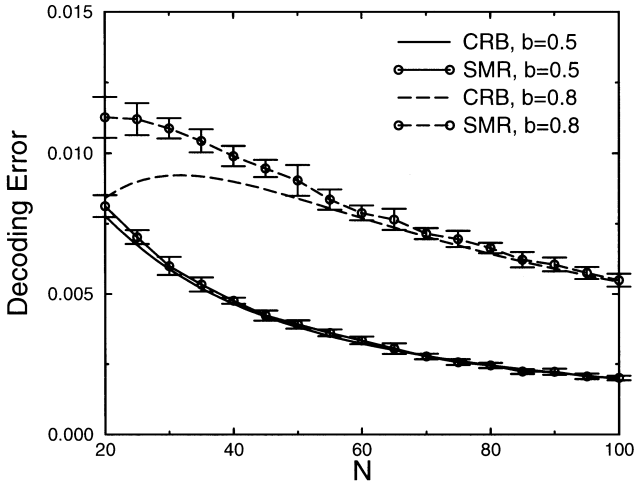
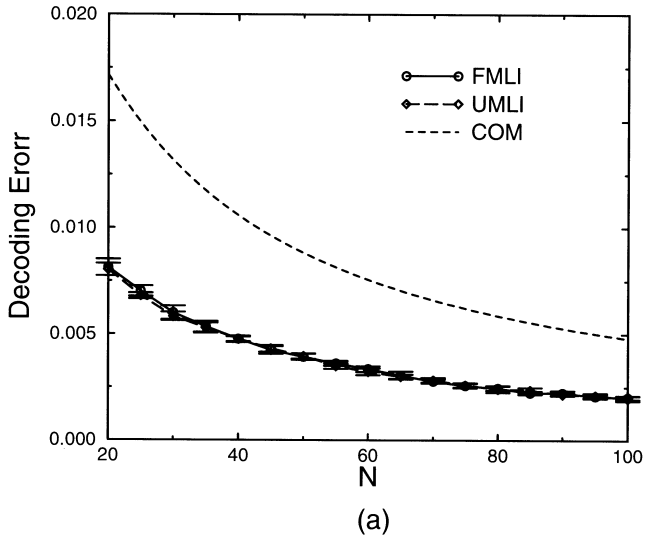


Figure 4: Comparison of the simulation results of the decoding errors of UMLI and FMLI in the limited-range correlation model with those obtained by using the Cramér-Rao bound (CRB) and the generalized Cramér-Rao bound (GCRB), respectively. The parameters are chosen as $a = 1$ and $\sigma = 0.1$. SMR denote the simulation results. In the simulation, 10 sets of data are generated, each averaged over 1000 trials. (a) FMLI. (b) UMLI.



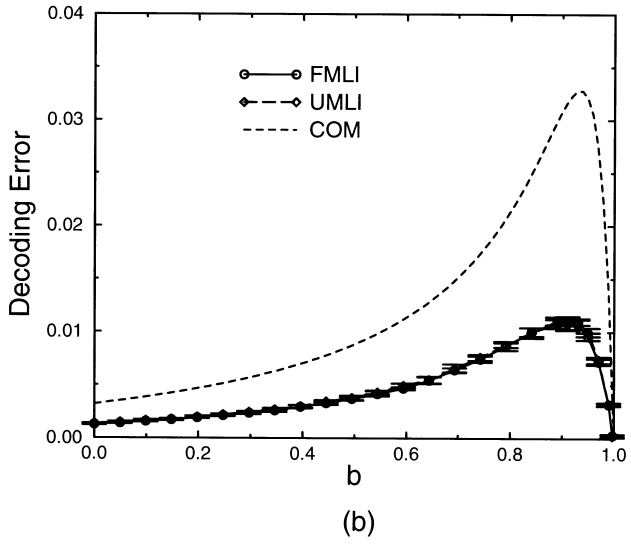


Figure 5: Comparison of the decoding errors of UMLI, FMLI, and COM for the limited-range correlation model and the gaussian tuning function. The parameters are chosen as $a = 1$ and $\sigma = 0.1$. The simulation results for FMLI and UMLI are very close. (a) The decoding errors change with the number of neurons, N. The correlation strength b is 0.5. (b) The decoding errors change with the correlation strength b. The number of neurons, N, is 50.

inefficient compared with the optimal decoding accuracy when the tuning width is small. Thus, cutoff benefits only COM. Our conclusion does not change except for this point.

In the above calculation, we have not considered the neuronal spontaneous activity. If this factor is included, for example, $\langle r_i \rangle = f_i(x)$ is replaced by $\langle r_i \rangle = f_i(x) + \gamma$, where γ is a small constant representing the background noise, the decoding error of COM will be enlarged, whereas the performance of UML and FMLI will not be affected. This can be understood from the same reason as that in cutoff. Thus, adding a spontaneous term will only make the superiority of UMLI to COM more significant.

The optimal linear estimator (OLE) is another decoding method, with computational complexity and accuracy between those of COM and FMLI (Salinas & Abbott, 1994). Since it uses information from the neural correlation, we do not compare it with UMLI.

4 Conclusions and Discussions

We have studied a population decoding paradigm in which MLI is based on an unfaithful model. This is motivated by the fact that the encoding process of the brain is not exactly known by the estimator and that a properly simplified decoding model can save computational cost without sacrificing much decoding accuracy.

As an example, we consider an unfaithful decoding model that neglects the pair-wise correlation between neuronal activities. Two different correlation structures are studied: the uniform and the limited-range correlations. We proved the asymptotic efficiency of FMLI and the quasi-asymptotic efficiency of UMLI for the two correlations.

The performance of UMLI is compared with that of FMLI and of COM. It turns out that UMLI has a lower decoding error than that of COM. Compared with FMLI, UMLI has an equivalent or larger decoding error, though with much less computational cost. In the large N limit, UMLI and FMLI have the same decoding error for both uniform and limited-range correlations. Furthermore, UMLI can be implemented by a biologically feasible recurrent network (Pouget et al., 1998; Deneve, Latham, & Pouget, 1999). Our future work seeks to understand the biological implication of UMLI.

Yoon and Sompolinsky (1999) studied a different correlation structure, referred to as the YS model. Unlike the limited-range correlation model, the correlation strength between two consecutive neurons in the YS model increases with N, and hence, the correlation range is finite even when N is infinity. Interestingly, Yoon and Sompolinsky have shown that the Fisher information in this case does not increase asymptotically in proportion to N but stays at a constant level. Since the correlation effect in the YS model is not negligible for large N, the law of large numbers (see equation 2.18) and the central limit theorem (see equation 2.20) do not hold. FMLI and UMLI are not consistent in this case. For simplicity, we study a simplified

YS model through setting $b = h^{1/N}$, for $0 < h < 1$, in the limited-range correlation (Abbott & Dayan, 1999) (It corresponds to the case of $a = 0$ in the general YS model. In this article, we use $a = 0$. However, we believe that $a \neq 0$ has an important effect when $b = h^{1/N}$.). The simulation results are shown in Figure 6. It shows that UMLI and FMLI are not consistent, and their decoding errors tend to constant levels in the large N limit (see Figure 6a). UMLI has a comparable performance as that of FMLI and a lower error than that of COM (see Figure 6b).

The effect of correlation on the decoding accuracies of FMLI, UMLI, and COM is investigated for the two correlation structures. It turns out that the correlation, depending on its form, can either improve or degrade the decoding accuracy. This observation agrees with the analysis of Abbott and Dayan (1999), which is done with respect to the optimal decoding accuracy, that is, the Cramér-Rao bound, not on specific decoding methods. Our result shows further that these properties hold in the three decoding methods. Through exchanging the roles of $\mathbf{A}$ and $\mathbf{A}^{-1}$, the limited-range correlation model can be easily generalized to be a new one, in which neurons are negatively correlated with nearest neighbors. It is easy to check that all results are in the same form except that $\mathbf{A}$ and $\mathbf{A}^{-1}$ are exchanged. Therefore, the effect of correlation becomes reverse, which improves the decoding accuracy when the strength is small and degrades when the strength is large.

This article has studied only the pair-wise correlation with a structure being uniform or of limited range. The real neuron correlation may be more complex and have orders of more than two. However, if the radial symmetry of the tuning functions and the translational invariance of the correlation are not violated so much, which seems to be true, we still expect that UMLI holds the beneficial properties discussed above.

Appendix: The Decoding Errors of UMLI and FMLI in the Case of Triangular Tuning Functions

For the triangular tuning function,

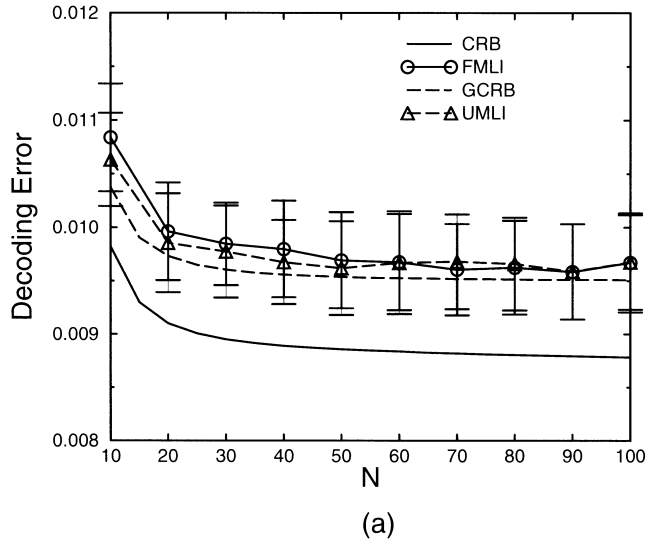
$$f_i'(x) = \begin{cases} \text{sign}(x - c_i)/a & |x - c_i| < a, \quad x \neq c_i \\ 0 & |x - c_i| > a \end{cases} \tag{A.1}$$

where $\text{sign}(x - c_i)$ denoting the sign of $(x - c_i)$. The function $f_i'(x)$ is singular at $c_i = x$ and $c_i = x \pm a$. Without loss of generality, we assume no such preferred stimuli exist.

Denote $r_i = f_i(x) + \xi_i$, for $i = 1, \ldots, N$, where $\{\xi_i\}$ are random numbers satisfying

$$\langle \xi_i \rangle = 0, \tag{A.2}$$

$$\langle \xi_i \xi_j \rangle = \sigma^2 A_{ij}. \tag{A.3}$$



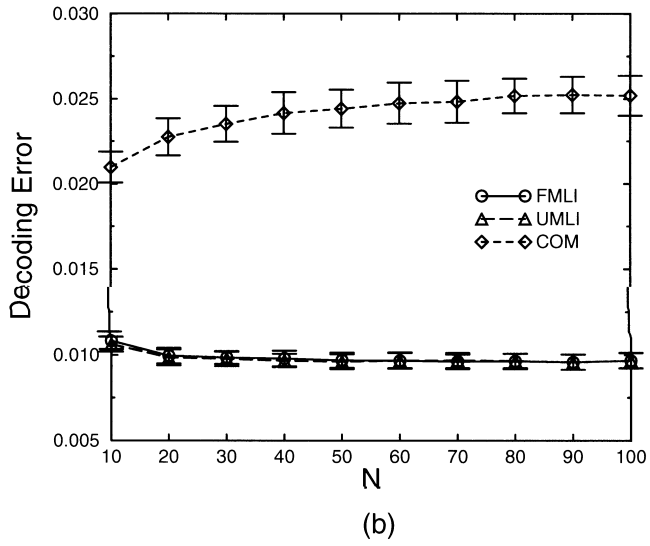


Figure 6: (a) Comparison of the simulation results of the decoding errors of UMLI and FMLI in the simplified YS model with those obtained by using the Cramér-Rao bound (CRB) and the generalized Cramér-Rao bound (GCRB), respectively. (b) Comparison of the decoding errors of UMLI, FMLI, and COM. The parameters are chosen as $a = 1$, $\sigma = 0.1$, $h = 0.2$. In the simulation, 10 sets of data is generated, each of which is averaged over $10,000$ trials.

The UMLI estimate is the solution of

$$\nabla \ln p(\mathbf{r}|\hat{x}) = 0. \tag{A.4}$$

From equation 2.9 we get

$$\sum_i (r_i - f_i(\hat{x}))f_i'(\hat{x}) = 0. \tag{A.5}$$

Suppose $\hat{x}$ is close enough to x, due to the piece-wise linearity of the triangular function; we then have

$$r_i - f_i(\hat{x}) = \xi_i - (\hat{x} - x)f_i'(x) \tag{A.6}$$

$$f_i'(\hat{x}) = f_i'(x). \tag{A.7}$$

Substituting equations (A.6) and (A.7) into (A.5), we get

$$\hat{x} = x + \frac{\sum_i \xi_i f_i'(x)}{\sum_i (f_i'(x))^2}. \tag{A.8}$$

Therefore,

$$\langle(\hat{x} - x)^2\rangle_{\mathrm{UMLI}} = \frac{\sigma^2 \sum_{ij} A_{ij} f_i'(x) f_j'(x)}{[\sum_i (f_i'(x))^2]^2}. \tag{A.9}$$

The FMLI estimate is the solution of

$$\nabla \ln q(\mathbf{r}|\hat{x}) = 0. \tag{A.10}$$

From equation 2.1 we get

$$\sum_{i,j} A_{ij}^{-1}(r_j - f_j(\hat{x}))f_i'(\hat{x}) = 0. \tag{A.11}$$

By using equations (A.6) and (A.7), we get

$$\hat{x} = x + \frac{\sum_{i,j} A_{ij}^{-1} \xi_j f_i'(x)}{\sum_{i,j} A_{ij}^{-1} f_i'(x) f_j'(x)}. \tag{A.12}$$

Therefore,

$$\langle(\hat{x} - x)^2\rangle_{\mathrm{FMLI}} = \frac{\sigma^2}{\sum_{ij} A_{ij}^{-1} f_i'(x) f_j'(x)}. \tag{A.13}$$

Acknowledgments

We are grateful to Noboru Murata for helping to solve the mathematic problems in this article, and we thank the two anonymous reviewers for their valuable comments. S. W. acknowledges helpful discussions with Peter Dayan and Danmei Chen.

References

Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation, 11*, 91–101.

Akahira, M., & Takeuchi, K. (1981). Asymptotic efficiency of statistical estimators: Concepts and high order asymptotic efficiency. In *Lecture Notes in Statistics 7*. Berlin: Springer-Verlag.

Anderson, C. H. (1994). Basic elements of biological computational systems. *International Journal of Modern Physics C, 5*, 135–137.

Baldi, P., & Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biol. Cybern., 59*, 313–318.

Brunel, N., & Nadal, J.-P. (1998). Mutual information, Fisher information, and population coding. *Neural Computation, 10*, 1731–1757.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience, 2*, 740–745.

Fetz, E., Yoyama, K., & Smith, W. (1991). Synaptic interactions between cortical neurons. In A. Peters & E. G. Jones (Eds.), *Cerebral cortex, 9*. New York: Plenum Press.

Gawne, T. J., & Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience, 13*, 2758–2771.

Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience, 2*, 1527–1537.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 243*, 1416–1419.

Johnson, K. O. (1980). Sensory discrimination: Neural processes preceding discrimination decision. *J. Neurophy., 43*, 1793–1815.

Lee, D., Port, N. L., Kruse, W., & Georgopoulos, A. P. (1998). Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *Journal of Neuroscience, 18*, 1161–1170.

Murata, M., Yoshizawa, S., & Amari, S. (1994). Network information criterion—determining the number of hidden units for an artificial neural network model. *IEEE. Trans. Neural Networks, 5*, 865–872.

Paradiso, M. A. (1988). A theory for use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics, 58*, 35–49.

Pouget, A., Deneve, S., Ducom, J.-C., & Latham, P. (1999). Narrow versus wide tuning curves: What's best for a population code? *Neural Computation, 11*, 85–90.

Pouget, A., Zhang, K., Deneve, S., & Latham, P. E. (1998). Statistically efficeint estimation using population coding. *Neural Computation, 10*, 373–401.

Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience, 1*, 89–107.

Seung, H. S., & Sompolinsky, H. (1993). Simple models for reading neuronal population codes. *Proceeding of the National Academy of Sciences USA, 90*, 10749–10753.

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology, 4*, 569–579.

Snippe, H. P. (1996). Parameter extraction from population codes: A critical assessment. *Neural Computation, 8*, 511–529.

Snippe, H. P., & Koenderink, J. J. (1992). Information in channel-coded systems: Correlated receivers. *Biological Cybernetics, 67*, 183–190.

Wu, S., Nakahara, H., Murata, N., & Amari, S. (2000). Population decoding based on an unfaithful model. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), *Advances in neural information processing systems 12* (pp. 192–198). Cambridge, MA: MIT Press.

Yoon, H., & Sompolinsky, H. (1999). The effect of correlations on the Fisher information of population codes. In M. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 167–173). Cambridge, MA: MIT Press.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Population interpolation of population codes. *Neural Computation, 10*, 403–430.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature, 370*, 140–143.

Received August 29, 1999; accepted June 15, 2000.