Adaptation using Local Information for Maximizing the Global Cost

Thomas Hoch, Gregor Wenning and Klaus Obermayer

TU Berlin, Department of Electrical Engineering and Computer Science, Franklinstr. 28/29, 10587 Berlin, {hoch,grewe,oby}@cs.tu-berlin.de

Abstract

Recently the information transmission properties of noisy, parallel summing threshold arrays, have been investigated and interpreted in a neural coding context (see Stocks, [1] [2]). The mutual information between certain stimuli and corresponding responses displays a maximum as a function of the noise level. This optimal noise level depends on the number N of neurons within the array, information that is not locally available for single neuron adaptation. We give an analytic expression for the optimal noise level, that only depends on locally available information. The result is based upon an approximation to the mutual information. In the large N limit both descriptions coincide.

Key words: Stochastic Resonance, Adaptation, Summing threshold array

1 Introduction

A noise induced maximum in the mutual information is a signature of stochastic resonance (see [3] for a review). Stochastic resonance has been extensively studied in the context of single neurons (see e.g. [4]). Recently the information transmission properties of noisy, parallel summing threshold arrays, have been investigated and interpreted in a neural coding context (see Stocks, [1] [2] [5]). The mutual information between certain stimulus and response distributions is maximized as a function of the noise level. This optimal noise level depends on the number N of array elements. Adaptation of single neurons to the optimal noise level would require the knowledge of N. But this is information about the global architecture of the system and it is not - in an obvious way - available locally at the single cell. In this study we give an analytic expression for the optimal noise level, which depends only on information which is - in principal - locally available at any single neuron. This local optimality criterium is based upon an approximation of the mutual information. In the

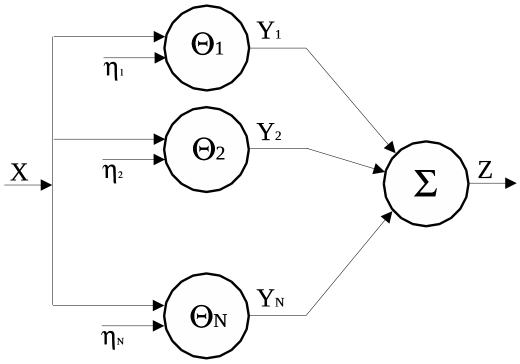

Fig. 1. The model consists of an array of threshold elements. Each neuron receives the same input (drawn from a gaussian distribution X) and is corrupted with additive gaussian noise $\eta_i, i \in \{1, .., N\}$, which is assumed to be independent of the signal and the other noise sources.

large N limit both descriptions yield the same optimal noise level. The quality of the approximation is investigated numerically.

This paper is organized as follows. First we introduce the abstract model which describes the input-output relationship of a summing array of parallel threshold elements. In section 3 the mutual information between a distribution of inputs and a distributions of outputs is introduced. Here we give an approximation to the standard mutual information, based on the work of Brunel et al. [6] and the model given in section 2. Using this approximation we deduce an analytical expression for the optimal noise level which depends only on information, which is locally available at any single neuron. Section 4 contains numerical results in which we demonstrate the quality of our approximation. Section 5, finally, concludes with a brief discussion.

2 Model

The model [1] consists of a parallel summing array of N threshold devices (fig. 1). The input to each threshold device is the sum of the signal X and the noise η_i, which is compared to a constant threshold Θ, to yield the following input-output relation:

$$y_i = sign(x + \eta_i - \Theta) \tag{1}$$

The signal X is the same for all threshold devices and is drawn from a gaussian distribution with variance σ_x. Each neuron is corrupted with gaussian noise η_i with variance $\sigma_\eta, i \in \{1, .., N\}$, which is assumed to be mutually independent of the signal X and the other noise sources. Hence, Z represents the number of devices n that are set to one for a given realization of X. Then the conditional probability $P(Z=n|x)$ can be calculated as follows,

$$P(n|x) = \binom{N}{n} P_{1|x}^{n} (1 - P_{1|x})^{N-n} \tag{2}$$

where the conditional probability $P_{1|x}$, that the output of a neuron is set to one, is the cumulative distribution

$$P_{1|x} = \int_{\Theta - x}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{\eta^2}{2\sigma_\eta^2}\right) d\eta \tag{3}$$

3 Approximation of the mutual information

The average mutual information is an information theoretic measure which quantifies the amount of information the output contains about the input [7]. The mutual information I^{MI} between $P(X)$ and $P(Z)$ is

$$I^{MI} = H(Z) - H(Z|X) \tag{4}$$

$$= -\sum_{n=0}^{N} P(n) \log_2 P(n) + \int_{-\infty}^{\infty} dx P_X(x) \sum_{n=0}^{N} P(n|x) \log_2 P(n|x) \tag{5}$$

N. G. Stocks demonstrates [2], that maximum information transfer occurs at an optimal noise level (σ_{opt}^{MI}). To yield an analytical expression for the optimal noise level, we adapted an approximation introduced by N. Brunel and J. Nadal [6]. They have shown that, in the limit of large N, the mutual information between the signal and the output becomes equal to the mutual information between the signal and an efficient gaussian estimator. The maximum likelihood estimator from $P(n|x)$ exists and is asymptotically unbiased, efficient and gaussian distributed around its mean value with variance $\frac{1}{F(x)}$, and $F(x)$ is the Fisher information:

$$F(x) = E\left[-\frac{\partial^2 \log_2 P_Z(n|x)}{\partial x^2}\right]_x = \left(\frac{\partial P_{1|x}}{\partial x}\right)^2 \frac{N}{P_{1|x}(1 - P_{1|x})} \tag{6}$$

Then we can calculate the amount of information the maximum likelihood estimator contains about the stimulus:

$$I(X, \hat{X}) \;=\; H(\hat{X}) \;-\; \int_{-\infty}^{\infty} dx P_X(x)\; H(\hat{X}|X) \tag{7}$$

In the limit of large N, $I(\hat{X}|X) \rightarrow I^F$, which is equal to

$$I^F \;=\; H(x) \;-\; \int\limits_{-\infty}^{\infty} dx P_X(x) \frac{1}{2} \log_2 \left(\frac{2\pi e}{F(x)} \right). \tag{8}$$

Since processing cannot increase information, the mutual information between the signal and the output is at least equal to I^F [6].

$$I^{MI}(X,Z) \;\geq\; I^F \tag{9}$$

We will use eq. (8) to obtain an analytical expression for the optimal noise level σ^F_{opt}. From the condition

$$\frac{\partial I^F}{\partial \sigma_\eta} = \frac{\partial}{\partial \sigma_\eta} \left(H(x) - \int\limits_{-\infty}^{\infty} dx P_x(x) \frac{1}{2} \log_2 \left(\frac{2\pi e}{F(x)} \right) \right) = 0 \tag{10}$$

we derived the optimal noise level by replacing the $\log_2(\frac{2\pi e}{F(x)})$ with its second order Taylor Series, which is expanded about $x_0 = 0$. This yields

$$\sigma^F_{opt} \simeq \sqrt{\left(1 - \frac{2}{\pi}\right) (\mu_x^2 + \sigma_x^2)}. \tag{11}$$

Note that the optimal noise level depends on the first and second moment of the input distribution.

4 Results

In fig. 2a the mutual information and I^F are plotted against $\sigma = \sigma_n/\sigma_x$ for various N and $\Theta = 0$. The curves indicate that for increasing N, I^F approximates the mutual information. Note that I^F can be negative, because the approximation is good only in the case $F(x) \gg 1$. That means, that the estimator is sharply peaked around its mean value. Furthermore, fig. 2a displays, that the optimal noise level of the Fisher information σ^F_{opt} is independent from the number of threshold devices, where as the optimum of the mutual information shows a dependency on the number of threshold elements, especially for small N. According to our approximation, eq. (8), the optimal noise level is overestimated compared to the optimum of the mutual information, as can be seen in fig. 2b. But due to the asymmetry of the mutual information, the distance between the maxima of I^{MI} and I^F is relatively small and decreases with increasing N. In fig. 3a the relative deviation from the maximum of the

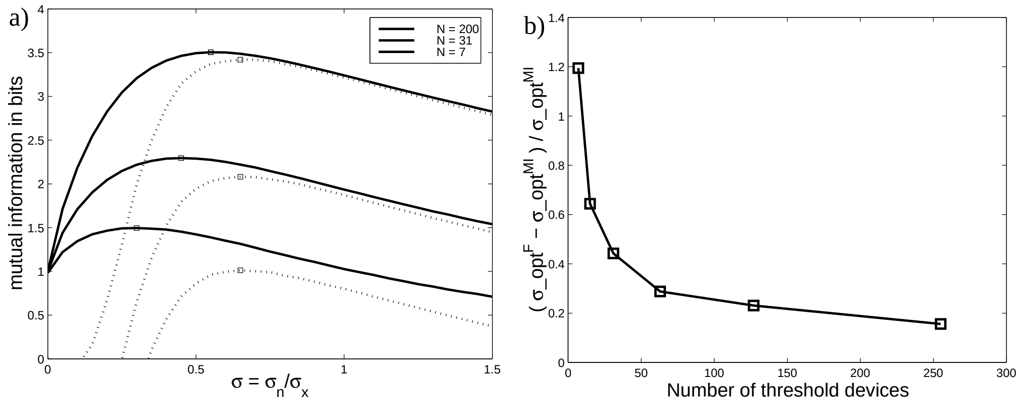

Fig. 2. a) displays the mutual information between the stimulus and the output (solid line) and the mutual information I^F according to eq. (8) (dashed line) are plotted against $\sigma = \frac{\sigma_n}{\sigma_x}$ for various N and $\Theta = \mu_x$. b) displays the relative distance between the optimum noise level of the mutual information and σ^F_{opt} of I^F.

mutual information is presented for increasing N. In fig. 3b, σ^F_{opt} calculated from eq. (11), is compared to the optimal noise level of I^F, obtained from numerical calculations. Because of the Taylor expansion, σ^F_{opt} is slightly below the optimal noise level for I^F.

5 Discussion

In this model study we examined stochastic resonance in a summing parallel threshold array. We compared optimal noise levels of the mutual information I^{MI} between a distribution of inputs and a corresponding distribution of outputs, and I^F, what is an approximation to I^{MI}. The quality of this approximation increases with the number N of parallel nonlinearities. The optimal noise level σ^{MI}_{opt} for I^{MI} depends on N, thus a neuron that should adapt to the optimal noise level needs to have some knowledge about the system size N. But this information is not locally available at a single neuron. The optimal noise level σ^F_{opt} in terms of I^F does not depend on N, but only on variables, which are in principle locally available at any single neuron.
For all N, σ^F_{opt} is larger than σ^{MI}_{opt} and with increasing N this difference decreases. Due to the asymmetry of the stochastic resonance curves (information vs. noise) a moderate overestimation of the optimal noise level does not degrade the amount of information transmitted in a dramatic way. Thus adaptation to σ^F_{opt} instead of σ^{MI}_{opt} is feasible and reasonable. Furthermore the analytic expression for the optimal noise level, in terms of I^F, does not only depend on the mean of the input distribution, as in [8], but also on the variance. More general, evaluating the integral in eq. (8) using more terms in the Taylor expansion yields a weighted sum of the moments of the input distribution.

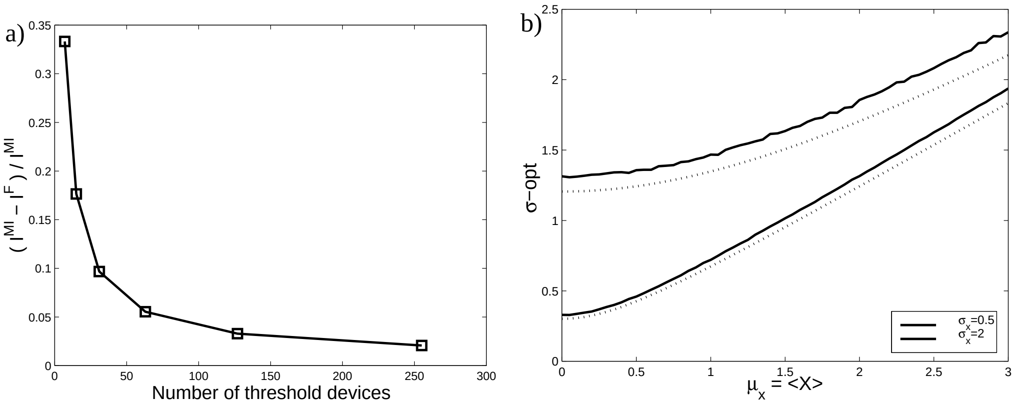

Fig. 3. a) displays the relative deviation of I^F from the maximum of the mutual information I^{MI}. b) Presents the optimal noise level according to eq.(8) (dotted line) and of I^F (solid line) for various μ_x and σ_x. Because of the Taylor expansion, σ_{opt} is slightly below the optimal noise level of I^F, obtained from numerical calculations.

Supported by Wellcome Trust (061113/Z/00).

References

[1] N. G. Stocks, *Suprathreshold stochastic resonance in multilevel threshold systems*, Physical Review Letters, **84**, 2000, p.2310-2314

[2] N. G. Stocks, *Information transmission in parallel threshold arrays: Suprathreshold stochastic resonance*, Physical Review E, **63**, 2001, 041114, p.1-9

[3] L. Gammaitoni, P. Hänggi, P. Jung and F. Marchesoni, *Stochastic resonance*, Reviews of Modern Physics, **70**(1), 1998, p.223-287

[4] D. F. Russel, L. A. Wilkens and F. Moss, *Use of behavioral stochastic resonance by paddle fish for feeding*, Nature, **402**, 1999, p.291-294

[5] N. G. Stocks, *Generic noise-enhanced coding in neuronal arrays*, Physical Review E, **64**, 2001, 030902, p.1-4

[6] N. Brunel and J. Nadal, *Mutual Information, Fisher Information, and Population Coding*, Neural Computation, **10**, 1998, p.1731-1757

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, 1991, 2nd edition

[8] G. Wenning and K. Obermayer, *Adjusting stochastic resonance in a leaky integrate and fire neuron to subthreshold stimulus distributions*, to appear in Neurocomputing, 2002