STATISTICAL MODELING OF BIOCHEMICAL DETECTION SYSTEMS

Sina Zahedi, Reza Navid, Arjang Hassibi

Department of Electrical Engineering, Stanford University, Stanford 94305-9510, CA, USA

ABSTRACT

We study the problem of counting the number of particles in a closed volume where the particles motion are modeled by a Brownian motion process. This problem arises in many biological and chemical sensing experiments, e.g., counting the number of analytes in blood sample, air pollutant measurement, etc. Finding the exact count of particles is challenging in these systems and one generally relies on an estimate based on the sample readouts. We study the statistical properties of the counting process in equilibrium and present the fundamental detection and estimation limitations. In particular, we demonstrate that the count process is inherently noisy and has a quantum-limit signal to noise ratio.

1. INTRODUCTION

Sensors by definition are devices that respond to thermal, electromagnetic, mechanical, or other physical stimulus by producing a signal usually of electrical nature. Biological sensors usually measure the number of particles in the close vicinity of the sensing element [1].

In most chemical and biological experiments and laboratory tests, the sensors are coupled with an aqueous environment where target particles can randomly move in a three dimensional volume. The random motion of the particles in absence of any external force is usually modeled by Brownian motion. As a result of the Brownian motion, some particles enter the sensing volume of a sensor, i.e., the finite volume which the sensor can count, and some particles in the sensing area move outside the area; hence generating some random fluctuations in the readout in the given test sample.

In this paper, we try to characterize the statistical properties of these random fluctuations. In practical situations, the output of a sensor is an analog time-varying signal which is also corrupted by the amplifier noise, i.e., transducer noise, which generally can be modeled as an additive white Gaussian noise with zero mean. In order to characterize the fundamental quantum limits of the sensing process, in this first study, we neglect the effect of amplifier noise and assume that the output of each sensor is a discrete random variable taking integer values.

The final goal of this paper is to find out the fundamental limitations imposed by the random nature of the quantum sensing process and comparing the effect of inherent noise of the process with additive noise of post-processors e.g., amplifier and analog to digital quantizer. Specifically, we compute the autocorrelation function of the observed process as the process evolves in time. Power spectral density (PSD) of the observed process can be computed as the Fourier transform of the autocorrelation function and is of particular importance since it can practically be measured in real experiments. Moreover, it provides some insight into the relative effect of post-processing noise on the overall performance of the measurement system [2].

Section II of this paper is dedicated to the model parameters and assumptions made in solving the problem. In Section III the stochastic analysis of the process has been presented. Analysis of time-dependent fluctuations are presented in Section IV and Section V concludes the paper.

2. PROBLEM DEFINITION AND MODEL

Consider the general sensor system model shown in Fig 1. We assume there are an unknown number of N particles in a given volume moving according to independent Brownian motions. We assume there are K disjoint sensors in the area. We denote the readout of the i'th sensor at time t by $n_i(t), 1 \leq i \leq K$, which is the number of particles in an effective volume around the i'th sensor.

We assume there are $n_0(t)$ particles in the volume not present in any of the sensing volumes such that for each time t we have

$$n_0(t) + \sum_{i=1}^{K} n_i(t) = N.$$

Consider $\mathbf{n}(t) = (n_0, n_1, \ldots, n_K)(t)$ which is a vector of $K+1$ elements. Time evolution of state vector $\mathbf{n}$ can be modeled as a Markov process $(\mathbf{n}(t), t \in \mathcal{R})$ with state space

$$\mathcal{S}_N = \left\{ \mathbf{n} \in \mathcal{Z}_+^{K+1} : \sum_{i=0}^{K} n_i = N \right\}.$$

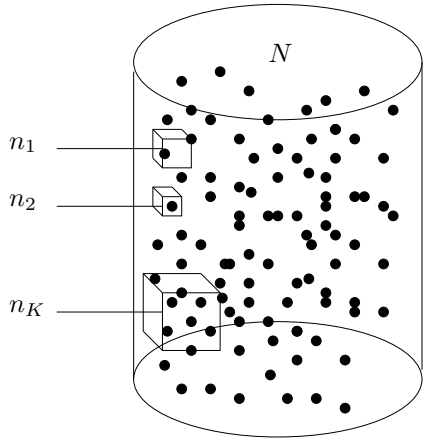

Fig. 1. N particles randomly moving in a volume with K sensors. Each sensor has an effective volume and each particle randomly moves into or out of the effective volume around each sensor.

Define transformed states

$$T_{0i}(\mathbf{n}) = (n_0 - 1, n_1, \ldots, n_{i-1}, n_i + 1, n_{i+1}, \ldots, n_K),$$
$$T_{i0}(\mathbf{n}) = (n_0 + 1, n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots, n_K).$$

For small enough δt, assume the probability that any single particle enters the i'th sensing volume is $p_i \delta t$ and the probability that any single particle inside the i'th sensing volume moves out is $q_i \delta t$. More precisely,

$$\mathrm{P}\{\mathbf{n}(t+\delta t) = T_{0i}(\mathbf{n}) | \mathbf{n}(t) = \mathbf{n}\} = p_i n_0 \delta t + o(\delta t),$$
$$\mathrm{P}\{\mathbf{n}(t+\delta t) = T_{i0}(\mathbf{n}) | \mathbf{n}(t) = \mathbf{n}\} = q_i n_i \delta t + o(\delta t),$$

where $\lim o(\delta t)/\delta t \to 0$ as $\delta t \to 0$. Now assume a Markov process with transition rates

$$q(\mathbf{n}, T_{0i}(\mathbf{n})) = p_i n_0,$$
$$q(\mathbf{n}, T_{i0}(\mathbf{n})) = q_i n_i,$$

with stationary equilibrium state probability distribution $\pi(\mathbf{n}|N)$. Distribution $\pi(\mathbf{n}|N)$ is then limiting, ergodic, and stationary, meaning that $\forall \mathbf{n} \in \mathcal{S}$,

$$\mathrm{P}\{\mathbf{n}(t) = \mathbf{n}|N\} \to \pi(\mathbf{n}|N), \text{ as } t \to \infty,$$

$$\frac{1}{T}\int_0^T I(\mathbf{n}(\tau) = \mathbf{n}|N) d\tau \to \pi(\mathbf{n}|N), w.p.1 \text{ as } T \to \infty,$$

$$\mathrm{P}\{\mathbf{n}(0) = \mathbf{n}|N\} = \pi(\mathbf{n}|N) \Rightarrow \mathrm{P}\{\mathbf{n}(t) = \mathbf{n}|N\} = \pi(\mathbf{n}|N),$$

$\forall t > 0$. As a result, the stationary equilibrium distribution $\pi(\mathbf{n}|N)$ is a sufficient first order conditional statistic of the process.

Next section of the paper is dedicated to finding this probability distribution function. As a result of the analysis in the next section, a posterior probability density of the number of particles $\pi(N|\mathbf{n})$ can be computed using a priori probability density function $\pi(\mathbf{n}|N)$.

3. ANALYSIS

The process as modeled in Section II is a form of a closed migration process where transitions are only allowed between free volume, marked by 0, and sensor volumes, $i = 1, \ldots, K$. The closed migration process naturally would arise from independent motion of particles [3]. No transition is allowed between sensor volumes in the model we consider in this paper. Although the analysis presented in this section can be slightly modified to consider this unrealistic case, we neglect this case here, since it does not provide any extra intuition.

First let's assume $N = 1$, i.e., a single particle performs a random walk with equilibrium distribution α_i where $\sum_{i=1}^K \alpha_i = 1$ and α_is are the solution of the following set of linear equilibrium equations known as full balance and partial balance equations:

$$\begin{aligned} \alpha_0 \sum_{i=1}^K p_i &= \sum_{i=1}^K \alpha_i q_i, \\ \alpha_j q_j &= \alpha_0 p_j, \quad j = 1, \ldots, K. \end{aligned}$$

Solving this set of linear equations give

$$\alpha_0 = \frac{1}{1 + \sum_{i=1}^K \frac{p_i}{q_i}}, \alpha_j = \frac{p_j}{q_j}\alpha_0.$$

Now we are ready to solve the balance equations for the general case where there are a given number of N particles in the volume. The full balance equation can be written as

$$\pi(\mathbf{n}|N) \sum_{i=1}^K [q(\mathbf{n}, T_{0i}(\mathbf{n})) + q(\mathbf{n}, T_{i0}(\mathbf{n}))] =$$
$$\sum_{i=1}^K [\pi(T_{0i}(\mathbf{n})|N) q(T_{0i}(\mathbf{n}), \mathbf{n}) + \pi(T_{i0}(\mathbf{n})|N) q(T_{i0}(\mathbf{n}), \mathbf{n})],$$

and the partial balance equations are

$$\begin{aligned} \pi(\mathbf{n}|N) \sum_{i=0}^K q(\mathbf{n}, T_{0i}(\mathbf{n})) &= \sum_{i=1}^K \pi(T_{0i}(\mathbf{n})|N) q(T_{0i}(\mathbf{n}), \mathbf{n}), \\ \pi(\mathbf{n}|N) q(\mathbf{n}, T_{j0}(\mathbf{n})) &= \pi(T_{j0}(\mathbf{n})|N) q(T_{j0}(\mathbf{n}), \mathbf{n}). \end{aligned}$$

Noting that $q(\mathbf{n}, T_{j0}(\mathbf{n})) = q_j n_j$, $q(\mathbf{n}, T_{0j}(\mathbf{n})) = p_j n_0$, $q(T_{j0}(\mathbf{n}), \mathbf{n}) = p_j(n_0 + 1)$ and $q(T_{0j}(\mathbf{n}), \mathbf{n}) = q_j(n_j + 1)$ it can be shown that

$$\frac{\pi(T_{j0}(\mathbf{n})|N)}{\pi(\mathbf{n}|N)} = \frac{n_j}{\alpha_j}\frac{\alpha_0}{n_0 + 1} = \frac{n_j}{n_0 + 1}\frac{q_j}{p_j},$$
$$\frac{\pi(T_{0j}(\mathbf{n})|N)}{\pi(\mathbf{n}|N)} = \frac{n_0}{\alpha_0}\frac{\alpha_j}{n_j + 1} = \frac{n_0}{n_j + 1}\frac{p_j}{q_j},$$

and hence the equilibrium distribution can be written as

$$\pi(\mathbf{n}|N) = B_N \prod_{i=0}^K \frac{\alpha_i^{n_i}}{n_i!}, \quad \forall \mathbf{n} \in \mathcal{S}_N,$$

where constant B_N is chosen such that $\sum_{\mathbf{n}\in\mathcal{S}_N}\pi(\mathbf{n}|N)=1$. Alternatively, we can write,

$$\pi(\mathbf{n}|N)=B_N\frac{\alpha_0^N}{n_0!}\prod_{i=1}^{K}\frac{(p_i/q_i)^{n_i}}{n_i!},\quad \forall\mathbf{n}\in\mathcal{S}_N.$$

For the rest of this paper, we assume the sensors are statistically equivalent, i.e., $p_1=p_2=\cdots=p_K=p$ and $q_1=q_2=\cdots=q_K=q$. With this simplifying assumption, we can write,

$$\pi(\mathbf{n}|N)=B_N\frac{\alpha_0^N}{n_0!}\left(\frac{p}{q}\right)^{N-n_0}\prod_{i=1}^{K}\frac{1}{n_i!},\quad \forall\mathbf{n}\in\mathcal{S}_N,$$

or equivalently

$$\pi(n_1,\ldots,n_K|N)=B_N\frac{\alpha_0^N\left(\frac{p}{q}\right)^{\sum_{i=1}^K n_i}}{\left(N-\sum_{i=1}^K n_i\right)!}\prod_{i=1}^{K}\frac{1}{n_i!}.$$

The above equation can be used to provide the best estimate of N given sensor observations $n_1,\ldots,n_K$. If N is uniformly distributed or if we have no prior knowledge of its distribution, maximum a posteriori estimate of N given sensor measurements is equal to the maximum likelihood estimate, hence,

$$\hat{N}=\arg\max_N\pi(N|n_1,\ldots,n_K)=\arg\max_N\pi(n_1,\ldots,n_K|N).$$

Therefore, the best estimate of the number of particles based on the set of observations $(n_1,\ldots,n_K)$ is given by

$$\hat{N}=\arg\max_N B_N\left(\frac{q}{q+Kp}\right)^N\frac{1}{\left(N-\sum_{i=1}^K n_i\right)!}.$$

In particular if we have only one sensor, i.e., $K=1$, we have

$$\pi(n|N)=B_N\frac{\left(\frac{q}{p+q}\right)^N}{n!(N-n)!}\left(\frac{p}{q}\right)^n.$$

Solving the equation $\sum_{i=1}^N\pi(i|N)=1$, we obtain $B_N=N!$. Therefore,

$$\hat{N}=\arg\max_N\frac{N!}{(N-n)!}\left(\frac{q}{p+q}\right)^N\approx n\left(1+\frac{q}{p}\right),$$

which is an un-biased estimator of N, i.e., $E(\hat{N}|N)=N$. Moreover, it can be shown that

$$E[(\hat{N}-N)^2|N]=\frac{Nq}{p}.$$

The conditional quantum-limit signal to noise ratio can be defined as [4]

$$\mathrm{SNR}_{QL}=\frac{N^2}{E[(\hat{N}-N)^2|N]}=\frac{Np}{q}.$$

This result is quite intuitive, in the sense that sensors with larger effective volume have larger p and smaller q and hence result in a better SNR. Moreover, SNR linearly grows with the number of particles, the better SNR can be obtained with a larger number of particles N.

It can also be shown that if all K sensors are statistically equivalent, $B_N=N!$ and

$$\hat{N}\approx\left(1+\frac{q}{p}\right)n.$$

It can also be shown that for the case of statistically equivalent sensors, $B_N=N!$, and

$$\hat{N}\approx\left(1+\frac{q}{Kp}\right)\sum_{i=1}^{K}n_i,$$

i.e., the individual sensor readouts are not important and as far as the estimation is concerned, the sensors together act like a single larger sensor reading the aggregate number of particles in all sensor volumes.

4. AUTO-CORRELATION AND TIME DEPENDENCE

In this section, we assume there is only one sensor (or one large equivalent sensor) with readout $n(t)$ over time and find the autocorrelation function of the random process $n(t)$. This analysis is of particular interest since it results in the power spectral density of the output signal which can be measured in practical experiments.

We assume time interval δt is chosen small enough such that $(1-p\delta t)(1-q\delta t)\approx 1-(p+q)\delta t$, i.e., at each time interval δt, no more than one particle enters or exits a sensory area. We would like to find the autocorrelation function $R_n(\tau)$ where

$$\begin{aligned}R_n(\tau)&=E[n(t)n(t+\tau)|N]\\&=E[n(0)n(\tau)|N]\\&=E_{n(0)}\{E[n(\tau)|n(0)]n(0)|N\}.\end{aligned}$$

First, let's focus on $E[n(\tau)|n(0)]$. Assuming that $j=|\tau|/\delta t$, it can be shown that

$$\begin{aligned}T_j&=E[n(j\delta t)|n(0)=n]\\&=E_{n((j-1)\delta t)}\{E[n(j\delta t)|n((j-1)\delta t),n(0)=n]\}\\&=E_{n((j-1)\delta t)}\{E[n(j\delta t)|n((j-1)\delta t)]|n(0)=n\}\\&=pN\delta t+(1-(p+q)\delta t)T_{j-1}.\end{aligned}$$

Noting that $T_0=n$ and solving the recursive equation for T_j, we obtain

$$T_j=pN\frac{1-(1-(p+q)\delta t)^j}{p+q}+n(1-(p+q)\delta t)^j.$$

As $\delta t \to 0$, we can write the above expression in terms of the more interesting parameter $|\tau| = j\delta t$ as

$$T_j \approx pN\frac{1-e^{-(p+q)|\tau|}}{p+q} + ne^{-(p+q)|\tau|},$$

and as a result $R_n(\tau)$ can be simplified to

$$R_n(\tau) = \left(\frac{Np}{p+q}\right)^2 \left(1 + \frac{q}{Np}e^{-(p+q)|\tau|}\right).$$

Taking the Fourier transform of the above expression, we obtain the power spectral density of random process $n(t)$

$$S_n(\omega) = \left(\frac{Np}{p+q}\right)^2 \left(2\pi\delta(\omega) + \frac{q/Np}{(p+q)^2+\omega^2}\right).$$

This power spectral density has a Lorentzian profile and shows that the behavior of the system is very similar to the optical detection systems where inherent signal shot-noise is taken into account [4]. If the amplification processes adds a white noise with variance N_0, it can be directly added to the power spectral density (see Fig. 2).

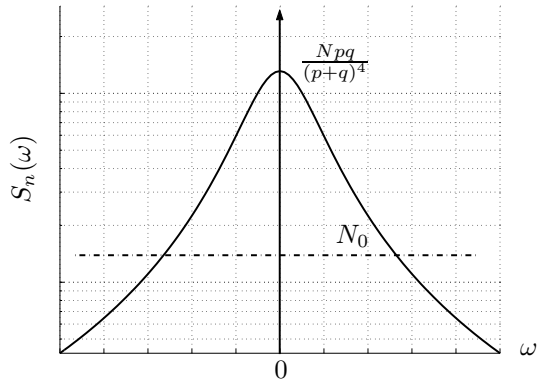

Fig. 2. Power Spectral Density (PSD) of the sensor output.

It can be seen from the figure above that in order to limit the effect of thermal noise added during the amplification process, the output signal can be filtered with a low-pass filter without considerably deteriorating the counting process.

Fig. 3 shows the power spectral density results of a simulation for $N = 1000$ and $N = 6000$ particles randomly walking on a $100 \times 100 \times 100$ unit box. The sensor volume is modeled by a $10 \times 10 \times 10$ cube in one side of the volume and the number of particles in the sensor volume has been counted an registered in a time interval of 256 time units. It can be seen that although the shot-noise level is higher for the case of $N = 6000$, as we expect, the signal to noise ratio is also higher.

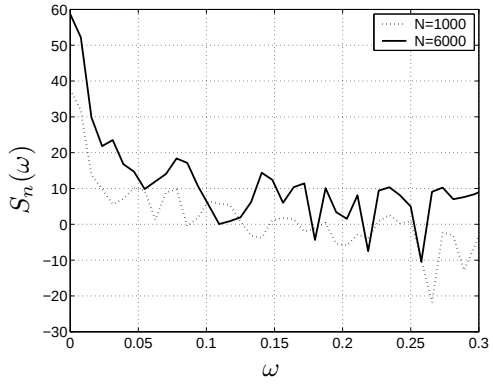

Fig. 3. Simulation results for power spectral density of particles within the sensor volume, $N = 1000$ and $N = 6000$.

5. CONCLUSIONS

The goal of biochemical sensors is to quantify the number of specific molecules present in the sample. In these systems, each individual sensing element counts the number of particles within its proximity. The readout of the system, originating from the number of particles present in the sensing volume becomes a random process as a result of random motion of particles. Here we have shown that the fluctuation of the readout signal is a function of the probability of particles moving in and out of the sensing volume. A Markov process model has been applied to derive the steady state power spectral density of the fluctuation, which demonstrates that the output signal PSD has in fact a Lorentzian profile. We have also provided the maximum likelihood estimate of the number of particles based on our observations which relates the best estimate to the physical parameters of the sensor. The estimation techniques presented here can be applied to various biochemical systems. In addition the model can potentially give insight into the design of high performance and low noise biochemical detectors.

6. REFERENCES

[1] M. J. Madou, R. Cubicciotti, "Scaling Issues on Chemical and Biological Sensors," *Proceedings of the IEEE*, vol. 91, pp. 830-838, June 2003.

[2] A. Hassibi, T. H. Lee, R. Navid, R. W. Dutton, S. Zahedi, "Effects of Scaling on the SNR and Speed," unpublished

[3] F. P. Kelly, *Reversibility and Stochastic Networks*, John-Wiley, 1979.

[4] R. M. Gagliardi, S. Karp, *Optical Communications*, 2nd edition, John-Wiley, 1995.