On the Capacity of Diffusion-Based Molecular Communications with SiNW FET-Based Receiver

Murat Kuscu, Student Member, IEEE and Ozgur B. Akan, Fellow, IEEE

Abstract—Molecular communication (MC) is a bio-inspired communication method based on the exchange of molecules for information transfer among nanoscale devices. Although MC has been extensively studied from various aspects, limitations imposed by the physical design of transceivers have been largely neglected in the literature. Recently, we have proposed a nanobioelectronic MC receiver architecture based on the nanoscale field effect transistor-based biosensor (bioFET) technology, providing noninvasive and sensitive molecular detection at nanoscale while producing electrical signals at the output. In this paper, we derive analytical closed-form expressions for the capacity and capacity-achieving input distribution for a memoryless MC channel with a silicon nanowire (SiNW) FET-based MC receiver. The resulting expressions could be used to optimize the information flow in MC systems equipped with nanobioelectronic receivers.

I. INTRODUCTION

Molecular communication (MC) is a promising means of realizing nanonetworks and the Internet of nano-bio things (IoBNT), which could find breakthrough applications in several domains, such as continuous health monitoring and smart drug delivery [1], [2]. This bio-inspired method is mainly based on the use of molecules to encode, transmit and receive information.

MC has been studied from several perspectives, including that of information theory [3], [4]. However, except for a few studies, the design of transceivers and corresponding physical limitations have been largely neglected in the literature. There exist mainly two approaches to design MC transceivers. The first approach is to use synthetic biology techniques for designing artificial cells capable of transmitting and receiving molecular messages [5]. The second one is based on nanobioelectronic architectures which exploit the extraordinary properties of novel nanomaterials, such as nanowires and graphene, with a bio-inspired approach. In our recent review of design options for the MC receiver, we justify that nanobioelectronic designs are feasible and more convenient in most cases [6]. Electrical output of a nanobioelectronic receiver provides the advantage of high-speed information processing, making the design more suitable for the complex communication schemes and protocols developed for MC [4].

In [7], we develop a communication theoretical model for a SiNW FET-based MC receiver and obtain the noise statistics. This nanobioelectronic design, shown in Fig. 1, is based on SiNW FET biosensors (bioFETs), which operate in a similar way to the conventional FETs. A SiNW bioFET consists of a recognition layer of surface receptors capable of binding the target ligands and a SiNW transducer channel which transduces the intrinsic charges of bound ligands into an electrical current flowing between the source and drain electrodes. In this paper, we develop an information theoretical model for a memoryless MC system equipped with the proposed SiNW FET-based MC receiver. We derive analytical closed-form expressions for the MC capacity and capacity-achieving input distribution, following the methodology of [8]. The resulting expressions incorporate the effect of transmitter, channel and receiver parameters, and could be used for optimizing the information flow in MC systems including nanobioelectronic receivers.

The remainder of the paper is organized as follows. In Section II, we present the existing model of MC system with a SiNW FET-based receiver. The capacity is derived in Section III, and the results are provided in Section IV. Finally, the concluding remarks are given in Section V.

II. MC SYSTEM MODEL

We consider a time-slotted MC system between a single transmitter nanomachine (TX) and a single receiver nanomachine (RX). TX and RX are assumed to be perfectly synchronized with each other in terms of time. Information is encoded into the number of molecules, i.e., ligands, released by the TX. The channel is assumed to be memoryless. This can be practically realized with a low symbol rate or using auxiliary enzymes degrading the residual molecules that cause inter-symbol interference (ISI).

The channel is assumed to be a free diffusion channel, where the trajectories of individual ligands are independent of each other. Thus, the ligand concentration as a function of time and space can be expressed by the Fick’s equation

$$\frac{\partial \rho (\vec{r}, t)}{\partial t} = D \nabla^2 \rho (\vec{r}, t), \quad (1)$$

where $\rho (\vec{r}, t)$ is the ligand concentration at location $\vec{r}$ and time $t$, and $D$ is the diffusion coefficient of the ligands. TX is assumed to be a point source located at the origin of 3D Cartesian coordinate system, i.e., $r_T = (0, 0, 0)$. It releases $N_{tx}$ number of molecules at the very beginning of a time slot. Thus, we can express the initial condition as $\rho (\vec{r}, 0) = N_{tx} \delta (\vec{r})$. Assuming that the distance $d$ between...
the TX and RX surface is much higher than the largest dimension of the RX surface, we can consider that all of the surface receptors are at equal distance from TX, and thus, are exposed to the same ligand concentration, which can be obtained by solving the Fick’s equation at the RX location with the initial condition imposed by the TX:

\[
\rho(d, t) = \frac{N_{tx}}{(4\pi D t)^{3/2}} \exp\left(-\frac{d^2}{4Dt}\right). \tag{2}
\]

We assume that RX samples the occupation states of the receptors when the ligand concentration at the receiver location is at its maximum, which is the case when \( t = \frac{d^2}{6D} \) [9]. The corresponding maximum ligand concentration at the RX location is then given by

\[
\rho_R = \left(\frac{3}{2\pi}\right)^{3/2} \frac{N_{tx}}{d^3} = \alpha_{ch} \times N_{tx}, \tag{3}
\]

where we define \( \alpha_{ch} = (3/2\pi e)^{3/2}d^{-3} \) as the channel attenuation parameter.

The interaction of ligands with the surface receptors is governed by the ligand-receptor binding kinetics. Assuming steady-state conditions, the probability of finding a single receptor at the bound state is given by

\[
P_b = \frac{k_1 \rho_R}{k_1 \rho_R + k_{-1}} = \frac{k_1 \alpha_{ch} N_{tx}}{k_1 \alpha_{ch} N_{tx} + k_{-1}}, \tag{4}
\]

where \( k_{-1} \) is the unbinding rate constant for the receptor-ligand pair; and \( k_1 \) is the binding rate constant in \( m^3 s^{-1} \). Generalizing the results to the case with \( N_r \) receptors, and assuming that all of the receptors observe independent concentrations, the distribution of number of bound receptors becomes binomial [10]:

\[
P(N_b = k) = \binom{N_r}{k} P_b^k (1 - P_b)^{N_r - k}, \tag{5}
\]

with mean and variance given by

\[
\mu_{N_b} = P_b N_r, \quad \sigma_{N_b}^2 = P_b(1 - P_b) N_r. \tag{6}
\]

Surface potential created on the NW-electrolyte interface as a result of binding of a single ligand is given as

\[
\psi_L = (q_{eff} \times N_e^-)/C_{eq}, \tag{7}
\]

where \( N_e^- \) is the number of free electrons per ligand molecule, and \( C_{eq} \) is the equivalent capacitance of the transducer. \( q_{eff} \) is the mean effective charge of a free ligand, which is degraded as the distance between the ligand electron and the transducer increases due to the Debye screening. The relation is given by \( q_{eff} = q \times \exp(-r/\lambda_D) \), where \( q \) is the elementary charge, and \( r \) is the average distance of ligand electrons in the bound state to the transducer’s surface [11], which is assumed to be equal to the average receptor length, i.e., \( r = l_{SR} \). The Debye length, \( \lambda_D \), quantizes the ionic strength of the medium according to

\[
\lambda_D = \sqrt{\frac{\epsilon_M k_B T}{2NAq^2c_{ion}}}, \tag{8}
\]

where \( \epsilon_M \) is the dielectric permittivity of the medium, \( k_B \) is the Boltzmann’s constant, \( T \) is the temperature, \( NA \) is Avogadro’s number, and \( c_{ion} \) is the ionic concentration of the medium [11]. The transducer capacitance is given by

\[
C_{eq} = \left(\frac{1}{C_{ox}} + \frac{1}{C_{nw}}\right)^{-1} + C_{dl}, \tag{9}
\]

where \( C_{dl} \) is the diffusion layer capacitance resulting from the double layer created by the medium counterions accumulated at the interface between oxide layer, i.e., SiO2 layer, and the electrolyte medium; \( C_{ox} \) is the capacitance of the oxide layer; and \( C_{nw} \) is the SiNW capacitance, which is again a double layer capacitance caused by the accumulation of carriers to the SiO2/SiNW interface [12], [13].

For a nanowire-on-insulator (NWoI) configuration of RX, we can consider the SiNW as a hemicylinder with an oxide layer of thickness \( t_{ox} \) covering the surface, and a diffusion layer of thickness \( \lambda_D \) covering the oxide layer [13]. Accordingly, the diffusion layer capacitance can be given by \( C_{dl} = (\epsilon_M/\lambda_D)w_Rl_R \), where \( w_R = \pi r_R \) is the effective width of SiNW with radius \( r_R \), and \( l_R \) is the SiNW length. The oxide capacitance is given by \( C_{ox} = C_{ox}^s w_Rl_R = (\epsilon_{ox}/l_{ox})w_Rl_R \), where \( \epsilon_{ox} \) is the permittivity and the thickness of the oxide layer, respectively. We will use \( C_{ox}^s \) to denote the oxide capacitance per area. For high values of hole density, e.g., \( p \sim 10^{18} \) cm\(^{-3} \), corresponding to the linear operation regime of the FET, the \( C_{nw} \) is obtained as \( C_{nw} = (\epsilon_S/\lambda_{nw})w_Rl_R \) [12], where \( \epsilon_S \) is the dielectric permittivity of SiNW, and \( \lambda_{nw} \) is the thickness of the double layer created in the inner surface of the NW, which is given by \( \lambda_{nw} = \sqrt{\epsilon_S k_B T/pq^2} \).

The potential induced at the SiNW/oxide layer interface is reflected into the output current of the RX. We assume that the p-FET is operated in the linear (Ohmic) region; thus, its transconductance is expressed by \( g_{FET} = \mu_p C_{ox}^s V_{SD}w_R/l_R \), where \( V_{SD} = V_{DS} \) is the source to drain voltage of the SiNW FET. Given the transconductance, the output current, \( I_{ox} \), resulting from the surface potential generated by the bound ligands is given as \( I_{ox} = \)}
where $V_{carriers}$ [14]. The PSD of the flatband-voltage noise scattering coefficient, and $\alpha$ is the threshold voltage of the SiNW FET. $\alpha_C$ is the Coulomb scattering coefficient, and $\mu_C$ is the mobility of the hole carriers [14]. The PSD of the flatband-voltage noise $S_{V,FB}$ is given by

$$S_{V,FB}(f) = S_{V,FB}^0(1 + \alpha_C\mu_C|V_{SG} - |V_{TH}|)^2,$$

where $V_{SG} = -V_{GS}$ is source to gate voltage, and $V_T$ is the oxide trap density, i.e., impurity concentration, of the SiNW channel [14]. At frequencies lower than $f_L$, it can be shown that the white noise behavior, i.e., $I_{V,FB}(f)$ for $|f| < f_L$. The reason is related to the low frequency cutoff set by the observation time $T_{obs}$ [15]. Considering that the received molecular signals are at the baseband, to be able to calculate the total noise power, we assume one-year operation time, i.e., $\approx 3 \times 10^7$ s, for the antenna such that the low cutoff frequency is $f_L = 1/T_{obs} \approx 1/\pi \times 10^{-7}$ Hz. At frequencies lower than $f_L$, the noise is assumed to show the white noise behavior, i.e., $S_{V,FB}^0(f) = S_{V,FB}(f_L)$ for $|f| < f_L$.

$$\sigma_{V,FB}^2 = \int_{-\infty}^{\infty} S_{V,FB}(f) df.$$

We can expect that a significant number of receptors, e.g., $> 1000$, are tethered to the top surface of a SiNW channel. For large number of surface receptors, the binomial distribution given in (5) can be approximated as Gaussian, i.e., $N_b \sim \mathcal{N}(\mu_{N_b},\sigma_{N_b}^2)$. 1/f noise is resulting from the bias current flowing through SiNW channel, thus, it is independent of the binding noise. It can be approximated to follow a Gaussian distribution [16]. Therefore, the overall noise process effective on the output current becomes a stationary Gaussian process with the variance

$$\sigma_{I,rx}^2 = \sigma_{V,FB}^2 + \sigma_{N_b}^2 g_{FET}^2 \psi_L N_b.$$

III. CAPACITY

We consider the overall system as a continuous-input continuous-output memoryless channel. Although information is encoded into discrete number of molecules, we can assume that the input is continuous given that it takes values from a very large input range. Then, the mutual information between the transmitted signal $N_{tx}$ and the received signal $I_{tx}$ can be written as

$$I(N_{tx};I_{tx}) = \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \log_2 f_{I_{tx}|N_{tx}}(y|x) dy dx$$

$$- \int_{-\infty}^{\infty} f_{N_{tx}}(x) \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \log_2 f_{I_{tx}|N_{tx}}(y|x) dy dx$$

where $N_{tx}^\text{min}$ and $N_{tx}^\text{max}$ denote the minimum and maximum number of molecules that TX can transmit. The input probability distribution satisfies the condition $\int_{N_{tx}^\text{min}}^{N_{tx}^\text{max}} f_{N_{tx}}(x) dx = 1$. Based on the discussion at the end of Section II, overall channel transition pdf $f_{I_{tx}|N_{tx}}$ can be given by

$$f_{I_{tx}|N_{tx}}(y|x) = \frac{1}{\sqrt{2\pi \sigma_{I,tx}^2}} e^{-\frac{(y-\mu_{I,tx})^2}{2\sigma_{I,tx}^2}}$$

where $\mu_{I,tx}$ and $\sigma_{I,tx}^2$ are functions of $N_{tx} = x$, as given in (10) and (14). Given the Gaussian pdf, the first integral on the RHS of (15) is obtained as [8]

$$\int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \log_2 f_{I_{tx}|N_{tx}}(y|x) dy = -\frac{1}{2} \log_2(2\pi e \sigma_{I,tx}^2).$$

(17)

We approximate the second integral on the RHS of (15) using Taylor series expansion of $\log_2 p(\mu_{I,tx})$ up to the second order [8]:

$$\int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \log_2 f_{I_{tx}|N_{tx}}(y|x) dy$$

$$= \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \log_2 f_{I_{tx}|N_{tx}}(y|x) dy$$

$$+ \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \frac{\partial \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y} \bigg|_{\mu_{I,tx}} dy$$

$$+ \frac{1}{2} \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \frac{\partial^2 \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y^2} \bigg|_{\mu_{I,tx}} dy.$$

The first integral in (18) becomes

$$\int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \log_2 f_{I_{tx}|N_{tx}}(y|x) dy$$

$$= \log_2 f_{I_{tx}|N_{tx}}(y|x) \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) dy = \log_2 f_{I_{tx}|N_{tx}}(y|x).$$

(19)

and the second integral translates into

$$\int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \frac{\partial \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y} \bigg|_{\mu_{I,tx}} dy$$

$$= \frac{\partial \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y} \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) dy$$

$$= 0.$$  

(20)

The third integral is simplified as follows

$$\int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) \frac{1}{2} \frac{\partial^2 \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y^2} \bigg|_{\mu_{I,tx}} dy$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} \frac{\partial^2 \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y^2} \int_{-\infty}^{\infty} f_{I_{tx}|N_{tx}}(y|x) dy$$

$$= \frac{1}{2} \int_{-\infty}^{\infty} \frac{\partial^2 \log_2 f_{I_{tx}|N_{tx}}(y|x)}{\partial y^2} \bigg|_{\mu_{I,tx}} dy \approx 0.$$  

(21)
The approximation in (21) follows from the fact that the variance is very small compared to the mean for the high values of the number of surface receptors [7], [8]. Using (19)-(21) we can approximate the second integral on the RHS of (15) as follows

$$\int_{-\infty}^{\infty} f_{I(x)}(y|x) \log_2 f_{I(x)}(y) dy \approx \log_2 f_{\mu_{I(x)}}. \quad (22)$$

Since $$\mu_{I(x)}$$ is a deterministic and single valued function of $$N_{I(x)}$$, $$f_{\mu_{I(x)}}$$ can be written in terms of the input distribution $$f_{N_{I(x)}}(x)$$ as follows

$$f_{\mu_{I(x)}} = f_{N_{I(x)}}(x) \frac{(\alpha_{ch}k_1 + k_{-1})^2}{N_rM}, \quad (23)$$

where $$M = \alpha_{ch}k_1k_{-1}g_{\text{FET}}\psi_L$$, and the mutual information becomes

$$I(N_{I(x)}; I_{tx}) = - \int_{N_{I(x)}^{\min}}^{N_{I(x)}^{\max}} f_{N_{I(x)}}(x) \left( \log_2 \frac{\sqrt{2\pi e} \sigma_{I_{tx}}^2}{N_rM} + \log_2 f_{N_{I(x)}}(x) + \log_2 \left( \frac{(\alpha_{ch}k_1 + k_{-1})^2}{N_rM} \right) \right) dx. \quad (24)$$

To find the capacity, the mutual information should be maximized over all input distributions satisfying $$f_{N_{I(x)}^{\max}}f_{N_{I(x)}}(x) dx = 1$$. We use the method of Lagrange multipliers defining the Lagrange function as

$$I_L = I(N_{I(x)}; I_{tx}) - \lambda \left( \int_{N_{I(x)}^{\min}}^{N_{I(x)}^{\max}} f_{N_{I(x)}}(x) dx - 1 \right), \quad (25)$$

where $$\lambda$$ is the Lagrange multiplier. The input distribution $$f_{N_{I(x)}^*}$$ maximizing $$I_L$$ is then found by functional derivative

$$\frac{\partial I_L}{\partial f_{N_{I(x)}}(x)} |_{f_{N_{I(x)}} = f_{N_{I(x)}^*}} = 0. \quad (26)$$

### Table I

<table>
<thead>
<tr>
<th>Default Values of Simulation Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max number of ligands TN transmits ($N_{I(x)}^{\max}$)</td>
</tr>
<tr>
<td>Min number of ligands TN transmits ($N_{I(x)}^{\min}$)</td>
</tr>
<tr>
<td>Transmitter-receiver distance (d)</td>
</tr>
<tr>
<td>Binding rate ($k_1$)</td>
</tr>
<tr>
<td>Unbinding rate ($k_{-1}$)</td>
</tr>
<tr>
<td>Average number of electrons in a ligand ($N_e$)</td>
</tr>
<tr>
<td>SiNW radius ($r_{SiNW}$)</td>
</tr>
<tr>
<td>Concentration of receptors on the surface ($\rho_{G2}$)</td>
</tr>
<tr>
<td>Length of a surface receptor ($l_{G2}$)</td>
</tr>
<tr>
<td>Temperature (T)</td>
</tr>
<tr>
<td>Relative permittivity of oxide layer ($\varepsilon_{ox}/\varepsilon_0$)</td>
</tr>
<tr>
<td>Relative permittivity of SiNW ($\varepsilon_{SiNW}/\varepsilon_0$)</td>
</tr>
<tr>
<td>Relative permittivity of medium ($\varepsilon_{medium}/\varepsilon_0$)</td>
</tr>
<tr>
<td>Ionic strength of electrolyte medium ($\mu_{ion}$)</td>
</tr>
<tr>
<td>Source-drain voltage ($V_{SD}$)</td>
</tr>
<tr>
<td>Source-gate voltage ($V_{SG}$)</td>
</tr>
<tr>
<td>Threshold voltage ($V_{TH}$)</td>
</tr>
<tr>
<td>Hole density in SiNW ($p$)</td>
</tr>
<tr>
<td>Tunneling distance ($l$)</td>
</tr>
<tr>
<td>Thickness of oxide layer ($t_{ox}$)</td>
</tr>
<tr>
<td>Oxide trap density ($\rho_{ox}$)</td>
</tr>
<tr>
<td>Effective mobility of hole carriers ($\mu_{ph}$)</td>
</tr>
<tr>
<td>Coulomb scattering coefficient ($\alpha_s$)</td>
</tr>
</tbody>
</table>

Substituting $$f_{N_{I(x)}^*}(x)$$ into (24), we obtain the capacity, in bits per use, as follows

$$C = \frac{1}{2} \log_2 \frac{N_r}{2\pi e} + \log_2 \left[ \sin^{-1} \left( \frac{L N_{I(x)}^{\max} - k_D/\alpha_{ch}}{L N_{I(x)}^{\min} - k_D/\alpha_{ch}} \right) \right] \left( N_{I(x)}^{max} - k_D/\alpha_{ch} \right). \quad (31)$$

### IV. Results

In this section, we investigate the characteristics of the capacity-achieving input distribution and the effect of main system parameters on the capacity. The default values for the controllable parameters used in the analyses are listed in Table I. The overall setting is the same as that used in [7].

The capacity-achieving input distribution $$f_{N_{I(x)}^*}$$ is demonstrated in Fig. 2. As expected, the distribution is more focused around the minimum number of molecules $$N_{I(x)}^{\min}$$.
TX transmits, which results in a less amount of noise in the receiver’s electrical output.

Next, we analyze the relation between the upper limit $N_{tx}^{max}$ of transmitted molecules and the capacity. As is seen in Fig. 3(a), the capacity significantly increases with the increasing $N_{tx}^{max}$ until it begins to saturate around 5 bits/use. This is originating from the fact that the receiver has a finite number of molecules; thus, it becomes saturated when $\rho R \gg K_D$, resulting in a reduced receiver sensitivity. Therefore, transmitting more molecules to increase the ligand concentration at the receiver location is not favorable when the receiver is saturated.

The second analysis is carried out for varying TX-RX distance. As demonstrated in Fig. 3(b), the capacity is maximum at intermediate distances, e.g., $d \approx 150 \mu$m, and begins to decrease when the distance is below or above this range. For a given $N_{tx}^{max}$, as the distance gets smaller, the receiver begins to operate near saturation because $\rho R$ significantly increases when the transmitter and receiver are close to each other. This is reflected to the output current, and results in a decrease in the sensitivity of the receiver so that it cannot discriminate different levels of ligand concentration corresponding to different symbols. As a result, the capacity decreases. For large distances, the ligand concentration gets significantly attenuated in the channel, thus, the signal-to-noise ratio (SNR) decreases [7], resulting in a lower capacity.

The ionic strength of the fluidic medium also substantially affects the capacity, as is seen in Fig. 3(c). The Debye length $\lambda_D$ decreases with the increasing ionic concentration, implying more effective screening of ligand charges. This in turn reduces the SNR, and thus, the capacity. Physiological conditions generally imply ionic concentrations higher than 100 mol/l$^3$. To compensate the attenuation of capacity, surface receptors with lengths comparable to Debye length should be selected.

Lastly, we analyze the capacity for varying oxide trap density $N_{ot}$, which is proportional to the impurity of the SiNW. Trap density affects the carrier mobility, and increases the $1/f$ noise. The negative effect of increasing trap density on the capacity is evident from Fig. 3(d).

![Fig. 3](attachment:fig3.png)

V. CONCLUSION

In this study, we developed an information theoretical model for a memoryless MC system equipped with a SiNW FET-based MC receiver, and derived analytical closed-form expressions for the capacity and capacity-achieving input distribution. The provided expressions enable the analysis of the effect of nanobioelectronic receiver parameters on the MC system capacity, and reveal the optimization pathways that can be targeted to improve reliability of the overall communication system.

REFERENCES