Abstract—Molecular Communications (MC) is a bio-inspired communication technique that uses molecules as a means of information transfer among nanoscale devices. As discrete molecules, i.e., ligands, are the carriers of information in MC, the distinct ligand types co-existing in the channel can be considered to have unique carrier frequencies, governing their reaction dynamics with the ligand receptors of the nanomachines. Sensing the dynamic molecular spectrum, i.e., multiplexed detection of the concentration of multiple ligand types of similar properties co-existing in the MC channel, provides opportunities for eliminating molecular interference resulting from external sources in crowded physiological environments, and multi-user interference (MUI) in nanonetworks, and adopting cognitive radio techniques for medium access in MC. In this paper, we develop a practical and low-complexity spectrum sensing method that can simultaneously estimate the individual concentrations of multiple types of ligands with single type of receptors by exploiting the amount of time the receptors stay bound and unbound in ligand-receptor binding reaction. We analyze its performance in terms of normalized mean squared error (NMSE) for varying number of co-existing ligand types, number of samples, similarity between ligands, and ligand concentration distribution, and show that it is possible to estimate the concentration of ligands up to $10^{-2}$. Additionally, for the transduction of receptor unbound and bound time durations into intracellular molecular signals for further processing, we propose a synthetic receptor design based on modified kinetic proofreading (KPR) scheme. Lastly, we discuss the implementation challenges of the estimator in engineered bacteria with synthetic analog computation tools in living cells.

Index Terms—Molecular communication, receiver, ligand receptors, molecular spectrum, spectrum sensing, maximum-likelihood estimation, method of moments, kinetic proofreading, synthetic biology, multiplexed detection, molecular division multiplexing, molecular division medium access.

I. INTRODUCTION

Molecular Communications (MC) is a bio-inspired communication technique based on using molecules to encode, transmit and receive information \cite{1}. MC has gained increasing popularity in the last decade due to its potential to enable artificial nanonetworks, i.e., networks of nanomachines, and the Internet of Nano Things (IoNT), an emerging technology consisting in nanonetworks connected to the Internet and promising for unprecedented applications, e.g., intrabody continuous health monitoring for early diagnosis and treatment, smart drug delivery, and artificial organs \cite{2}, \cite{3}, \cite{4}. MC has been extensively studied from various theoretical aspects, e.g., detection, channel modeling, and modulation \cite{5}; however, there is still no practical implementation of an MC system at nanoscale. The key to realize nanoscale artificial MC networks is to close the gap between the theory and practice by addressing the peculiarities resulting from discrete nature of molecules, limited capabilities of nanoscale devices, and highly stochastic, nonlinear, and time-varying dynamics of the MC channels, which bring about new challenges fundamentally different from those we tackle in conventional electromagnetic (EM) wireless communications.

In practice, MC channels, especially in physiologically-relevant environments, can be crowded by many types of molecules that may have similar characteristics making their discrimination nontrivial. These molecules can be resulting from natural processes, e.g., intrabody cell signaling, that are generally not relevant to the MC application, leading to natural interference. They can also result from another MC system co-existing in the same medium leading to multi-user interference (MUI) \cite{6}. As the molecules are the carriers of information in MC, the type of molecules employed can be regarded as the carrier frequency in an analogy with EM communications. Therefore, the set of all ligand types that potentially co-exist in the channel, with the power intensities being the instantaneous concentration of corresponding ligand types constitutes the dynamic molecular spectrum of the channel.

The knowledge of the channel state information (CSI) in terms of dynamic molecular spectrum occupancy is crucial for developing reliable detection and modulation methods in the time-varying presence of interferer molecules of similar characteristics with the messenger molecules. It is also important for developing cognitive medium access (MA) techniques for MC nanonetworks making the most of the available molecular spectrum. Although channel estimation techniques are proposed for MC with passive and transparent observers in \cite{7}, \cite{8}, and spectrum sensing techniques for co-existing MC nanonetworks that utilize the same type of molecules is considered in \cite{9}, simultaneous estimation of concentration of different molecule types, which is enabled by the spectrum sensing technique developed in this paper, has not been considered in the MC literature before.

This study is focused on engineered bacteria-based biological MC transceivers with synthetic ligand receptors on their surface. The receptors constitute the interface between the exterior and interior of the bacteria, interact with the external ligands, i.e., molecules in the MC channel, based on ligand-receptor binding reaction, and transduce the binding
events into intracellular signals [10]. The binding rate of ligands to the receptors depend on the transport properties and the concentration of ligands as well as the activation energy of the binding reaction, whereas the unbinding rate reflects the affinity of the ligands with receptors, and can be considered specific to the particular ligand-receptor pair. In other words, the unbinding rate of a ligand can be regarded as its carrier frequency. Receptors, in practice, can provide specificity for the target ligands only to a finite extent, and other types of ligands can also bind the same receptors, though typically having lower affinities [11]. The correlation between the unbinding rate and the ligand type is the key property that is exploited in this paper to develop a molecular spectrum sensing technique.

MC with ligand receptors has been addressed from different aspects. Channel models are developed between point transmitters and reactive receivers with ligand receptors in [12], [13]. Detection techniques are proposed for concentration shift keying (CSK) modulated MC with ligand receptors in [14], based on sampling the instantaneous number of bound receptors. Recently, the continuous history of bound and unbound states has been suggested to bear more information over the external ligand concentration [15], [16]. In this direction, maximum a posteriori (MAP) detection methods are proposed for MC in [15], [17]. In our previous work [18], receptor unbound time duration statistics is shown to provide a larger dynamic input range for the detector to cope with saturation at high ligand concentrations resulting from intersymbol interference (ISI). Maximum likelihood (ML) estimation of the concentration of two different ligand types based on sampling the receptor bound time durations is studied in [11], [19], [20], [21]. These studies also argue practical implementations of the ML estimators exploiting kinetic proofreading (KPR) mechanism, which is an active cellular mechanism that increases specificity, suggested to exist in T-cell receptors that can sense very low concentrations of foreign agents with extreme specificity as part of the immune system [22]. Based on a similar approach, we have previously investigated the performance of ML detection with bound time durations in the presence of single type of interferers, and shown that it outperforms other MC detection schemes that use the samples of receptor unbound time or instantaneous receptor states [23]. However, none of the previous studies has considered the problem of sensing the concentration of more than two types of ligands at the same time.

In this paper, we develop a practical spectrum sensing method to concurrently estimate the concentrations of all ligand types co-existing in the channel. The proposed method consists of two estimators. The first one is an unbiased maximum-likelihood (ML) estimator of the total ligand concentration, which uses the amount of time the receptors stay unbound, as proposed in [16], [18]. This estimator exploits the fact that as the total ligand concentration increases, the receptors bind more frequently to the ligands, and thus, the unbound periods of the receptors get shorter. The second estimator is also unbiased, and estimates the concentration ratio of each ligand type from the amount of time the receptors stay bound to a ligand, using method of moments. This estimator is based on the fact that the receptors bind more strongly to the ligands of higher affinity, and thus, the duration of bound time periods are correlated with the type of ligands [11]. We also develop a more practical version of the concentration ratio estimator, which is biased, however, requires less number of computations. The product of the total concentration and the ratio estimators provides the required dynamic spectrum information, i.e., the instantaneous concentration of each ligand type. We evaluate the performance of the spectrum sensing technique in terms of normalized mean squared error (NMSE) averaged over all co-existing ligands, for varying number of ligand types in the mixture, varying number of samples, and similarity between the ligand types, and varying concentration distributions of different ligand types within the mixture.

The estimators should operate inside synthetic cells by making use of second messengers, i.e., intracellular signaling molecules, for arithmetic calculations. This requires the transduction of unbound and bound time durations into concentration of second messengers, which are then processed through analog computing. To this end, we propose a synthetic receptor design with a multitude of internal states, that utilizes a modified version of conventional KPR mechanism [22]. The proposed receptor is able to be activated by the intracellular activation signal at the start of a sampling period, and encode the observed unbound and bound time durations into the concentration of different types of second messengers.

The remainder of this paper is organized as follows. In Section [1] we discuss the opportunities of spectrum sensing in MC focusing on its potential in developing reliable detection methods, adaptive and multi-functional receivers, and medium access techniques. We present the mathematical framework of the proposed spectrum sensing method in Section [III]. The performance of the technique is evaluated in Section [IV]. We provide a practical discussion on implementation of the proposed method in Section [V]. Lastly, we conclude the paper in Section [VI] by discussing open research directions.

II. OPPORTUNITIES OF SPECTRUM SENSING IN MC

Exploiting the cross-talk between different types of ligands for spectrum sensing with single type of receptors is important for improving the adaptivity and reliability of the MC devices, increasing the capacity of the MC channels, and enabling the effective use of molecular spectrum for medium access without requiring substantial amount of additional computational resources and receptors. The proposed spectrum sensing method can prove effective especially towards the following directions in the MC research:

- Development of reliable detection methods for CSK modulated signals based on eliminating the interference of similar ligands released by external sources: Current studies focusing on CSK-based MC with ligand receptors assume that the receptors are ideal, such that they only react with the ligand type that carries the information [19]. However, in practice, the specificity of receptors is not perfect, and they can react with multiple types of molecules, though with different reaction rates, especially in physiologically relevant conditions. Eliminating the interference by sampling the instantaneous
receptor states is not viable, when the channel is time-varying, e.g., the concentration of interferer molecules change between signaling intervals. The spectrum sensing method proposed in this paper does not require a priori knowledge of the probability distribution of ligand concentrations; therefore, it can enable robust and reliable detection under time-varying conditions.

- Development of reliable detection methods for molecule shift keying (MoSK) and ratio shift keying (RSK) modulated MC signals: These modulation techniques rely on the transmission of multiple types of ligands. In MoSK, the information is encoded into the concentration of different ligand types, which are transmitted in separate signaling intervals. On the other hand, in RSK, the information is encoded into the ratio of concentrations of different ligand types transmitted at the same signaling interval. The current studies focusing on both modulation methods assume that there is an ideal receptor for each ligand type, and the cross-talk between different ligand-receptor pairs is neglected. However, this is not the case in practice, as the cross-talk between ligands always exists. The proposed spectrum sensing method can be employed to eliminate the cross-talk between different ligand-receptor pairs and increase the capacity of the channel. Additionally, with the use of the proposed method, both MoSK and RSK modulated MC signals can be accurately detected by utilizing only a single type of receptor. This can also enable the transmitter to increase the cardinality of the set of transmitted molecule types for boosting the channel capacity, without necessitating the deployment of extra receptors in the receiving side.

- Development of interference-free molecular division medium access (MDMA) techniques: MDMA is based on the idea of using different types of molecules in different MC channels co-existing in the same environment. In this way, multitude of MC network nodes can concurrently use the same medium for information transmission; however, the MUI cannot be avoided, as the specificity of the receptors is not infinite. Moreover, as in EM communication, the spectrum is a limited resource; however, this time, it is limited by the ligand types that an MC transmitter is capable of generating and transmitting, and an MC receiver is able to detect. In these circumstances, as similar to the cognitive radio techniques studied for conventional EM communications, the spectrum sensing technique can be opportunistically utilized to dynamically estimate the channel state in terms of available, i.e., unoccupied, frequencies prior to transmission, to avoid crowding the medium with a particular type of molecule and degrading the communication performance. On the receiver side, the spectrum sensing method can provide the receiver with the required adaptivity in detecting different types of molecules transmitted. This also enables the receiving node to simultaneously communicate multiple transmitting nodes through molecular division multiplexing by preventing cross-talk from affecting the reliability of the communication.

- Multi-functionality: Lastly, the proposed technique can enable multi-functional MC devices that can simultaneously perform the sensing of multiple types of molecules and communication using the same receptors. This can also help reduce the energy and molecular costs, and simplify the design of biological MC devices for MC nanosensor network applications.

III. SPECTRUM SENSING BASED ON LIGAND-RECEPTOR BINDING REACTION

In ligand-receptor binding reaction taking place on the surface of an MC device, e.g., engineered bacteria, receptors randomly bind ligands in their vicinity. A receptor can be either in the Bound (B) or Unbound (U) state, with exponential waiting times depending on the binding and unbinding rate of ligand-receptor pair. The state of a single receptor exposed to a concentration of single type of ligands is governed by the following two state Continuous-Time Markov Process (CTMP), i.e.,

\[
U \xleftrightarrow{k^-} B,
\]

where \(c_L(t)\) denotes the time-varying ligand concentration in the vicinity of receptors, \(k^+\) and \(k^-\) are the binding and unbinding rates for the ligand-receptor pair. Note that the transition rate from unbound to bound state is modulated by ligand concentration \(c_L(t)\). In diffusion-based MC, due to the low-pass characteristics of the diffusion channel, the bandwidth of the \(c(t)\) is typically low compared to the characteristic frequency of the binding reaction, i.e., \(f_B = c_L(t)k^+ + k^-\); thus, the ligand-receptor reaction is usually assumed to be at equilibrium with a stationary ligand concentration, which we denote simply by \(c_L\). In equilibrium conditions, the probability of observing a receptor in the bound state is given by

\[
p_B = \frac{c_L}{c_L + K_D},
\]

where \(K_D = k^-/k^+\) is the dissociation constant, which stands as a measure of the affinity between a ligand-receptor pair. If there are multiple receptors that are exposed to the same ligand concentration and not interacting with each other, the number of bound receptors becomes a binomial random variable with success probability of \(p_B\), and thus, its mean and variance are given by

\[
E[n_B] = \frac{c_L}{c_L + K_D}N_R,
\]

\[
\text{Var}[n_B] = p_B(1 - p_B)N_R,
\]

where \(N_R\) is the total number of receptors.

Sampling the number of bound receptors at a given time instant previously proved effective in inferring the concentration of ligands, when the receiver is away from saturation, i.e., when \(p_B < 1\). However, when there are \(M\) different ligand types in the environment, as shown in Fig. [1](a), which can bind the same receptor with different affinities, i.e., with different dissociation constants, the equilibrium bound state probability becomes

\[
p_B = \frac{\sum_{i=1}^{M} c_i/K_{D,i}}{1 + \sum_{i=1}^{M} c_i/K_{D,i}},
\]
and cannot be used to infer the individual ligand concentrations \( c_i \) due to the interchangeability of the summands \([11]\). The required insight into the individual ligand concentrations in the case of a mixture can be gained by examining the continuous history of binding and unbinding events over receptors, which is exemplified in Fig. \(1(b)\). The likelihood of observing a series of \( N \) binding-unbinding events at equilibrium can be written as

\[
p\left(\{\tau^B, \tau^U\}_N\right) = \frac{1}{Z} e^{-\sum_{i=1}^{N} \tau_i^B (\sum_{j=1}^{M} k_j^+ c_j e^{-k_j^- \tau_i^B})},
\]

where \( Z \) is the probability normalization factor, \( \tau_i^U \) and \( \tau_i^B \) are the \( i \)th observed unbound and bound time durations, respectively, \( k_j^+ \) and \( k_j^- \) are the binding and unbinding rate for the \( j \)th ligand type, respectively, and \( M \) is the number of ligand types co-existing in the channel \([11], [18]\). In the diffusion-limited case, i.e., where the reaction rates are much higher than the characteristic rate of diffusion, the binding rate can be given by \( k^+ = 4 Da \) for circular receptors, with \( D \) and \( a \) being the diffusion constant of molecules and the effective receptor size, respectively. Assuming that the ligands are of similar sizes, their diffusion coefficients \( D \) can be taken equal for all ligand types, and thus, the likelihood function \(6\) can be reduced to

\[
p\left(\{\tau^B, \tau^U\}_N\right) = \frac{1}{Z} e^{-T_U \times \text{c}_\text{tot} (k^+ \text{c}_\text{tot})} \prod_{i=1}^{N} \left( \sum_{j=1}^{M} \alpha_j k_j^- e^{-k_j^- \tau_i^B} \right),
\]

where \( T_U = \sum_{i=1}^{N} \tau_i^U \) is the total unbound time, \( \text{c}_\text{tot} = \sum_{i=1}^{N} c_i \) is the total ligand concentration in the vicinity of the receptors, and \( \alpha_i = c_i/\text{c}_\text{tot} \) is the concentration ratio of the \( i \)th ligand. The log-likelihood function for the set of unbound/bound time durations can be written as the sum of three terms, i.e.,

\[
L(\{\tau^B, \tau^U\}_N) = \ln p(\{\tau^B, \tau^U\}_N),
\]

\[
= L_0 + L_1(\text{c}_\text{tot}) + L_2(\alpha),
\]

where \( L_0 \) comprises the terms that do not depend on \( \text{c}_\text{tot} \) or \( \alpha \), while \( L_1 \) and \( L_2 \) are functions of total concentration \( \text{c}_\text{tot} \) and ligand concentration ratios \( \alpha \) denoted here as an \((M \times 1)\) vector, respectively, and given by

\[
L_1(\text{c}_\text{tot}) = N \ln(\text{c}_\text{tot}) - k^+ \text{c}_\text{tot} T_U,
\]

\[
L_2(\alpha) = \sum_{i=1}^{N} \ln \left( \sum_{j=1}^{M} \alpha_j k_j^- e^{-k_j^- \tau_i^B} \right).
\]

\( L_1(\text{c}_\text{tot}) \) tells us that the total unbound time \( T_U \) is informative of the total ligand concentration \( \text{c}_\text{tot} \), whereas \( L_2(\alpha) \) shows that the bound time durations \( \{\tau^B\}_N \) are informative of the ligand concentration ratios \( \alpha \). This suggests that by inferring the total ligand concentration and concentration ratios of ligand types separately by exploiting the total unbound time

\[
T_U \text{ and bound time durations } \{\tau^B\}_N \text{, respectively, we can estimate the individual concentrations of each ligand type.}
\]

An ML estimator for the total ligand concentration can be found by solving \( \partial L_1(\text{c}_\text{tot})/\partial \text{c}_\text{tot} = 0 \) for \( \text{c}_\text{tot} \) that maximizes the likelihood. The result is the estimator \( \hat{\text{c}}_\text{tot} = N/T_U \). Note that \( T_U \), as the sum of \( N \) independent exponential random variables, \( \tau^U \)‘s, is gamma distributed, making its reciprocal \( 1/T_U \) inverse gamma distributed with mean \((k^+ \text{c}_\text{tot})/(N - 1)\). Hence, the mean of the estimator becomes \(E[\hat{\text{c}}_\text{tot}] = (N/k^+) \times E[1/T_U] = c_{\text{tot}} \times N/(N - 1)\), rendering it biased unless \( N \) is very large. Therefore, we prefer here using its unbiased version, which is obtained by modifying only the numerator of the biased estimator \([18]\), i.e.,

\[
\hat{\text{c}}_\text{tot} = \frac{N - 1}{k^+ T_U}.
\]

Accordingly, the mean squared error (MSE) of this unbiased estimator is equal to its variance \([18]\), i.e.,

\[
\text{MSE}[\hat{\text{c}}_\text{tot}] = \text{Var}[\hat{\text{c}}_\text{tot}] = \frac{\text{c}^2_{\text{tot}}}{N - 2} \text{ for } N > 2.
\]

Estimation of the concentration ratios of co-existing ligand types can be performed in the same manner, by solving
\[ \frac{\partial \mathcal{L}_2(\alpha)}{\partial \alpha_i} = 0 \] for the \( i \)th ligand, i.e.,

\[ 0 = \sum_{i=1}^{N} \frac{k_i^- e^{-k_i^- T_i} }{\sum_{j=1}^{M} \alpha_j k_j^- e^{-k_j^- T_j} }, \quad (13) \]

which has no closed-form analytical solution; therefore, the ML estimation should be found via numerical approaches. This makes the ML estimation of concentration ratios not viable for resource limited low-complexity biological MC devices. To overcome this problem, next, we investigate estimation based on method of moments, by converting the observed statistics from the exact amount of bound time durations into the number of binding events lasting between specific time thresholds.

When the receptors are exposed to single type of ligands, the bound time durations follow exponential distribution with the unbinding rate, i.e., \( \tau_B \sim \text{Exp}[k^-] \). In the case of mixture of different ligand types, the probability distribution of bound time durations becomes a mixture of exponential distributions, i.e.,

\[ p(\tau_B) = \sum_{i=1}^{M} \alpha_i e^{-k_i^- \tau_B}. \quad (14) \]

The problem of Bayesian inference from mixture of exponential distributions does not lend itself to analytical solutions, and is usually tackled by computationally expensive iterative algorithms, e.g., expectation-maximization (EM) algorithm [51], [32], which are not feasible for resource-limited biomanufacturing. In this study, we propose a practical method by assuming that the unbinding rates of different ligand types existing in the channel are known to the estimator. As demonstrated in Fig. 2, we define time thresholds, \( T_i \)'s, corresponding to each ligand type, after sorting the ligand types in the decreasing order of unbinding rate, i.e., increasing order of their affinity with the receptors. Later, we will show that this transduction scheme is suitable for biological MC devices, as it can be implemented by active receptors based on well known KPR scheme.

The probability of observing a ligand binding event with a duration in between two time thresholds can be written as

\[ p_t = \int_{T_{i-1}}^{T_i} p(\tau_B) d\tau_B = \sum_{i=1}^{M} \alpha_i \left( e^{-k_i^- T_{i-1}} - e^{-k_i^- T_i} \right), \quad (15) \]

with \( T_0 = 0 \) and \( T_M = +\infty \). in matrix notation, it is given by

\[ p = S\alpha, \quad (16) \]

where \( p \) is an \((M \times 1)\) vector with the elements \( p_i \), and \( S \) is an \((M \times M)\) matrix with the elements

\[ s_{i,j} = e^{-k_j^- T_{i-1}} - e^{-k_j^- T_i}. \quad (17) \]

The number of binding events with durations in between time thresholds follows binomial distribution, with a mean and variance given by

\[ E[n] = pN, \quad (18) \]

\[ \text{Var}[n] = (p \odot (1 - p)) N, \quad (19) \]

where \( \odot \) denotes the Hadamard product, i.e., \( (A \odot B)_{i,j} = (A)_{i,j} (B)_{i,j} \).

We can apply the method of moments for the estimation of individual concentration ratios, by employing only the first moment, i.e., by matching the mean number of binding events corresponding to a time interval to the observed number of binding events for the same interval:

\[ n = \hat{p} N = S\hat{\alpha}N, \quad (20) \]

where a hat denotes the estimated parameters. The resulting estimator for the ligand concentration ratio vector is given by

\[ \hat{\alpha} = \left( \frac{1}{N} \right) W n, \quad (21) \]

where \( W = S^{-1} \), i.e., the inverse of \( S \) matrix, which is also an \((M \times M)\) matrix with elements \( w_{i,j} \). Accordingly, the estimated individual ligand concentration ratios, i.e.,

\[ \hat{\alpha}_l = \left( \frac{1}{N} \right) \sum_{i=1}^{M} n_i w_{l,i}, \quad (22) \]

are the weighted sums of \( M \) correlated binomial random variables with the weights \( w_{l,i} \). Hence, the variance of the ratio estimator can be written as

\[ \text{Var}[\hat{\alpha}_l] = \frac{1}{N^2} \sum_{i=1}^{M} N \sum_{j=1}^{M} w_{l,i} w_{l,j} \text{Cov}[n_i, n_j], \quad (23) \]

with

\[ \text{Cov}[n_i, n_j] = \begin{cases} \text{Var}[n_i], & \text{if } i = j, \\ -p_i p_j N, & \text{otherwise}. \end{cases} \quad (24) \]

Note that this is an unbiased estimator, as the expected value of the estimator is equal to the true value of the estimated parameter \( \alpha \), i.e.,

\[ E[\hat{\alpha}] = \left( \frac{1}{N} \right) W E[n] = W p = S^{-1} p = \alpha. \quad (25) \]

We also introduce a biased version of the ratio estimator, which has a simplified design, enabled when the threshold values are set sufficiently large. In this case, we can neglect the noisy contributions of the ligand types that have higher binding rates than the ligand type that is being estimated. When the thresholds are much larger than the corresponding unbinding rates, i.e., \( T_i \gg 1/k_i^- \), \( S \) matrix can be approximated by an upper triangular matrix, i.e.,

\[ H = S|_{T_i \gg 1/k_i^-}, \quad (26) \]

with the matrix elements

\[ h_{i,j} = \begin{cases} s_{i,j}, & \text{if } i < j, \\ e^{-k_i^- T_{i-1}}, & \text{if } i = j, \\ 0, & \text{otherwise}. \end{cases} \quad (27) \]

This approximation results in the following ratio estimator,

\[ \hat{\alpha}^* = \left( \frac{1}{N} \right) Rn, \quad (28) \]
Finally, the estimators for the concentrations of individual ligand types can be given as the product of the estimators for the total concentration and ligand concentration ratios, i.e.,

\[ \hat{c} = \hat{c}_{\text{tot}}, \quad \hat{c}^* = \hat{c}_{\text{tot}}^* \]

and the corresponding variances are found via

\[
\text{Var}[\hat{c}] = \text{Var}[\hat{c}_{\text{tot}}] \text{Var}[\hat{c}] + \text{Var}[\hat{c}_{\text{tot}}] (E[\hat{c}] \odot E[\hat{c}]) + \text{Var}[\hat{c}] E[\hat{c}_{\text{tot}}]^2,
\]

\[
\text{Var}[\hat{c}^*] = \text{Var}[\hat{c}_{\text{tot}}] \text{Var}[\hat{c}^*] + \text{Var}[\hat{c}_{\text{tot}}] (E[\hat{c}^*] \odot E[\hat{c}^*]) + \text{Var}[\hat{c}^*] E[\hat{c}_{\text{tot}}]^2.
\]

The bias for the simplified estimator of individual concentrations is given by

\[
\Delta[\hat{c}^*] = E[\hat{c}^*] - c = c_{\text{tot}} (E[\hat{c}^*] - \alpha) = c_{\text{tot}} \Delta[\hat{\alpha}^*].
\]

The resulting MSE for both estimators are given as follows

\[
\text{MSE}[\hat{c}] = \text{Var}[\hat{c}],
\]

\[
\text{MSE}[\hat{c}^*] = \text{Var}[\hat{c}^*] + (\Delta[\hat{c}^*] \odot \Delta[\hat{c}^*]).
\]

IV. PERFORMANCE ANALYSIS

We evaluate the performance of the proposed spectrum sensing method in terms of the normalized MSE (NMSE) averaged over all ligands in the mixture, i.e.,

\[
\langle \text{NMSE}[\hat{c}] \rangle = \frac{\text{NMSE}[\hat{c}]}{M} = \frac{1}{M} \sum_{i=1}^{M} \text{MSE}[\hat{c}_i].
\]

The average NMSE can be calculated for the simplified estimator \( \hat{c}^* \) in the same way. Note that with the normalization, we render the analysis independent of the total concentration. The performance of the proposed method only depends on the number of unbound and bound time duration samples, relative affinities of the ligand types with the receptors, number of ligand types, and the concentration ratios.

We define an additional parameter to formulate the affinity of ligand types with the receptors. For the simplicity of the analysis, we assume, without the loss of generality, that the unbinding rates of ligand types are in decreasing order, i.e.,

\[ k_1 > k_2 > \ldots > k_M. \]

We set the similarity parameter \( \chi = 5 \), such that the ratio of the unbinding rates between the two most similar ligands is \( \chi = 5 \).

The time thresholds are taken as proportional to the inverse of the unbinding rates of the corresponding ligand types, i.e.,

\[ T_i = \nu / k_i, \]

where \( \nu \) is the proportionality constant. For each ligand type, the isolation rates can be given as the product of the estimators for the total concentration and ligand concentration ratios, i.e.,

\[ \hat{c} = \hat{c}_{\text{tot}}, \quad \hat{c}^* = \hat{c}_{\text{tot}}^* \]

where \( R = H^{-1} \), is also an upper triangular matrix. The elements of \( R \) can be recursively calculated as follows

\[
r_{i,j} = k_j \left( I_{i=j} - \sum_{\gamma=1}^{j-i} r_{i+\gamma,j} \theta_{i+\gamma,i} \right),
\]

where \( I_{i=j} \) is the indicator function which is equal to 1 if \( i = j \), and 0 otherwise; \( k_j = \exp(-k_i T_j), \) and \( \theta_{i,j} = \exp(-k_i T_{j-i}) - \exp(-k_i T_{j-i}) \). The estimator for the concentration ratio of the \( l \)-th ligand type can be written as the sum of \( M-l+1 \) terms, i.e.,

\[
\hat{c}_l^* = \frac{1}{N} \sum_{i=0}^{M-l} n_{M-i} r_{l, M-l-i}.
\]

This substantially simplifies the ratio estimation of the ligand types with the highest affinities, which, in most cases, are the ones that are most relevant for information transfer in MC. Similar to [23], the variance of this estimator can be written as

\[
\text{Var}[\hat{c}_l^*] = \frac{1}{N^2} \sum_{i=0}^{M-l} \sum_{j=0}^{M-l} r_{l, M-l-i} \text{Cov}[n_{M-i, n_{M-j}}],
\]

where \( \text{Cov}[n_{M-i, n_{M-j}}] \) can be calculated using [24]. The mean of this estimator is given by

\[
E[\hat{c}_l^*] = \left( \frac{1}{N} \right) R E[n] = Rp.
\]

As is clear from [22], this is a biased estimator, due to the residuals resulting from the approximation. The resulting bias can be computed as follows

\[
\Delta[\hat{c}_l^*] = E[\hat{c}_l^*] - \alpha = Rp - S^{-1} p = (R - W)p.
\]

The MSE of this estimator can then be written as

\[
\text{MSE}[\hat{c}_l^*] = \text{Var}[\hat{c}_l^*] + (\Delta[\hat{c}_l^*] \odot \Delta[\hat{c}_l^*]).
\]
system setting analyzed, we perform the optimization of $\nu$ for the minimum average NMSE, i.e.,

$$\nu_{\text{opt}} = \arg \min_{\nu > 0} \{\text{NMSE}[\hat{e}]\}, \quad (41)$$

and provide the optimized value of the average NMSE together with the performance of unbiased and simplified estimators. The obtained values of $\nu_{\text{opt}}$ are different for each setting; however, we find that they concentrate around $\nu_{\text{opt}} = 3$ (data not shown). Therefore, in the performance evaluation of our unbiased estimator, we employ $\nu = 3$. For the simplified biased estimator, however, the value of $\nu$ is constrained by the fact that the simplification is based on the assumption that $T_i \gg k_i^-$. In our analysis, we conclude that setting $\nu = 5$ is sufficient for the validity of the assumption. Moreover, throughout the analysis we set the default number of samples, and default number of ligand types in the channel as $N = 10000$ and $M = 5$, respectively.

Given the default system setting above, next we evaluate the sensing performance for varying number, similarity, and ratio distribution of ligand types, and varying number of samples.

A. Effect of Number of Ligand Types in the Mixture

The first analysis is carried out for varying number of ligand types $M$. This is a critical parameter that depends on the interference characteristics of the MC channel, or the utilized medium access scheme. The results are provided in Fig. 3 for unbiased estimator, optimized unbiased estimator, and simplified biased estimator, where we assume that the concentration ratios of ligand types are equal, i.e., $\alpha_i = 1/M$ for $i \in \{1, 2, \cdots, M\}$. As expected, the NMSE is increasing with increasing $M$, however, the spectrum sensing method demonstrates an acceptable performance even when the channel is crowded by 10 different types of ligands. The results also reveal that the error performance of the unbiased estimator with $\nu = 3$ is very close to that of the optimized estimator for each setting. The performance of the simplified estimator follows the same trend; however, its error is almost an order of magnitude larger than the unbiased estimator when $M$ is high.

B. Effect of Similarity between Ligand Types

The similarity of the ligands co-existing in the channel has a significant effect on the performance of the spectrum sensing, as demonstrated in Fig. 4. An increase in the similarity, reflected by the decreasing $\chi$, reduces the capability of the sensing method to discriminate between different types of ligands from the bound time duration data. The results reveal that it is not possible to accurately sense the spectrum with the unbiased estimators when $\chi < 2$ and $M \geq 5$. Interestingly, however, the simplified biased estimator shows superior performance in this range of similarity, implying that neglecting the stochastic contribution of the ligands with lower affinities results in better error performance, when the ligands manifest very similar affinities with the receptors.

C. Effect of Number of Unbound/Bound Duration Samples

The number of samples affects the performances of both the ratio estimator $\hat{\alpha}$ and the total concentration estimator $c_{\text{tot}}$. As a result, the overall impact on the estimation of individual concentrations is remarkable, as demonstrated in Fig. 5. The relation between the average NMSE and the number of samples follows the same trend for all of the estimators, and the unbiased estimator has acceptable accuracy even when the number of samples $N = 500$, and $M = 5$.

D. Effect of Concentration Ratios of Ligands

Lastly, we evaluate the sensing performance for the case of heterogeneous distribution of concentration ratios, i.e., $\alpha_i = c_i/c_{\text{tot}}$. In particular, we change the ratio of the ligand with the highest affinity with the receptors, and keep the ratios of the other remaining ligand types homogeneously distributed, i.e.,

$$\alpha_i = \frac{1 - \alpha M}{M - 1}, \quad \text{for } i \in \{1, 2, \cdots, M - 1\}. \quad (42)$$

The results are provided in terms of average NMSE for unbiased and biased estimators in Fig. 6 and in terms of
NMSE of the unbiased and biased estimators of concentration of only the highest-affinity ligand in Fig. 7. Given that there are 5 different types of ligands in the channel, the average NMSE is maximized when the weights are almost homogeneously distributed, i.e., when $\alpha_M \approx 0.2$. Interpreting both results together, we see that while the accuracy of the estimation of $c_M$ significantly increases for very high values of $\alpha_M$, the overall performance of the spectrum sensing deteriorates. In an MC application, we can expect that the molecules of interest, i.e., information-carrying molecules, would be the ligands that have the highest affinity with the employed receptors. Hence, the results show that the proposed spectrum sensing method can be effectively used to eliminate the interference of lower-affinity ligands for improving the detection performance.

V. DISCUSSION ON IMPLEMENTATION

In this section, we discuss the proposed spectrum sensing method from a practical perspective. Our focus is on synthetic biology-enabled bacteria-based implementation approach, as biosensor-based approaches for MC receiver, overviewed in \[5\], \[28\], do not allow inspecting the states of individual receptors.

The key element in the spectrum sensing is the biological receptors, which are the interface between the exterior and interior of a living cell, and transduce the external signals represented by the concentration of ligands into intracellular signals in the form of concentration of second messengers inside a living cell. The transduced signals need to be further processed for the estimation to be achieved.

The proposed estimator relies on two statistics, total unbound time $T_U$, and the number of binding events $n_i$ of durations within $[T_i-1, T_i]$. Our aim is to provide a potential approach for the design of synthetic receptors that can encode both the unbound time and bound time information into the concentration of intracellular molecules. The proposed receptor comprises $(5 + M)$ states, in which it shows different reaction characteristics, as shown in Fig. 8(a). The receptor utilizes both a modified kinetic proofreading (KPR) scheme with $M$ states for obtaining the number of binding events for each time interval, and an activation mechanism that we previously proposed for obtaining the total ligand concentration \[18\].

The sampling process starts with the generation of the activation molecules, $A$, generated in an impulsive manner by the receiver, when it decides to sample receptor states, as demonstrated in Fig. 8(a). The generation of activation signal is controlled by the rate $s(t)\psi^+$. Shortly after the activation, the cell generates deactivation molecules $A^-$ with a rate of $d(t)\psi^-$. The deactivation molecules degrade the activation molecules within the cell at a rate $\rho$, such that the duration of the overall sampling process can be controlled. The inactive receptors $(U_A, B_0)$ first become semi-active when they react with the activation molecules $A$ (see also Fig. 8(b)). The binding of a semi-active unbound receptor $U_A^*$ transforms it into a semi-active bound receptor $B_A^*$. Upon the first unbinding event, a semi-active bound receptor $B_A^*$ goes into the active unbound state $U_A$, where it can catalyze an intracellular reaction, i.e., $S \xrightarrow{\mu} S^*$. The overall concentration of the resulting $S^*$ molecules becomes proportional to the total unbound time $T_U$. 

![Fig. 5. Average NMSE with varying number of unbound and bound time duration samples, $N$, for unbiased $\hat{c}$, optimal unbiased $\hat{c}_{opt}$, and simplified biased $\hat{c}^*$ estimators.](image)

![Fig. 6. Average NMSE with varying concentration ratio of the highest-affinity ligand type, $\alpha_M$, for unbiased $\hat{c}$, and simplified biased $\hat{c}^*$ estimators.](image)

![Fig. 7. NMSE in the estimation of concentration of the highest-affinity ligand type, $c_M$, with varying concentration ratio of the highest-affinity ligand type, $\alpha_M$, for unbiased $\hat{c}$, and simplified biased $\hat{c}^*$ estimators.](image)
Upon binding of an active unbound receptor $U_A$ to a ligand, it switches into the active bound state $B^A_1$, and activates the modified KPR scheme. Note that the knowledge of the number of samples taken for the estimation of total concentration is critical for accuracy. A wise way to obtain a stable accuracy is to fix the number of samples to the number of receptors, i.e., $N = N_R$. To ensure that the inactivated receptors are not re-activated during the same sampling process, the generation control signals $s(t)$ and $d(t)$ should be impulse-like.

In the KPR mechanism, the active bound receptor protein sequentially visits its $M$ internal states in an irreversible manner during the bound time period by undergoing a series of conformational changes at specific transition rates, as shown in Fig. 8 (a). In each internal state, the receptor is allowed to return directly to the initial unbound state $U_I$ through unbinding of the bound ligand at the unbinding rate of that particular ligand type. In our modified KPR scheme, while returning to the initial unbound state, the receptor releases an intracellular molecule $D_i$, type of which is specific to the last occupied internal state. The transition delay between internal states can be set through modifying the state transition rates $\beta_i$, given that $E[T_i] = 1/\beta_i$. If the delays are set to match the time thresholds considered in Section III, the resulting number of second messengers, $D_i$, produced from the internal states $B^A_i$ gives the number of binding events of durations within corresponding time ranges.

Once the transduction of total unbound time $T_{U_I}$ and the number of binding events $n_i$ for each interval into the concentration of second messengers is completed, the arithmetic operations required for the estimator can be realized through intracellular chemical reaction networks (CRN) that can perform analog computations \[^{33}\]. Accordingly, the unbiased estimator of concentration weights $\hat{\alpha}$ requires the weighted sum of concentration of $(M - 1)$ different types of second messengers for each ligand type, whereas the simplified biased estimator $\hat{\alpha}^*$ requires the same with only $(M - 1 - i)$ terms for the $i^{th}$ ligand ranked in decreasing order of unbinding rates. On the other hand, the unbiased ML estimator of the total ligand concentration, i.e., $\hat{\alpha}_{tot} = \frac{N_R}{k + c_i}$, necessitates the division of a constant term by the concentration of second messengers representing $T_{U_I}$, which is hard to implement with CRNs. Instead, the overall estimators of individual concentrations, i.e., $\hat{\alpha} = \hat{\alpha}_{tot}\hat{\alpha}$ and $\hat{\alpha}^* = \hat{\alpha}_{tot}\hat{\alpha}^*$, can be represented by their logarithms, such that the estimation problem can be solved only by summation/substraction and logarithm operations. Note that, Implementation of logarithm function with CRNs proved feasible in \[^{33}\]; however, the design and the performance analysis of the estimators with CRNs are left as an open issue. Nevertheless, we can expect that the additional randomness resulting from the stochastic CRNs would have an substantial impact on the overall spectrum sensing performance. Therefore, the analyses carried out in Section LV should be taken as providing upper performance limits for the proposed spectrum sensing method.

VI. CONCLUSION

In this paper, we develop a spectrum sensing technique for MC with ligand receptors for the first time in the literature. In light of the results, we discuss that the proposed technique can be utilized for developing reliable MC detection, modulation, and medium access methods, as it proved effective in sensing the individual concentrations of multiple ligand types by eliminating the interference from the others. The technique is practical and low-complexity, and can be implemented in resource-limited biological MC devices, e.g., engineered bacteria. In this direction, we also discuss a synthetic receptor design, built upon the kinetic proofreading mechanism, that can transduce the required statistics of ligand-receptor binding reaction into intracellular signals. This study is not exhaustive, and there remain many challenges and opportunities calling for future research. For example, interesting generalizations can be obtained by investigating the cases where ligand unbinding rates are not known a priori to the estimators, or by studying non-equilibrium cases where the concentration of ligands are changing rapidly compared to the binding kinetics. As
discussed throughout the paper, the study can be extended with the applications of the proposed method in developing reliable detection methods for CSK, MoSK and RSK modulated MC signals, and molecular division medium access techniques.

REFERENCES

[1] O. B. Akan, H. Ramezani, T. Khan, N. A. Abbasi, and M. Kuscu, “Fundamentals of molecular information and communication science,” Proceedings of the IEEE, vol. 105, no. 2, pp. 306–318, 2017.

[2] I. F. Akyildiz and J. M. Jornet, “The internet of nano-things,” IEEE Wireless Communications, vol. 17, no. 6, 2010.

[3] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, “The internet of bio-nano things,” IEEE Communications Magazine, vol. 53, no. 3, pp. 32–40, 2015.

[4] M. Kuscu and O. B. Akan, “The internet of molecular things based on ftri,” IEEE Internet of Things Journal, vol. 3, no. 1, pp. 4–17, 2016.

[5] M. Kuscu, E. Dinc, B. A. Bilgin, H. Ramezani, and O. B. Akan, “Transmitter and receiver architectures for molecular communications: A survey on physical design with modulation, coding and detection techniques,” Proceedings of the IEEE, under review.

[6] E. Dinc and O. B. Akan, “Theoretical limits on multiuser molecular communication in internet of nano-bio things,” IEEE Transactions on Nanobioscience, vol. 16, no. 1, pp. 266–270, 2017.

[7] V. Jamali, A. Ahmadzadeh, C. Jardin, H. Sticht, and R. Schober, “Channel estimation for diffusive molecular communications,” IEEE Transactions on Communications, vol. 64, no. 10, pp. 4238–4252, 2016.

[8] A. Noel, K. C. Cheung, and R. Schober, “Joint channel parameter estimation via diffusive molecular communication,” IEEE Transactions on Molecular, Biological and Multi-Scale Communications, vol. 1, no. 1, pp. 4–17, 2015.

[9] A. Alizadeh, H. R. Bahrami, M. Maleki, N. H. Tran, and P. Mohseni, “On the coexistence of nano networks: Sensing techniques for molecular communications,” IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 2018.

[10] W. Bialek, Biophysics: searching for principles. Princeton University Press, 2012.

[11] T. Mora, “Physical limit to concentration sensing amid spurious ligands,” Physical Review Letters, vol. 115, no. 3, p. 038102, 2015.

[12] Y. Deng, A. Noel, M. Elkashlan, A. Nallanathan, and K. C. Cheung, “Modeling and simulation of molecular communication systems with a reversible adsorption receiver,” IEEE Transactions on Molecular, Biological and Multi-Scale Communications, vol. 1, no. 4, pp. 347–362, 2015.

[13] A. Ahmadzadeh, H. Arjandi, A. Burkovski, and R. Schober, “Comprehensive reactive receiver modeling for diffusive molecular communication systems: Reversible binding, molecule degradation, and finite number of receptors,” IEEE Transactions on Nanobioscience, vol. 15, no. 7, pp. 713–727, 2016.

[14] M. Kuscu and O. B. Akan, “Modeling and analysis of sinr ftri-based molecular communication receiver,” IEEE Transactions on Communications, vol. 64, no. 9, pp. 3708–3721, 2016.

[15] C. T. Chou, “A markovian approach to the optimal demodulation of diffusion-based molecular communication networks,” IEEE Transactions on Communications, vol. 63, no. 10, pp. 3728–3743, 2015.

[16] R. G. Endres and N. S. Wingreen, “Maximum likelihood and the single receptor,” Physical Review Letters, vol. 103, no. 15, p. 158101, 2009.

[17] C. T. Chou, “Maximum a-posteriori decoding for diffusion-based molecular communication using analog filters,” IEEE Transactions on Nano-biotechnology, vol. 14, no. 6, pp. 1054–1067, 2015.

[18] B. A. Bilgin, O. B. Akan, “Maximum likelihood detection with ligand receptors for diffusion-based molecular communications in internet of bio-nano things,” IEEE Transactions on Nanobioscience, vol. 17, no. 1, pp. 44–54, 2018.

[19] V. Singh and I. Nemenman, “Simple biochemical networks allow accurate sensing of multiple ligands with a single receptor,” PLoS Computational Biology, vol. 13, no. 6, e1005490, 2017.

[20] J.-B. Lalanne and P. Francois, “Chemosensation in fluctuating environments: Receptor coupling, buffering, and antagonism,” Proceedings of the National Academy of Sciences, vol. 112, no. 6, pp. 1898–1903, 2015.

[21] E. D. Sigga and M. Vergassola, “Decisions on the fly in cellular sensory systems,” Proceedings of the National Academy of Sciences, vol. 110, no. 39, pp. E3704–E3712, 2013.

[22] T. W. McKeithan, “Kinetic proofreading in t-cell receptor signal transduction,” Proceedings of the National Academy of Sciences, vol. 92, no. 11, pp. 5042–5046, 1995.

[23] G. Muzio, M. Kuscu, and O. B. Akan, “Selective signal detection with ligand receptors under interference in molecular communications,” in 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, 2018, pp. 1–5.

[24] M. S. Kuran, H. B. Yilmaz, T. Tugcu, and I. F. Akyildiz, “Modulation techniques for communication via diffusion in nanonetworks,” in Communications (ICC), 2011 IEEE International Conference on. IEEE, 2011, pp. 1–5.

[25] H. ShahMohammadian, G. G. Messier, and S. Magierski, “Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels,” Nano Communication Networks, vol. 3, no. 3, pp. 183–195, 2012.

[26] N.-R. Kim and C.-B. Chae, “Novel modulation techniques using isomers as messenger molecules for nano communication networks via diffusion,” IEEE Journal on Selected Areas in Communications, vol. 31, no. 12, pp. 847–856, 2013.

[27] L. P. Giné and I. F. Akyildiz, “Molecular communication options for long range nanonetworks,” Computer Networks, vol. 53, no. 16, pp. 2753–2766, 2009.

[28] M. Kuscu and O. B. Akan, “On the physical design of molecular communication receiver based on nanoscale biosensors,” IEEE Sensors Journal, vol. 16, no. 8, pp. 2228–2243, 2016.

[29] O. B. Akan, O. B. Karli, and O. Ergul, “Cognitive radio sensor networks,” IEEE network, vol. 23, no. 4, 2009.

[30] M. B. Bereshkovskii and A. Szabo, “Effect of ligand diffusion on occupancy fluctuations of cell-surface receptors,” The Journal of Chemical Physics, vol. 139, no. 12, p. 09B610_1, 2013.

[31] V. Hasselblad, “Estimation of finite mixtures of distributions from the exponential family,” Journal of the American Statistical Association, vol. 64, no. 328, pp. 1459–1471, 1969.

[32] N. P. Jewell, “Mixtures of exponential distributions,” The Annals of Statistics, pp. 479–484, 1982.

[33] R. Daniel, J. R. Rubens, R. Sarapeshkar, and T. K. Lu, “Synthetic analog computation in living cells,” Nature, vol. 497, no. 7451, p. 619, 2013.

[34] C. T. Chou, “Chemical reaction networks for computing logarithm,” Synthetic Biology, vol. 2, no. 1, p. yxx002, 2017.