

Optimizing the Spatial Topology of Bacterial Relay Systems: Delay Minimization

Mustafa Can Gursoy*, Sonali Gupta[†], Ophelia S. Venturelli[†], and Urbashi Mitra*

*Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, USA

[†] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

E-mails: {mgursoy, ubli}@usc.edu, {sgupta97, venturelli}@wisc.edu

Abstract—Diffusion-based molecular communication (DBMC) between spatially separated bacterial colonies has limited range due to slow diffusive propagation. To this end, relay-aided DBMC with bacterial colonies as nodes is considered in this paper. A deterministic framework that governs the overall system behavior is provided for amplify-and-forward (AF) type relays. Motivated by real-life constraints in practical implementation, the framework is expanded to cover a maximum saturation limit on emission intensity, yielding the AF-with saturation (AFS) relay model. For n -hop bacterial DBMC with AFS relays, a trade-off between diffusion delay and relay processing time is investigated, which hints to an optimal number of relays that minimizes end-to-end delay. A tractable objective function for the end-to-end delay is provided by approximating the system as a cascade of n one-hop links. Numerical results show that the approximation is tight, and up to 50% decrease in end-to-end delay can be achieved by optimizing the number of relays.

Index Terms—Diffusive molecular communications, relays, end-to-end delay, bacterial molecular communication

I. INTRODUCTION

Diffusion-based molecular communication (DBMC) has heavy inter-symbol interference (ISI) due to the characteristics of the diffusion process [1]. As in radio communications, unmitigated ISI can affect data rates and reliability. In DBMC, ISI is exacerbated by increasing ranges. As a potential solution, relay-aided DBMC has been extensively considered. In [2], the molecular multiple access, broadcast, and relay channels are defined. Amplify-and-forward (AF) relays are examined in [3]–[5], estimate-and-forward (EF) in [6], and decode-and-forward (DF) in [7]–[10].

The above studies consider error rates as performance metrics, and focus mostly on the communication theoretic side of the design. On the other hand, end-to-end delay is seldom considered as a performance metric in the relay-aided DBMC literature. In [11], delay and link reliability are considered in the context of multihop molecular communication, where the information is encoded within the genetic material of viruses. Similarly, [12] considers routing and the associated delay when encoding information within the plasmid of a bacterium, whose motility carries the information (as opposed to diffusion-based propagation). In [13], a rate-delay trade-off within a two-hop DBMC system is investigated, when the

information is encoded information within the body of a large molecule, assuming that a single molecule's arrival is sufficient for successful information transfer. In contrast to [11]–[13], we focus on (relay-aided) DBMC where each node is a bacterial colony within a chamber and concentration-based signaling is employed. In this regard, the closest line of work to our study is [14]–[16]. However, as much of the DBMC relay work, this prior art focuses on error probability and capacity metrics.

Synthetic biology offers significant promise towards engineering living cells including bacteria to achieve desired functions in applications ranging from the production of compounds that benefit human health or enhance plant growth and resistance to invasion [17]. A key challenge is programming the collective behavior of bacteria by exploiting information processing systems including quorum sensing [18]–[20]. Predictable design of complex synthetic circuits has been limited by unintended interactions with the host cell genome, leading to negative impacts on cellular fitness [21]. To reduce metabolic burden, circuits could be partitioned among different sub-populations or community members to achieve a community-level function [22]. Further, these sub-populations could be engineered for long-range signal communication [23]. Designing such functions [24], [25] necessitates understanding how quickly or slowly microbial communities activate. This activation time question can be answered through delay optimization in spatial networks of microbial communities. To the best of our knowledge, our study is the first work that considers delay minimization in bacterial relay-aided DBMC. The models we propose herein are inspired by experimental results [24], [26] on synthetically engineered bacterial relay systems.

The contributions of this paper are as follows:

- 1) Considering bacterial colonies as nodes in a relay-aided DBMC setting, we provide a model framework that governs the overall system behavior for AF-type relays.
- 2) Motivated by experimental data, we extend our framework to incorporate biologically consistent constraints such as a maximum saturation limit on emission intensity (as exhibited by real world bacterial colonies), and propose the amplitude-and-forward with saturation (AFS) relay model.
- 3) For the AFS model, we approximate the end-to-end system as a cascade of identical one-hop links. Based on the approximate model, we provide a tractable objective

This work has been funded in part by one or more of the following grants: ONR N00014-15-1-2550, NSF CCF-1817200, ARO W911NF1910269, Cisco Foundation 1980393, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, ONR 503400-78050.

function that can be used for optimizing the number of relays numerically.

- 4) Numerical results show that the approximate model closely follows the actual model, and demonstrates the existence of an optimal number of relays that minimizes end-to-end delay.

The remainder of this paper is organized as follows: Section II presents the system model and the corresponding signal model. Section III characterizes the evolution of the system observations as a function of the initial state. Section IV provides an approximation to the end-to-end system, and presents an objective function for optimization. Section V presents numerical results and Section VI concludes the paper.

II. SYSTEM AND SIGNAL MODEL

The system of interest in this paper is an n -hop molecular communication system between a transmitter node (TX, N_0) and a receiver node (RX, N_n), with $n - 1$ relay nodes in between ($N_1, \dots, N_{n-1}$). All nodes are assumed to be identical. Throughout the paper, we model each node to be passive observers in a driftless, unbounded, 1-D medium. Each node is assumed to observe a region of length L . The distance between the centers of the transmitter and the receiver is denoted by d , and the relays are equally spaced between TX and RX. Thus, the distance between the centers of node i and node j is equal to $d_{ij} = d \frac{|i-j|}{n}$. The system considered in the paper is presented in Figure 1.

For this system model, given unit impulse emission from N_i , the concentration at the center point of N_j is given by $c_{ij}(t) = \frac{1}{(4\pi Dt)^{1/2}} \exp(-\frac{d_{ij}^2}{4Dt})$, where D denotes the diffusion coefficient of the utilized molecule [27], [28]. When $L \ll d_{ij}$ (which is satisfied throughout the paper), the total concentration observed within a node's body can be approximated to be uniform [29], yielding $h_{ij}(t) \approx Lc_{ij}(t)$. Using said uniform concentration approximation (UCA, [29]), the channel impulse response (CIR) between N_i and N_j can be expressed as

$$h_{ij}(t) = \frac{L}{\sqrt{4\pi Dt}} \exp\left(-\frac{d_{ij}^2}{4Dt}\right). \quad (1)$$

We consider the *collective* behavior of large bacterial colonies as opposed to a single cell/nano-machine. Thus, each node represents the collective response of the colony at the node as a whole. Naturally, the concentration scale of the system of interest is significantly larger than typically considered for nano-scale applications. Throughout the paper, we assume that the concentration scale is large enough that the effects of perturbations around the mean (*i.e.*, CIR) are negligible when characterizing the observed number of molecules/concentrations. Thus, we employ a *deterministic* model in this paper. Under this assumption, using a discrete time model with time increments Δt , for each (i, j) pair with $i \neq j$, the series $\{h_{ij}(k\Delta t)\}_{k \geq 1}$ fully characterizes the effect of N_i on N_j . Henceforth, for brevity in notation, we refer to time $k\Delta t$ with the discrete index k , *e.g.*, $h_{ij}(k\Delta t) = h_{ij}[k]$.

Let $r_j[k]$ denote the observed local concentration at node j at time k , and $x_j[k]$ denote its total generated signal. We model

$r_j[k]$ to consist of two components: the aggregate diffusion-caused effect of all nodes other than N_j ($r_j^{\text{dif}}[k]$), and the effect of self-feedback within the colony ($r_j^{\text{SF}}[k]$). That is,

$$r_j[k] = r_j^{\text{SF}}[k] + r_j^{\text{dif}}[k]. \quad (2)$$

Within the scope of this paper, every node utilizes the same molecule type. We assume that a fraction $\beta \in (0, 1)$ of the molecules emitted at time $(k-1)$ from N_j are “trapped” within the j th node, contributing to the arrival count of N_j at time k . This consideration accounts for the self-feedback loop of the bacterial colony at N_j , and implies

$$r_j^{\text{SF}}[k] = \beta x_j[k-1]. \quad (3)$$

Note that we assume a self-feedback with a memory of one time-slot and with unit delay. The extension to more memory self-feedback is left as future work.

The remaining $(1 - \beta)$ fraction of the transmitted concentration is “leaked out” of the colony, yielding the emission of N_j at time k to be $(1 - \beta)x_j[k]$. The diffusion-induced component of $r_j[k]$ is the superposition of the effects of all other transmitting nodes in the system¹. Due to the convolutional nature of the DBMC channel, this component is then equal to

$$r_j^{\text{dif}}[k] = (1 - \beta) \sum_{\substack{i=0 \\ i \neq j}}^n \sum_{l=1}^k x_i[k-l] h_{ij}[l]. \quad (4)$$

From Equations (3) and (4), we see that $r_j[k]$ is affected by previous emissions of all nodes in the system. Denoting $\mathbf{r}[k] = [r_0[k] \ \dots \ r_n[k]]^\top$, the characterization of the system evolution through considering $\mathbf{r}[k]$ as the state at time k is addressed in the sequel.

III. SYSTEM EVOLUTION CHARACTERIZATION

Herein, we characterize $\mathbf{r}[k]$ in terms of the initial state $\mathbf{r}[0]$, for AF type relays. Combining (3) and (4), we have

$$r_j[k] = \beta x_j[k-1] + (1 - \beta) \sum_{\substack{i=0 \\ i \neq j}}^n \sum_{l=1}^k x_i[k-l] h_{ij}[l]. \quad (5)$$

For AF-type relays, denoting the amplification factor/gain as α and $\mathbf{x}[k] = [x_0[k] \ \dots \ x_n[k]]^\top$, we can write

$$\mathbf{x}[k] = \alpha \mathbf{r}[k], \quad (6)$$

leading to

$$r_j[k] = \alpha \beta r_j[k-1] + \alpha(1 - \beta) \sum_{\substack{i=0 \\ i \neq j}}^n \sum_{l=1}^k r_i[k-l] h_{ij}[l]. \quad (7)$$

Setting $h_{ii}[1] := \frac{\beta}{1-\beta}$ and $h_{ii}[k] := 0$ for all $k \neq 1$, we get

$$r_j[k] = \delta \sum_{i=0}^n \sum_{l=1}^k r_i[k-l] h_{ij}[l] = \delta \sum_{l=1}^k \mathbf{h}_j[l]^\top \mathbf{r}[k-l], \quad (8)$$

¹We assume that the effects of diffusive self-interference (SI) are accounted for within the internal self-feedback loop in (3).

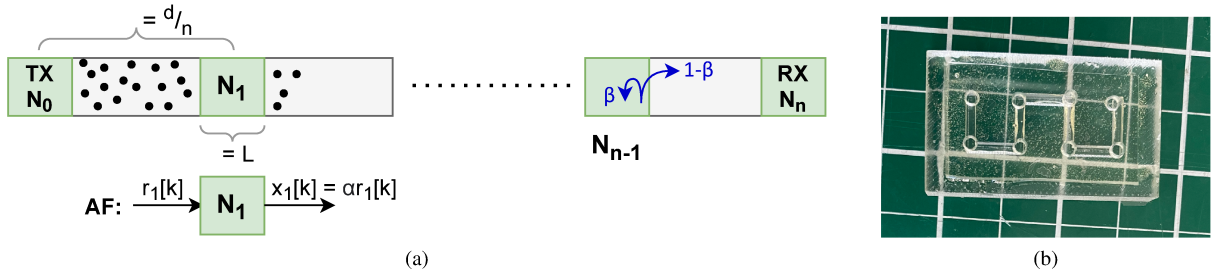


Fig. 1. System model of interest. An idealized n -hop DBMC in 1-D (left). A practical, microfluidic channel-based n -hop system in 1-D with $n = 7$ (right).

where $\mathbf{h}_j[l] = [h_{0j}[l] \ h_{1j}[l] \ \dots \ h_{nj}[l]]^\top$ is the vector that holds the channel coefficients going into node j , and the parameter $\delta = \alpha(1 - \beta)$. Therefore, the overall system update equation can be written as

$$\mathbf{r}[k] = \delta \sum_{l=1}^k \mathbf{H}[l]^\top \mathbf{r}[k-l], \quad (9)$$

with $\mathbf{H}[l] = [\mathbf{h}_0[l] \ \dots \ \mathbf{h}_n[l]]$. Note that (9) fully characterizes $\mathbf{r}[k]$ as a function of all previous observations including the initial state $\mathbf{r}[0]$ and a finite number of pre-determined matrices $\mathbf{H}[l]$ that depend on system parameters. Building upon (9), we provide the following theorem that characterizes $\mathbf{r}[k]$ solely using δ , $\mathbf{H}[1], \dots, \mathbf{H}[k]$, and the initial state $\mathbf{r}[0]$.

Theorem 1. For the system defined by Equations (2)-(6), the observation vector (state) at time k can be written as

$$\mathbf{r}[k] = (\mathbf{H}^{0 \rightarrow k})^\top \mathbf{r}[0], \quad (10)$$

$$\text{where } \mathbf{H}^{0 \rightarrow k} = \sum_{b=1}^k \delta^b \sum_{c=1}^{(k-1)} \mathbf{H}_c^{(b)}, \quad \mathbf{H}_c^{(b)} \in \mathcal{H}^{(b)}[k] \quad (11)$$

$$\mathcal{H}^{(b)}[k] = \left\{ \prod_{i=1}^b \mathbf{H}[q_i] : \sum_{i=1}^b q_i = k, \ q_i \geq 1 \right\}. \quad (12)$$

Proof. The proof is given in Appendix A. $\square$

Theorem 1 characterizes the system for an arbitrary input vector $\mathbf{r}[0]$ (including cases where the relay nodes or RX have non-zero initial state) and for purely AF-type relays. This model is accurate when describing $\mathbf{r}[k]$, but has challenges:

Remark 1. For each time index k , evaluating $\mathbf{H}^{0 \rightarrow k}$ requires evaluating $\binom{k-1}{b-1}$ expressions with b multiplicands. This incurs a complexity $\mathcal{O}((k-1)^{\min(b-1, k-b)})$ when evaluating each b -term summand, and makes computing $\mathbf{r}[k]$ increasingly difficult as k increases.

Remark 2. Due to the positive self-feedback at each relay node, whenever $\alpha\beta > 1$, $r_j[k]$ exponentially increases in magnitude since $r_j[k] \geq \alpha\beta r_j[k-1]$ per Equation (7), yielding an unstable system (in the sense that bounded $r_0[k]$ yields unbounded $r_j[k]$ for $j \neq 0$). However, as can also be confirmed from Figure 2, actual relay colonies exhibit a *self-saturation* that limits their maximum output.

Motivated by Remark 2, we extend our model to capture the saturation effect. Throughout the manuscript, we call

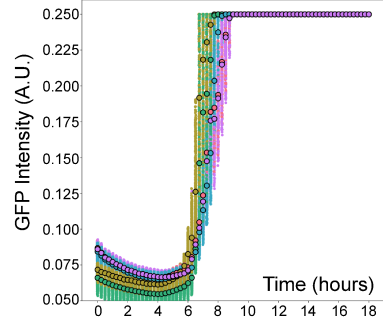


Fig. 2. Fluorescence of green fluorescent protein (GFP) controlled by LuxI regulated promoter as a function of time. Wet-lab experiment where bacteria are grown on agar and imaged using fluorescent time-lapse microscopy (each plot corresponds to a different spatial region within the imaged colony). The bacteria produces N-acyl-homoserine lactone (AHL) that activates the expression of the AHL synthetase and GFP signal, resulting in a positive feedback loop (up to a self-saturation that limits maximum output).

this extended model the *amplify-and-forward with saturation* (AFS) model.

Henceforth, we focus on the particular case where the transmitter is subject to a constant input M throughout the experiment (adjusted exogenously), and all relay nodes and RX have a zero initial state, implying $\mathbf{r}[0] = [M \ 0 \ \dots \ 0]^\top$. Note that since TX observes $r_0[k] = M$ for all k , we have $x_0[k] = \alpha M$. Motivated by this and the assumption that all nodes are identical, we limit the maximum allowable emission intensity to αM in the AFS model. Thus, we have

$$x_j[k] = \min(\alpha r_j[k], \alpha M), \quad \forall j \in \{0, \dots, n\}. \quad (13)$$

IV. CHARACTERIZING THE DELAY AND OPTIMIZATION

Herein, we address the problem of optimizing the number of relays between a fixed TX and RX using AFS-type relays. In other words, we seek to find the optimal n , given a fixed d . Recall that the initial time $k = 0$ is the first instant when the TX is stimulated (*i.e.*, $r_0[0] = M$ and $r_0[k] = 0$ for all $k < 0$). We define the overall activation time (*i.e.*, end-to-end delay) of the system, T_a , as the first instant (*i.e.*, smallest k) when the last colony is *fully* activated (reaches saturation). For an n -hop AFS system, the end-to-end delay is then defined as

$$T_a(n) = \min\{k \in \mathbb{Z}^+ : r_n[k] \geq M\}. \quad (14)$$

Thus, the optimal n that minimizes the delay is expressed as

$$n^* = \arg \min_n T_a(n) = \arg \min_n \min_k \{k : r_n[k] \geq M\}. \quad (15)$$

Mathematically, as each node is modeled as a passive receiver, the introduction of extra relay nodes do not negatively affect the pre-existing nodes' observations. Therefore, extra relays can only help, which yields the trivial solution $n^* = \infty$ to minimize delay. However, in real life, bacterial colonies that are close by compete for nutritional resources [30]. This phenomenon imposes a cost associated with introducing extra relays, as increasing n decreases $\frac{d}{n}$, causing more severe competition for food among adjacent colonies. Motivated by the tradeoff between the "health" (fitness) of the colony and the signal amplification it can produce, we incorporate this phenomenon into the gain parameter α , which now becomes a function of n . In particular, denoting the 1-hop gain by α_1 , we assume $\alpha = \frac{\alpha_1}{n^p}$ for some exponent $p \in \mathbb{R}^+$.

A smaller α implies that each relay takes a longer time to reach sufficiently large emission intensities once it starts receiving a non-zero signal, due to slower self-feedback. This introduces a *processing delay* that increases with increasing n . Thus, the effect of n on α yields a trade-off between the benefit of a relay (diffusion delay) and processing delay, which hints to the existence of a finite n^* that minimizes overall delay by balancing these two sources of delay.

Unfortunately, the optimization problem described by (15) is computationally challenging, since n governs α , the channel coefficients $h_{ij}[k]$, as well as the size of the $\mathbf{r}[k]$ vectors. Furthermore, as the saturation in (13) introduces a non-linearity to the system, (9) is no longer valid for explaining system behavior, rendering Theorem 1 inapplicable for AFS relays. Since each node affects every other node, this interconnected evolution of the system severely challenges optimization.

To overcome these computational challenges, we approximate the end-to-end delay of the system as

$$T_a \approx n \times T_{01}, \quad (16)$$

where T_{01} denotes the one-hop delay between the TX and the first relay (i.e., N_0 and N_1). We note that this approximation is particularly accurate when

- 1) The one-hop distance $\frac{d}{n}$ **is large**: This regime describes the case where nodes are not too close to one another. When it is satisfied, the signal at node j is dominated by the contributions of the closest nodes N_{j-1} and N_{j+1} , since a large $d_{j,j \pm j'}$ causes the effects of $N_{j \pm j'}$ (where $j' > 1$) to be negligible compared to $N_{j \pm 1}$. Recalling $\mathbf{r}[0] = [M \ 0 \ \dots \ 0]^\top$, the activation/saturation is likely to be "one-by-one" and the effect of N_{j+1} would be negligible compared to N_{j-1} , thus it is reasonable to further approximate $r_j^{\text{dif}}[k]$ as

$$r_j^{\text{dif}}[k] \approx (1 - \beta) \sum_{l=1}^k x_{j-1}[k-l] h_{j-1,j}[l]. \quad (17)$$

For a small one-hop distance, N_j 's effect on $N_{j \pm j'}$ becomes non-negligible and considering (17) becomes

inaccurate. This phenomenon is also observed experimentally in [26, Figure 5.6].

- 2) The parameter α **is large**: This regime implies that each node reaches saturation rapidly upon receiving a non-zero signal. To provide insight, consider $\alpha \rightarrow \infty$. Then, immediately upon node $j-1$ receiving any non-zero signal (say at time k), node $j-1$ would reach saturation at time $k+1$, that is $x_{j-1}[l] = \alpha M, \forall l \geq k+1$. Assuming that Point 1 above is satisfied, this implies that from N_j 's perspective, the N_0-N_1 link is approximately equivalent to the $N_{j-1}-N_j$ link. Then, the aggregate end-to-end delay is well-approximated by the sum of n identical, one-hop delays that are $\approx T_{01}$.

We will validate these approximations in the next section.

Using the approximation in (16), and noting that $T_{01} = \min_k \{k \in \mathbb{Z}^+ : r_1[k] \geq M\}$, the optimization problem in (15) can be reformulated as

$$n^* = \arg \min_n n \times \min_k \{k \in \mathbb{Z}^+ : r_1[k] \geq M\}. \quad (18)$$

We underscore that $r_1[k]$ is still a function of n in the hypothetical one-hop scenario used for approximation, as n affects both $h_{01}[k]$ through $\frac{d}{n}$, and α . Given $x_0[k] = \alpha M, \forall k \geq 0$, combining (7) with (17) and expanding the recursion for time k , the general expression governing $r_1[k]$ can be written as

$$r_1[k] = M \frac{1-\beta}{\beta} \sum_{i=1}^k q_i (\alpha\beta)^i, \quad (19)$$

where $q_i = \sum_{j=1}^{k-i+1} h_{01}[j]$. Then, the one-hop delay can be expressed as

$$\begin{aligned} T_{01} &= \min_k \left\{ k : M \frac{1-\beta}{\beta} \sum_{i=1}^k q_i (\alpha\beta)^i \geq M \right\} \\ &= \min \left\{ k : \sum_{i=1}^k q_i (\alpha\beta)^i \geq \frac{\beta}{1-\beta} \right\}. \end{aligned} \quad (20)$$

Therefore, recalling $\alpha = \frac{\alpha_1}{n^p}$, the optimal number of hops for this approximate case can be found by

$$n^* = \arg \min_n n \times \min_k \left\{ k : \sum_{i=1}^k q_i (\alpha_1 \beta n^{-p})^i \geq \frac{\beta}{1-\beta} \right\}. \quad (21)$$

Although the solution of (21) still needs numerical evaluation, this objective function is considerably easier to compute than (15). This is the case for the following three reasons: (a) the expression does not depend on the size of the vector $\mathbf{r}[k]$, (b) the problem only requires the values of $h_{01}[k]$ in contrast to $h_{ij}[k]$ for all $i \neq j$, and (c) the problem only requires the evaluation of this single link as opposed to considering the system as a whole.

V. NUMERICAL RESULTS

In this section, we provide numerical results to assess the approximation accuracy of (16) (hence the optimization in (21)) compared to the simulated system evolution governed by the model equations (2)-(9), and to aid discussion on the value of n^* as a function of system parameters. Table I presents parameter values considered throughout the section.

TABLE I
CONSIDERED SYSTEM PARAMETERS. DEFAULT VALUES SHOWN IN BOLD.

Parameter	Value
Δt (s)	60
D ($\text{m}^2 \text{s}^{-1}$)	10^{-10}
M (molecules)	10^{15}
L (m)	10^{-3}
d (m)	0.1, 0.15, 0.2 , 0.25
α_1	20, 50, 100 , 200, 500
β	0.7
p	0.4, 0.7, 1 , 1.3, 1.6

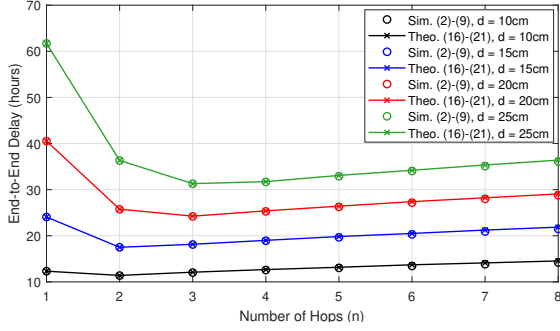


Fig. 3. T_a vs. n for varying d . All other parameters take their default values.

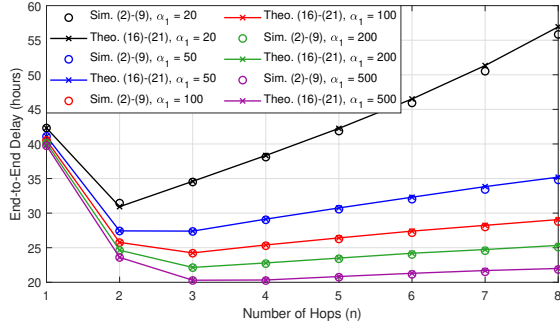


Fig. 4. T_a vs. n for varying α_1 . All other parameters take default values.

Figure 3 shows the end-to-end delay T_a versus n for different TX-RX distance (d) values. The results of Figure 3 confirm the existence of the trade-off between diffusion delay and processing delay (due to non-zero time to reach saturation) conjectured in Section IV. As expected, increasing distance increases diffusion delay, which causes the optimal number of hops, n^* , to shift rightward with increasing d .

In Figure 4, we consider T_a versus n for different gain (α_1) values. The results of Figure 4 demonstrate that n^* increases with increasing α_1 , and vice versa. A smaller α_1 implies that the system has a longer processing delay. As increasing n would further increase this delay, the results suggest that the system is better-off by avoiding larger n values, and settling for a larger diffusion delay to prevent α from further decreasing. We note that even though a plot with respect to β could not be provided due to space restrictions, our observations suggest that it has common qualitative results with α_1 .

Recalling $\alpha = \frac{\alpha_1}{n^p}$, a larger p implies a sharper decrease

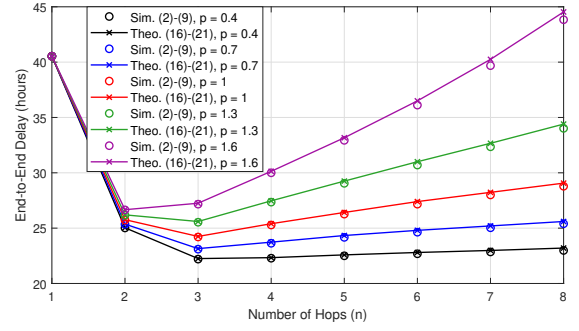


Fig. 5. T_a vs. n for varying p . All other parameters take their default values.

in α for the same n . Thus, increasing p increases the cost of introducing additional relays. The results of Figure 5 confirm this trend, as n^* shifts leftward with increasing p , which suggests that similar to Figure 4, the system favors incurring more diffusion delay to prevent α from sharply decreasing.

Lastly, it can be observed from Figures 3-5 that considering the overall system as a cascade of n individual one-hop links, and using Equations (16)-(21) to theoretically optimize the system yields a tight approximation to the actual end-to-end delay. We note that the approximation is particularly accurate when n is smaller, and is only slightly looser as n increases. This confirms the corresponding explanation in Section IV, as increasing n would both decrease $\frac{d}{n}$ and $\alpha = \frac{\alpha_1}{n^p}$, the two quantities which improve the accuracy when they are large. Overall, the approximate objective function accurately follows the trend of the true T_a , and can indeed be used to decrease complexity when optimizing n .

VI. CONCLUSIONS

This paper considers spatially separated bacterial colonies as nodes for relay-aided DBMC. A deterministic framework that characterizes system evolution has been provided for AFS-type relays. The framework has been extended to incorporate biological constraints such as a maximum saturation limit on emission intensity, yielding the *amplify-and-forward with saturation* (AFS) relay model. The end-to-end system has been approximated as a cascade of one-hop links for AFS-type relays, and a tractable objective function has been provided for numerical optimization. As future work, we consider extending our framework to cover stochastic or imperfect observations, chemical reactions within nodes, and physically implement an experimental relay-aided bacterial DBMC system.

APPENDIX A PROOF OF THEOREM 1

The proof follows from induction.

Base Step ($k = 1$): For this case, we need to show

$$\mathbf{H}^{0 \rightarrow 1} = \sum_{b=1}^1 \delta^b \sum_{c=1}^{\binom{1-1}{b-1}} \mathbf{H}_c^{(b)}, \quad \mathbf{H}_c^{(b)} \in \mathcal{H}^{(b)}[1],$$

which is equal to $\delta \mathbf{H}_1^{(1)}$, where $\mathbf{H}_1^{(1)} \in \mathcal{H}^{(1)}[1]$. The set $\mathcal{H}^{(1)}[1]$ has a single element in it by definition, which is $\mathbf{H}[1]$. Thus, we have $\mathbf{H}^{0 \rightarrow 1} = \delta \mathbf{H}[1]$, which confirms with (9).

Inductive Step: Here, we assume the theorem holds for all time indices up to $(k - 1)$, and prove that it holds for k . To do so, we first note that since we assume the statement is true for $(k - 1)$ and $\mathbf{r}[k]$ relates to $\mathbf{r}[k - 1]$ with δ , it is trivial to show that the (outer) summation over b in (11) has k summands (with maximum scalar multiplier δ^k). Thus, to prove Theorem 1, it is sufficient to show that $\forall b \in \{1, \dots, k\}$, the b th summand in (11), which is the summand with scalar multiplier δ^b , has the corresponding $\mathcal{H}^{(b)}[k]$ defined in (12).

- *Special case with $b = 1$:* For $b = 1$, we have $\mathcal{H}^{(1)}[k] = \{\mathbf{H}[k]\}$ by definition, which is trivially true since $\mathbf{r}[k]$ is only dependent on $\mathbf{H}[k]$ through $\mathbf{r}[0]$ (see (9)).
- *The general case with $b \in \{2, \dots, k\}$:* For an arbitrary $m \in \{b - 1, \dots, k - 1\}$, the $(b - 1)$ th summand of $\mathbf{H}^{0 \rightarrow m}$ can be written as

$$\delta^{b-1} \sum_c \mathbf{H}_c^{(b)}, \quad \mathbf{H}_c^{(b-1)} \in \mathcal{H}^{(b-1)}[m],$$

where the set $\mathcal{H}^{(b-1)}[m]$ contains all $(b - 1)$ -term combinations with time indices adding up to m .² Then, from (9), since $\mathbf{r}[m]$ appears in $\mathbf{r}[k]$ only through $\mathbf{H}[k - m]$, the path from time zero to k through m yields all (b) -term combinations with time indices adding up to k that has $\mathbf{H}[k - m]$ as the last multiplicand. Considering all $m \in \{b - 1, \dots, k - 1\}$ yields all b -term combinations (with scalar multiplier δ^b) that have $\mathbf{H}[k - m]$ as the last multiplicand for all $m \in \{b - 1, \dots, k - 1\}$. This is equivalent to all possible b -term combinations that add up to k , concluding the inductive step.

Lastly, to conclude the proof, we note that the cardinality of the set $|\mathcal{H}^{(b)}[k]| = \binom{k-1}{b-1}$, since by definition, Equation (12) is equivalent to the “occupancy problem” with k balls and b bins with no empty bins [31, Proposition 1.6].

REFERENCES

- [1] N. Farsad, H. B. Yilmaz, A. Eckford, C. B. Chae, and W. Guo, “A comprehensive survey of recent advancements in molecular communication,” *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, Feb. 2016.
- [2] B. Atakan and O. B. Akan, “On molecular multiple-access, broadcast, and relay channels in nanonetworks,” in *Proc. Int. Conf. Bio-Inspired Models of Netw. Inf. Comput. Sys.*, Nov. 2008, pp. 1–8.
- [3] A. Ahmadzadeh, A. Noel, A. Burkovski, and R. Schober, “Amplify-and-forward relaying in two-hop diffusion-based molecular communication networks,” in *IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.
- [4] J. Wang, M. Peng, Y. Liu, X. Liu, and M. Daneshmand, “Performance analysis of signal detection for amplify-and-forward relay in diffusion-based molecular communication systems,” *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1401–1412, Feb. 2019.
- [5] Z. Cheng, Y. Tu, J. Yan, and Y. Lei, “Amplify-and-forward relaying in mobile multi-hop molecular communication via diffusion,” *Nano Commun. Netw.*, p. 100375, Aug. 2021.
- [6] S. K. Tiwari and P. K. Upadhyay, “Estimate-and-forward relaying in diffusion-based molecular communication networks: Performance evaluation and threshold optimization,” *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 3, no. 3, pp. 183–193, Sep. 2017.
- [7] X. Wang, M. D. Higgins, and M. S. Leeson, “Relay analysis in molecular communications with time-dependent concentration,” *IEEE Commun. Lett.*, vol. 19, no. 11, pp. 1977–1980, Nov. 2015.
- [8] P. Hou, A. W. Eckford, and L. Zhao, “Analysis and design of two-hop diffusion-based molecular communication with ligand receptors,” *IEEE Access*, vol. 8, pp. 189 458–189 470, Oct. 2020.
- [9] A. Ahmadzadeh, A. Noel, and R. Schober, “Analysis and design of multi-hop diffusion-based molecular communication networks,” *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 2, pp. 144–157, Jun. 2015.
- [10] N. Varshney, A. Patel, W. Haselmayr, A. K. Jagannatham, P. K. Varshney, and A. Nallanathan, “Impact of intermediate nanomachines in multiple cooperative nanomachine-assisted diffusion advection mobile molecular communication,” *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4856–4871, Jul. 2019.
- [11] F. Walsh and S. Balasubramaniam, “Reliability and delay analysis of multihop virus-based nanonetworks,” *IEEE Trans. Nanotechnol.*, vol. 12, no. 5, pp. 674–684, Sep. 2013.
- [12] S. Balasubramaniam and P. Lio, “Multi-hop conjugation based bacteria nanonetworks,” *IEEE Trans. NanoBiosci.*, vol. 12, no. 1, pp. 47–59, Mar. 2013.
- [13] B. D. Unluturk, D. Malak, and O. B. Akan, “Rate-delay tradeoff with network coding in molecular nanonetworks,” *IEEE Trans. Nanotechnol.*, vol. 12, no. 2, pp. 120–128, Mar. 2013.
- [14] A. Einolghozati, M. Sardari, and F. Fekri, “Design and analysis of wireless communication systems using diffusion-based molecular communication among bacteria,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6096–6105, Dec. 2013.
- [15] —, “Relaying in diffusion-based molecular communication,” in *IEEE Int. Symp. on Inf. Theory (ISIT)*, Jul. 2013, pp. 1844–1848.
- [16] —, “Decode and forward relaying in diffusion-based molecular communication between two populations of biological agents,” in *IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 3975–3980.
- [17] O. S. Venturelli, R. G. Egbert, and A. P. Arkin, “Towards engineering biological systems in a broader context,” *J. Mol. Biol.*, vol. 428, no. 5, pp. 928–944, Feb. 2016.
- [18] N. Michelusi, J. Boedicker, M. Y. El-Naggar, and U. Mitra, “Queueing models for abstracting interactions in bacterial communities,” *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 584–599, Mar. 2016.
- [19] M. M. Vasconcelos, U. Mitra, O. Camara, K. P. Silva, and J. Boedicker, “Bacterial quorum sensing as a networked decision system,” in *IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.
- [20] A. Tamsir, J. J. Tabor, and C. A. Voigt, “Robust multicellular computing using genetically encoded nor gates and chemical ‘wires,’” *Nature*, vol. 469, no. 7329, pp. 212–215, Jan. 2011.
- [21] Y. Qian, H.-H. Huang, J. I. Jiménez, and D. Del Vecchio, “Resource competition shapes the response of genetic circuits,” *ACS Synth. Biol.*, vol. 6, no. 7, pp. 1263–1272, Jul. 2017.
- [22] M. Tei, M. L. Perkins, J. Hsia, M. Arcak, and A. P. Arkin, “Designing spatially distributed gene regulatory networks to elicit contrasting patterns,” *ACS synth. Biol.*, vol. 8, no. 1, pp. 119–126, 2018.
- [23] J. K. Kim, Y. Chen, A. J. Hirning, R. N. Alnahhas, K. Josić, and M. R. Bennett, “Long-range temporal coordination of gene expression in synthetic microbial consortia,” *Nat. Chem. Biol.*, vol. 15, no. 11, pp. 1102–1109, Nov. 2019.
- [24] S. Gupta, T. D. Ross, M. M. Gomez, J. L. Grant, P. A. Romero, and O. S. Venturelli, “Investigating the dynamics of microbial consortia in spatially structured environments,” *Nat. Commun.*, vol. 11, no. 1, pp. 1–15, May 2020.
- [25] R. H. Hsu, R. L. Clark, J. W. Tan, J. C. Ahn, S. Gupta, P. A. Romero, and O. S. Venturelli, “Microbial interaction network inference in microfluidic droplets,” *Cell Syst.*, vol. 9, no. 3, pp. 229–242, Sep. 2019.
- [26] S. Gupta, “Manipulating spatiotemporal variables to program coordinated behaviors in microbial consortia,” Ph.D. dissertation, University of Wisconsin-Madison, 2021.
- [27] A. Einstein, *Investigations on the Theory of the Brownian Movement*. Courier Corporation, 1956.
- [28] L.-S. Meng, P.-C. Yeh, K.-C. Chen, and I. F. Akyildiz, “On receiver design for diffusion-based molecular communication,” *IEEE Trans. Sig. Proc.*, vol. 62, no. 22, pp. 6032–6044, Sep. 2014.
- [29] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, “Channel modeling for diffusive molecular communication—a tutorial review,” *Proc. IEEE*, vol. 107, no. 7, pp. 1256–1301, Jun. 2019.
- [30] M. E. Hibbing, C. Fuqua, M. R. Parsek, and S. B. Peterson, “Bacterial competition: surviving and thriving in the microbial jungle,” *Nat. Rev. Microbiol.*, vol. 8, no. 1, pp. 15–25, Jan. 2010.
- [31] S. M. Ross, *A first course in probability*, 6th ed. Prentice Hall, 2002.

²For $m < b - 1$, no $(b - 1)$ -term combinations are present in $\mathbf{H}^{0 \rightarrow m}$.