Towards Practical and Scalable Molecular Networks

Jiaming Wang
jw27@illinois.edu
UIUC
United States

Sevda Öğüt
sevda.ogut@epfl.ch
EPFL
Switzerland

Haitham Al Hassanieh
haitham.alhassanieh@epfl.ch
EPFL
Switzerland

Bhuvana Krishnaswamy
bhuvana@ece.wisc.edu
UW-Madison
United States

ABSTRACT

Molecular networks have the potential to enable bio-implants and biological nano-machines to communicate inside the human body. Molecular networks send and receive data between nodes by releasing molecules into the bloodstream. In this work, we explore how we can scale molecular networks from a single transmitter single receiver paradigm to multiple transmitters that can concurrently send data to a receiver. We identify unique challenges in enabling multiple access in molecular networks that prevent us from using standard multiple access protocols. These challenges include the lack of synchronization and feedback, the non-negativity of molecular signals, the extremely long tail of the molecular channel leading to high ISI (Inter-Symbol-Interference), and the limited types of molecules that can be used for communication. We present MoMA (Molecular Multiple Access), a protocol that enables a molecular network with multiple transmitters. We introduce packet detection, channel estimation, and encoding/decoding schemes that leverage the unique properties of molecular networks to address the above challenges. We evaluate MoMA on a synthetic experimental testbed and demonstrate that it can scale up to four transmitters while significantly outperforming the state-of-the-art.

CCS CONCEPTS

• **Networks** → **Cyber-physical networks**; **Network protocol design**.

KEYWORDS

molecular communication, micro-implants, medium access control, code division multiple access

ACM Reference Format:
Jiaming Wang, Sevda Öğüt, Haitham Al Hassanieh, and Bhuvana Krishnaswamy. 2023. Towards Practical and Scalable Molecular Networks. In *ACM SIGCOMM 2023 Conference (ACM SIGCOMM '23), September 10, 2023, New York, NY, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3603269.3604881

ACM SIGCOMM '23, September 10, 2023, New York, NY, USA

ACM ISBN 979-8-4007-0236-5/23/09…$15.00
https://doi.org/10.1145/3603269.3604881

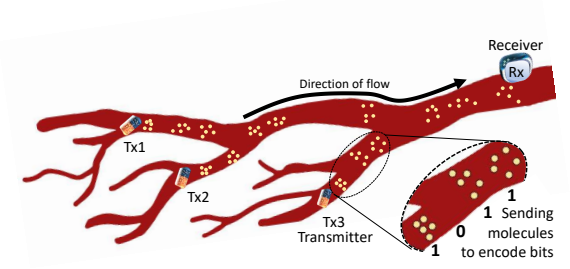


Figure 1: Molecular network with multiple transmitters.

1 INTRODUCTION

The Internet of Bio-Nano Things (IoBNT) promises to revolutionize medicine and healthcare [2]. It consists of biological computing machines such as micro and nano-implants that can collect sensor information inside the human body and coordinate monitoring and treatment. The past decade has witnessed huge leaps towards enabling this vision. Advances in bio-engineering, synthetic biology, and nanotechnology have led to biosensors that collect and process data [39, 41, 43], nano-scale Lab-on-a-Chip that run medical tests inside the body [24, 46, 60], the use of bacteria to design biological nano-machines that can detect pathogens [13, 29, 47], and all the way to nano-robots that can swim through the bloodstream to perform targeted drug delivery and treatment [11, 24, 48].

However, networking such bio-implants to deliver data to the labs-on-chip and communicate with nano-robots has gained limited attention. Traditional techniques like wireless radios, which work well for large implants such as pacemakers, defibrillators, and pill cams [52], cannot be scaled in form factor to micro and nano-dimensions [38, 49, 66]. Moreover, wireless signals do not propagate well in body fluids. As a result, molecular communication (MC) has emerged as the most suitable paradigm for networking nano-implants [1, 5–7, 32, 45]. The idea behind MC is to encode "1" and "0" bits by releasing molecular particles into the bloodstream. In its simplest form, one can encode "1" bit by releasing particles and a "0" bit by releasing nothing similar to ON-OFF-Keying in wireless networks. The small form factor and bio-compatibility of MC make it the most promising approach which led researchers to design bio-transceivers that can send and receive particles using synthetic cells or genetically engineered bacteria as well as biological circuits that use cells to emulate NAND and NOR logic gates. [12, 15, 40, 44, 50, 51, 56, 58].

Despite these advances in MC, networking research has been limited to mostly theoretical work [14] with very few experimental studies or setups [42, 63, 64]. While there is a long way to enable the deployment of such systems, this paper takes early steps towards more practical and scalable molecular networks. In particular, we focus on identifying the practical challenges in scaling molecular networks beyond a single transmitter and designing protocols that enable multiple molecular transmitters to send data to a central receiver as shown in Fig. 1.

Building multiple access protocols for molecular networks is difficult. First, it is hard to synchronize molecular transmitters due to the large propagation delay of molecules [34] and the added complexity of implementing a receiver in addition to the transmitter on the nano-implant as we explain in more detail in Sec. 3. As a result, schemes like TDMA (Time Division Multiple Access) cannot be implemented and packets can be transmitted at any time, i.e., while the receiver is decoding a packet, new packets might arrive. Second, the number of available distinct molecules that can be released by a biological nano-machine is typically limited, and designing a receiver that can distinguish many molecules is a challenging problem [29, 53, 57]. Hence, techniques like MDMA (Molecule-Division Multiple-Access) [18] where each transmitter uses a different molecule (similar to FDMA where each wireless radio uses a different frequency) cannot scale.

Third, the molecular signal is non-negative as it corresponds to the concentration of released particles. This makes it hard to directly implement techniques CDMA (Code Division Multiple Access) as signals only add up constructively and correlating with the CDMA code cannot cancel interference from other transmitters. In contrast, CDMA in wireless networks like cellular and GPS use "+1" and "-1" which sum up destructively to cancel interference. To address this, past work adopts Optical Orthogonal codes (OOC) [54, 68] used in fiber optic networks where the signal is also non-negative. However, as we explain in Sec. 8 and show in our results in Sec. 7, OOC does not work well in molecular networks. It requires very long codes which can significantly reduce the data rate and its highly unbalanced code creates large fluctuations in concentration making it hard to detect and decode new packets. Finally, due to the diffusion of molecular particles, the molecular communication channel has a very long tail which results in high inter-symbol-interference (ISI) [26]. Unlike, fiber optic networks which can compensate for ISI in hardware [21, 25] and still use CDMA albeit with long codes, high ISI is unavoidable in molecular networks which necessitates accurate estimation and compensation of the molecular channel.

In this paper, we present MoMA (Molecular Multiple Access), a protocol that enables multiple molecular transmitters to send their data to a central receiver that is capable of identifying and accurately decoding the data packets. MoMA opts for a CDMA-based scheme that uses Gold codes as it does not require synchronization and enables easy addressing of the different transmitters. However, MoMA modifies the code and packet structure and intertwines packet detection, channel estimation, and decoding of data bits to address the above challenges. Due to the non-negativity of the molecular signal and high ISI, failing to detect a single packet or inaccurately estimating a single transmitter's channel will bias the entire concentration of the received signal and completely corrupt the decoding. Hence, while decoding, MoMA constantly checks for new packets and re-estimates the channel.

To enable accurate packet detection, MoMA ensures that the signal power (concentration) throughout a data packet is stable and balanced while the power in the preamble fluctuates quickly. To do so, we modify the CDMA code to include only balanced codes i.e., the number of "1"s and "0" differ by at most 1. Moreover, as opposed to the standard approach [54, 68] of multiplying the code with the data bit which leads to no transmissions for "0" bits, MoMA takes an element-wise XOR of the code and the complement of the data bit, i.e., it sends the code as is for bit "1" and sends its complement for bit "0" which balances the power across the packet. For the preamble, MoMA repeats each element of the code a few times leading to a large sequence of pulse transmissions or a large sequence of no transmissions. Such fluctuations make the preamble easily distinguishable and allow the receiver to detect new incoming packets that start in the middle of previous packets being decoded.

To enable accurate channel estimation, MoMA jointly estimates the channels of all detected transmitters using adaptive filters. It also incorporates the unique properties of the molecular channel such as non-negativity and the long tail into the optimization function. Channel estimation is also used to further improve packet detection by rejecting falsely detected packets if their channel deviates too much within consecutive samples of the preamble or if deviates too far from the statistical model of the channel presented in Sec. 2, i.e., the channel cannot look random and cannot drastically change within the preamble.

Finally, MoMA introduces a second molecule to each transmitter i.e., each transmitter sends two molecules. Unlike, MDMA which requires N distinct molecules to support N transmitters, MoMA only requires two which is still practical [53, 57]. Using two molecules helps eliminate false positives and false negatives in packet detection, improves the accuracy of channel estimation, and increases throughput by sending different data streams on different molecules.

We implement and evaluate MoMA using a synthetic experimental testbed adapted from prior work [16, 63] and extended it to support multiple transmitters. The testbed, described in detail in Sec. 6, uses a network of tubes and pumps to emulate a constant flow of liquid with transmitters releasing particles into the liquid. Our results reveal the following: MoMA can scale up to four transmitters even in the case when the four packets from the transmitters always collide with random offsets. It can achieve a per transmitter throughput of around 0.89 bps which is 20× larger than prior work that uses OOC with CDMA for molecular networks [64] and 1.7× larger compared to a baseline that combines CDMA with MDMA. It also improves the BER (bit error rate) by 200× compared to [64] and by 10× compared to using OOC codes instead of MoMA's modified Gold codes. Moreover, MoMA's channel estimation improves BER by 10×, and the use of an additional molecule improves the packet detection rate by 20%.

Contributions: The paper has the following contributions

- We present MoMA, a medium access protocol for molecular networks that enable multiple unsynchronized transmitters to transmit their packets at any time to a receiver that is able to accurately decode colliding packets.

- We introduce packet encoding, packet detection, channel estimation, and packet decoding techniques that are customized to molecular networks and leverage the unique properties of the molecular channel.
- We build a synthetic experimental testbed with four transmitters and evaluate the performance of MoMA to show its ability to support multiple transmitters and demonstrate significant improvement in performance.

Limitations: We acknowledge that our current testbed might not capture all the challenges associated with designing protocols for molecular networks. In-vivo testing of micro-implants and microfluids in wet-labs is needed to achieve practical and deployable molecular networks. However, we take the first steps towards this vision and believe our insights for designing molecular networks will hold as the underlying diffusion and fluid dynamics models in our testbed are fundamental to molecular communication. Further discussion of limitations is provided in Sec. 9.

Ethical Issues: While the vision of IoBNT itself might invoke privacy and health concerns once deployed in practice, this paper focuses on exploring communication protocols and does not include any human subjects or potentially sensitive data. Hence, we believe that the technical work done in this paper does not raise any ethical concerns.

2 BACKGROUND

2.1 Molecular Communication Channel

As described earlier, the transmitter in molecular communication will encode data bits "1" and "0" by either releasing or not releasing particles similar to ON-OFF-Keying (OOK) in wireless communications where the transmitter encodes data bits by either turning its transmission ON or OFF. The particles propagate through the flowing liquid and are then detected and decoded by the receiver.[1]

The propagation of the released particles from the transmitter to the receiver is governed by three phenomena: advection, diffusion, and turbulence. Advection refers to the transport of molecules along a bulk flow. Diffusion refers to the random motion of particles as they collide with neighboring substances in the medium and randomly spread throughout space. Turbulence is motion caused by spatial or temporal changes in pressure and flow velocity. It usually appears as small-scale unpredictable flows and brings randomness similar to diffusion, but at a higher intensity [23].

Thus, the propagation of particles is usually described as the statistical behavior using Fick's Law and the advection-diffusion equation [37] as follows:

$$\frac{\partial C}{\partial t} + \nabla \cdot (vC) = \nabla \cdot (D \nabla C) + R \tag{1}$$

where $C(x,t)$ denotes the concentration of particles as a function of location x and time t. v denotes the velocity of advection, ∇ and $\nabla\cdot$ are respectively the gradient and divergence operators, and $R(x,t)$ denotes the input particles released by the transmitter at a certain location and time. Finally, D denotes the diffusion coefficient that describes how fast the particles spread out in the environment, which jointly quantifies diffusion and turbulence. It can be solved under different boundary conditions for various shapes of the channel. For example, assuming a point transmitter releases a pulse of K particles at location $x = 0$ and time $t = 0$ in an infinite 1-D environment, Eq. 1 becomes [17]:

$$\frac{\partial C}{\partial t} + \frac{\partial}{\partial x}(vC) = D\frac{\partial^2 C}{\partial x^2} + K\delta(0,0) \tag{2}$$

which has a closed-form solution[2] as

$$C(x,t) = \frac{K}{\sqrt{4\pi D t}} \exp\left\{-\frac{(x-vt)^2}{4Dt}\right\} \tag{3}$$

The above equation represents the channel impulse response (CIR) of the molecular communication channel. Fig. 2 shows an example of this CIR for two different flow velocities (v). As can be seen, the CIR has a very long tail resulting in significant inter-symbol-interference (ISI) between the transmitted data bits which severely limits the data rates in molecular networks [26].

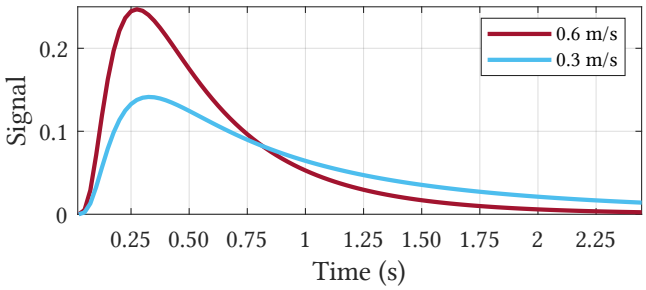


Figure 2: The channel impulse response in molecular communication for two different speeds of the flowing liquid.

Past work has also shown that this channel differs from channels in wired and wireless networks in several aspects: (1) It exhibits non-causal ISI, i.e., future symbols can also interfere with current symbols, (2) The coherence time of the channel is short relative to its delay spread, i.e., the channel changes fast and within the same packet transmission and (3) The channel has signal-dependent noise, i.e., transmitting more particles results in more noise in the system [63]. Past work has shown how to address these challenges for a single transmitter molecular communication system [63]. However, these challenges are exacerbated once we move to multiple transmitters as we will describe in the following sections.

2.2 CDMA Gold Codes

CDMA (Code Division Multiple Access) is a popular scheme used in wireless networks that allows multiple transmitters to transmit data at the same time by modulating their data bits with a unique binary code. The code should have good auto-correlation and cross-correlation properties, i.e., a transmitter's code c_i correlates very well with itself resulting in a peak if c_i is present in the signal. It also correlates very badly with the codes of other transmitters canceling out their signal. Ideally, the codes are orthogonal and the dot product $c_i \cdot c_j = 0$ if $i \neq j$. However, to achieve orthogonal codes, the length of the code L_c will be exponential in the number

[1] Prior work has proposed more complex MC encoding schemes like concentration shift keying [31] which is similar to pulse amplitude modulation (PAM) in wireless networks or even encoding huge amounts of data into protein and DNA structures [15, 51, 69]. However, these schemes are yet to be implemented and so for the purpose of this paper, we focus on OOK since it is the simplest and likely the most practical approach.

[2] This solution also assumes a passive receiver, i.e., the receiver does not absorb or destroy the particles [33].

of needed codes G, i.e., $L_c = 2^G$. This exponentially reduces the data rate and is typically avoided in practice [20].

Gold codes are binary codes that are not perfectly orthogonal but still maintain good correlation properties while having a code length L_c that is linear in the number of codes G. As a result, they are widely used in CDMA systems, e.g. GPS [22]. Gold codes are generated using linear-feedback shift registers of size n that generate $G = 2^n + 1$ codes of length $L_c = 2^n - 1$ [19]. The codes have high auto-correlation: $c_i \cdot c_i = L_c$ and low cross-correlation:

$$\max_{i,j,i\neq j} c_i \cdot c_j = \begin{cases} 2^{(n+2)/2} + 1 & \text{if } n \text{ even} \\ 2^{(n+1)/2} + 1 & \text{if } n \text{ odd} \end{cases} \propto \sqrt{L_c} \tag{4}$$

In a set of Gold codes about half of the codes are balanced, i.e., the number of +1s and −1s in the code differ by at most 1. For example, the set of Gold codes with $n = 3$ are listed below where only the first 3 are balanced.

$$\begin{aligned} c_0 &= -1,+1,+1,+1,+1,-1,-1 & c_1 &= -1,+1,-1,+1,-1,+1,+1 \\ c_2 &= +1,-1,+1,+1,-1,-1,+1 & c_3 &= +1,+1,-1,-1,-1,-1,-1 \\ c_4 &= -1,-1,+1,-1,-1,+1,-1 & c_5 &= +1,+1,+1,-1,+1,+1,+1 \\ c_6 &= -1,-1,-1,-1,+1,-1,+1 \end{aligned} \tag{5}$$

It is worth noting that Gold codes have poor performance for any n that is a multiple of 4. Hence, we are limited to using $n = 3, 5, 6, 7, 9, \cdots$ Finally, the individual bits of the code $c_i[m] = \pm 1$ for $m \in [0, L_c - 1]$ are referred to as chips.

3 OVERVIEW AND CHALLENGES

MoMA's goal is to enable multiple transmitters in molecular networks. We consider the most common case where multiple micro-implants transmit data to a central receiver [4, 14]. The receiver can potentially be a larger implant that is more powerful. For the purpose of this paper, we only consider one-way communication from the micro-implants to the larger implant in which case the micro-implants can be as simple as possible or can even be implemented using biological cells [12, 50, 51]. We also assume that the receiver is placed downstream in the direction of blood flow making it easier for the particles to propagate from the micro-implants to the receiver. We adopt ON-OFF-Keying as the modulation scheme by releasing particles to encode a "1" bit and releasing nothing to encode a "0" bit which also emphasizes our goal to keep the transmitters simple and push the complexity to the receiver.

Achieving multiple access in such molecular networks requires addressing the following challenges:

- **Non-negative signal:** Unlike typical wired and wireless signals, molecular communication signals are non-negative as they represent the concentration of particles. Hence, it is not possible to encode "+1" and "−1" chips of the CDMA code in a molecular signal and we must resort to using "1" and "0" instead. To understand why this is a problem, consider how CDMA works in today's wireless networks. Once we correlate the received signal with the code of a transmitter c_i, the "+1" and "−1" chips sum up destructively to cancel the signal from other transmitters. As a result, we can decode the packet of transmitter i while treating all the other signals as noise. This is no longer true when we transmit "1" and "0". The interference from other transmitters is always non-negative and will bias the decoder to think that the transmitter i is releasing particles even when it is not, leading to major errors in decoding. This problem is made worse when (1) the interfering transmitters have stronger signal strength (i.e. better CIR) and (2) when the decoder does not detect the packets of other interfering transmitters. Hence, it is not possible to decode packets independently and the decoder must detect all the transmitted packets as well as estimate their CIR in order to be able to decode the packets.
- **Lack of Synchronization:** Since the transmitters in a molecular network are not synchronized and cannot receive feedback from the receiver, we cannot implement typical multiple access schemes like TDMA. Furthermore, the transmitters are likely to transmit their packets at any time, i.e., a new packet might arrive while the receiver is already in the middle of decoding other packets. Simply discarding such packet collision will lead to further degradation in what is already a relatively low data rate. Note that even if one were to afford the added complexity and equip micro-implants with receivers such that they can get feedback from a central implant, achieving synchronization is still difficult since (1) unlike wireless signals that travel at the speed of light, the propagation delay of molecular signals is very large, on the order of multiple transmitted bits and (2) it is not possible to diffuse data upstream in the opposite direction of blood flow. Having the particles loop around (through the heart, lungs, etc.) to reach the micro-implant will lead to an even larger propagation delay and very low signal strength.
- **High ISI:** As can be seen from Fig. 2, the molecular channel has a very long tail that decays very slowly. As a result, molecular communication systems suffer from very high inter-symbol-interference (ISI) which further complicates the design and makes it hard to decode signals from different transmitters independently using CDMA. It also makes it difficult to decode without accurately estimating the channel of each transmitter.

The above challenges make accurate packet detection and channel estimation very difficult and at the same time essential for accurate decoding. In the following sections, we describe in detail how MoMA addresses these challenges to scale molecular networks to multiple transmitters.

4 MOMA PROTOCOL

In this section, we describe how the MoMA transmitters encode their data packets to enable multiple access. As explained earlier, MoMA uses a CDMA-based multiple access scheme where each transmitter is assigned a CDMA code that is used to encode its data bits. The code also serves as a means to address (identify) different transmitters. Transmitters are not synchronized and can transmit at any time. Hence, their packets can collide with random offsets. To enable accurate packet detection at the receiver even when packets from many MoMA transmitters collide, MoMA ensures that the preamble demonstrates significant fluctuations in power whereas the data portion of the packet is balanced with a stable power across the data. MoMA also uses two molecules per transmitter to increase the likelihood of correctly detecting the packet. Below we describe in detail how the MoMA transmitter encodes its packets.

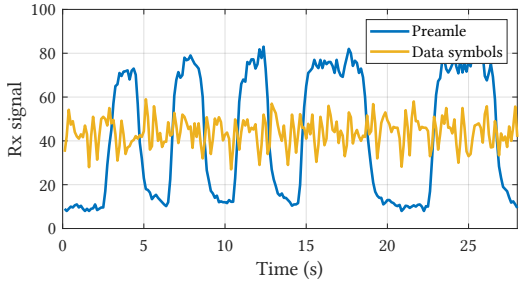


Figure 3: Fluctuation in power in the preamble vs. the data symbols.

4.1 Codebook

The codebook of MoMA is constructed from Gold codes described in Sec. 2. We only use Gold codes that are balanced, i.e., the number of "1" and "-1" in the code differ by at most one. In general, for parameter n, less than half of the $G = 2^n + 1$ generated codes are balanced. Hence, if we have N transmitters, we will generate Gold codes with parameter $n = \lceil \log_2(N+1)+1 \rceil$. However, as mentioned in Sec. 2, Gold codes have poor performance when n is a multiple of 4 which is the case when $4 \leq N \leq 8$. In this case, using $n = 5$ would generate $G = 33$ codes of length $L_c = 31$ which will unnecessarily cut the data rate in half. Instead, we use $n = 3$ to generate codes of length $L_c = 7$ and append each code with a Manchester code such that every sequence becomes perfectly balanced. This results in codes of length $L_c = 14$ instead of 31. Finally, the chips of each code are represented as "1" and "0" instead of "+1" and "-1" to encode releasing molecules or releasing nothing.

4.2 Packet construction

The i-th transmitter is assigned a code $c_i = [c_i[0], \ldots, c_i[L_c - 1]]^T$. The preamble p_i of each packet has the length $L_p = R \times L_c$, which is generated by repeating each chip of the code R times. Let $\vec{1}_R^T$ be a vector of all "1" of length R. Then,

$$p_i = \left[c_i[0]\vec{1}_R^T, \ldots, c_i[L_c - 1]\vec{1}_R^T \right]^T \tag{6}$$

The consecutive pulses of "1"s or "0' create large fluctuations in signal power making it easier to detect the preamble.

The data symbols d_i are encoded by taking the element-wise XOR between the code and the complement of the data bit as follows:

$$d_i = \begin{cases} c_i \oplus \vec{0}_R^T & \text{if data bit is 1} \\ c_i \oplus \vec{1}_R^T & \text{if data bit is 0} \end{cases} \tag{7}$$

Effectively, we send the code if the data bit is 1 and we send the complement of the code if the bit is 0. Unlike past work that sends nothing if the bit is 0, MoMA's approach ensures that the power is balanced across the packet which improves both decoding and packet detection. Fig. 3 shows the power (concentration) of the received signal for both the preamble and the data portion of the packet when $R = 16$. As can be seen, the consecutive "1"s in the preamble lead to a buildup of concentration of molecules at the receiver, and the consecutive "0"s lead to a sharp drop creating large changes in power. On the other hand, the power in the data symbols is stable and does not vary drastically as the number of "1"s and "0"s are balanced, and due to high ISI, there is not enough buildup or drop in concentration.

Two points are worth noting. First, the total power of the preamble and the data symbols is the same, i.e., we are not sending the preamble at a higher power. We are simply rearranging the 1s and 0s to create a larger fluctuation in power. Second, one might assume that using the complement to send the 0 bit would hurt the decoding as sending nothing for L_c chips would cause a noticeable drop in concentration. However, note that this figure shows a single transmitter. When the signals from multiple transmitters combine, such large fluctuations in data symbols make it harder to decode as we show in our results section.

4.3 Multiple molecules

MoMA uses multiple molecules to help improve packet detection, decoding, and throughput. One option is to use multiple molecules, instead, to perform MDMA (Molecule Division Multiple Access) where each transmitter uses a different molecule and none of the packets interfere. However, as described earlier, the number of possible types of molecules is limited, and having a single receiver to decode all types is challenging [29, 53, 57]. Another approach is to combine MDMA with CDMA. Suppose we have N transmitters and use M molecules, we divide the transmitters into groups of N/M transmitters and use a CDMA code within each group. The packets from the different groups do not interfere. This is effectively similar to reducing the size of the network by $M\times$ which reduces the length of the code by $M\times$ and increases the data rate by $M\times$. We refer to these two approaches as MDMA and MDMA+CDMA. We compare with them in our results in Sec. 7 to show that MoMA's approach yields better performance.

MoMA's approach is instead to have each transmitter use multiple molecules. Each transmitter is assigned a different code for each molecule and the assignment is legal as long as any two transmitters do not share the same code on the same molecule. An example of a legal assignment would be: transmitter i can use c_1 on molecule 1 and c_3 on molecule 2 whereas transmitter j uses c_6 on molecule 1 and c_1 on molecule 2. Besides, each transmitter can send different data streams on different molecules which also increases the data rate by $M\times$. While there is no gain in data rate compared to using $MDMA + CDMA$, MoMA is able to achieve higher overall throughput and better scaling for two reasons: (1) Using two molecules significantly improves packet detection as the probability of missing the packet on multiple molecules decreases exponentially with the number of molecules. (2) Different codes might have different performance depending on the channel impulse response and the underlying data. Since the codes cannot be changed after deployment, having a bad code-channel combination can significantly harm the data rate of a transmitter. Using two codes on two molecules significantly reduces the probability of such a bad combination on both molecules.

It is worth noting that MoMA's approach can yield further scaling gains if we allow some transmitters to share the same code on the same molecule. As long as, they do not share the same code on all molecules, it is still possible to distinguish and decode the data from the transmitters. For a codebook of size G, this allows us to scale the possible number of transmitters from $O(G)$ to $O(G^M)$. We explain this possibility for further scaling in Appendix B in more detail and present preliminary results that show that it is possible

to distinguish and decode transmitters if they use the same code on one of two molecules.

5 MOMA DECODER

In this section, we will explain how the MoMA receiver is able to decode the colliding packets from the transmitters. Since packets can arrive at any time, the MoMA receiver must constantly keep checking for new packets while at the same time updating its estimates of each transmitters' CIR and decoding the data bits. To do so, MoMA uses a sliding window design. In each window, the decoder (1) performs packet detection to discover new packets; (2) estimates the CIR of new packets and updates CIR of already detected packets; (3) and decodes the data symbols. It does this for each molecule and combines measurements across molecules to improve packet detection. However, as previously mentioned, packet detection, channel estimation, and decoding are quite intertwined as we will describe in detail in the following subsections.

5.1 Packet Detection

Packet detection is particularly important. As we will show in Sec. 7, missing detection of a single packet will lead to an exploding BER. Hence, we opt for packet detection that favors false positives (detecting a packet that is not there) over false negatives (missing a packet). We can then use multiple molecules and channel estimation to eliminate detection errors. For simplicity, we will describe it for a single molecule and then extend it to multiple molecules. The packet detection process works as follows. A full pseudocode can be found in Appendix. A. We maintain a list of transmitters detected in the previous windows S_d. In each window:

Step 1. Initialize an empty list S_c of newly detected transmitters.

Step 2. Decode the data of transmitters in $S_d \cup S_c$ assuming no new packets will arrive in this window.

Step 3. Update the CIR of each transmitter and use it along with the decoded bit sequence to reconstruct the received signal y_d of these transmitters.

Step 4. Subtract the reconstructed signal y_d from the actual received signal y to get the residual signal y_r which might contain undetected transmitter packets.

Step 5. For each transmitter i that is not in $S_d \cup S_c$, compute the correlation of its preamble p_i with the residual signal y_r and find the peak of the correlation. If the peak is in the window of the preamble length, then a new packet might be present which indicates that the decoding in Step 2 was wrong.

Step 6. Iterate between (1) estimating the CIR of all TXs in $S_d \cup S_c \cup \{i\}$ and (2) decoding all these TXs' data. The iteration terminates when the decoding converges, i.e., we decode the same bits every time.

Step 7. Divide the preamble of the packet of transmitter i into two halves and estimate two CIRs of i from each half. If the two CIRs pass the similarity test described below, add transmitter i to S_c as a newly detected transmitter.

Step 8. Loop back to Step 5 for the next transmitter that exhibits a peak in correlation.

Step 9. If no more new transmitters were added to S_c, add S_c to S_d and move to the next window. Otherwise, loop back to step 2.

In the above packet detection scheme, the similarity test in step 7 aims to remove false positives. It is based on the fact that the CIR should not change drastically in a preamble period. After computing the CIR using the first and second half of the preamble, we compute the ratio of the total power of the two estimates. We then compute the correlation coefficient of the two estimates. The test fails if the either the correlation or power ratio are below a threshold since the CIR should follow the model in Sec. 2 and should not look random.

To extend the above protocol to multiple molecules, we run the entire process in parallel on each molecule but we average the peaks across molecules in step 5 and average the correlation coefficient in the similarity test in step 7. This reduces the probability of missing packets in step 5 and improves our ability to reject false positives.

Finally, as can be seen, the above packet detection is intertwined with channel estimation and decoding which we will describe in the following subsections.

5.2 Channel Estimation

In traditional systems like wireless, channel estimation typically takes place after detecting a new packet by using the preamble. However, past work has shown that the channel coherence time in MC is on the same order as the delay spread (the tail of the channel) [63]. Hence, the channel must be re-estimated and updated regularly throughout the packet to ensure accurate decoding. In the context of multiple access, this means that we cannot rely on CIRs estimated in the previous window and we need to re-estimate the CIR in every window as shown in the previous section. Moreover, since the received signal is the sum of packets from different transmitters, we cannot estimate the CIR of each transmitter independently, i.e., we must jointly estimate the CIR of all transmitters.

We estimate the channel for each molecule independently. In Appendix B, we show how we can leverage multiple molecules to improve the channel estimation. Let y be the received signal, h_i and x_i be the channel impulse response and transmitted signal of transmitter i respectively, and n the noise vector. For simplicity, we will sum in the below equations over all transmitters $i \in [1, \cdots, N]$. However, in practice, we should only sum over the transmitters that were detected during packet detection, i.e., transmitters in the set $S_d \cup S_c$ from the previous section. Then,

$$\begin{aligned} y &= \sum_{i=1}^{N} h_i * x_i + n = \sum_{i=1}^{N} X_i h_i + n = Xh + n \\ \text{where} \quad X &= [X_1, \ldots, X_N] \\ h &= \left[h_1^T, \ldots, h_N^T\right]^T \end{aligned} \tag{8}$$

Although linear matrix inversion can be directly used to compute the least squares channel estimate from Eq. 8, it does not account for the unique features of the molecular channel and leads to suboptimal decoding performance as we show in Sec. 7. To account for the characteristics of the MC channel, we formulate channel estimation as an optimization problem whose loss term is composed of the following:

- **Least Squares loss.** This is the term that directly corresponds to the linear inversion of Eq. 8, which describes the distance between the actual received signal and the expected received signal constructed from the estimated CIR. Suppose the length

of the signal y is L_y, then

$$\mathcal{L}_0 = \frac{1}{L_y}\|y - Xh\|^2 \tag{9}$$

- **Non-negativity loss.** This arises from the fact that the molecular signal is fundamentally the concentration of particle, which is strictly non-negative. Hence, we penalize the negative terms in the estimated CIR. Suppose the length of CIR h_i is L_h, then

$$\mathcal{L}_1 = \frac{1}{L_h}\sum_{i=1}^{N}\|\text{ReLu}(-h_i)\|^2 \tag{10}$$

- **Weak head-tail loss.** As can be seen from Fig. 2, the CIR of a molecular channel should start with a weak head and end with a weak tail. Thus, we penalized the non-zero taps that are far away from the peak CIR sample. Let $a \odot b$ denote the element-wise product of vectors a and b, then

$$\begin{gathered}\mathcal{L}_2 = \frac{1}{L_h^2}\sum_{i=1}^{N}\|g_i \odot h_i\|^2 \\ \text{where } g_i = [1 - q_i, \ldots, L_h - q_i]^T \\ \text{and } q_i \text{ is the index of the peak sample}\end{gathered} \tag{11}$$

- **CIR similarity loss.** CIRs of the molecular network reveal high correlation when they share the same molecule diffusion coefficient and/or the same transceiver distance. Defining $x' = x/D$ and $t' = t/D$, the diffusion model Eq. 2, can be simplified as

$$\frac{\partial C}{\partial t'} + \frac{\partial}{\partial x'}(vC) = \frac{\partial^2 C}{\partial x'^2} + DK\delta(0,0) \tag{12}$$

This indicates that the CIR $C(x,t)$ can be achieve from another one $C(x',t')$ by scaling in the amplitude and the time scale. Theoretically, the channel estimation can include a step of parameter fitting that unifies the common parameter in the estimated CIRs and improves the performance. However, an accurate channel model is hard to generalize to the various practical molecular environments. But, when the diffusion coefficients D of the molecules do not diverge much, the two CIRs can be compared directly with only scaling in amplitude. Thus, we only take an preliminary check among CIRs of the same transmitter and penalize those deviating from the average in this work. We introduce another loss term, similarity loss ($\mathcal{L}_3$), in the channel estimation algorithm. Note that this loss term, unlike the other loss terms mentioned above, is only applicable to multi-molecule channel and cannot be computed separately for each molecule.

$$\begin{gathered}\mathcal{L}_3 = \frac{1}{L_h}\sum_{i=1}^{N}\sum_{j=1}^{M}\left\|\vec{h}_{ij} - a_{ij}\vec{h}_i^a\right\|^2 \\ \text{where } \vec{h}_i^a = \frac{1}{M}\sum_{j=1}^{M}\frac{\vec{h}_{ij}}{a_{ij}}\text{, and } a_{ij} = \|\vec{h}_{ij}\|\end{gathered} \tag{13}$$

By adding the above loss terms with different weights W_1, W_2, we formulate the molecular channel estimation problem as the below optimization. [3]

$$\min_{h_i, i=1,\ldots,N} \mathcal{L}_0 + \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3 \tag{14}$$

[3]The weights of the loss terms are not perfectly tuned in this work, but might lead to potential performance improvement.

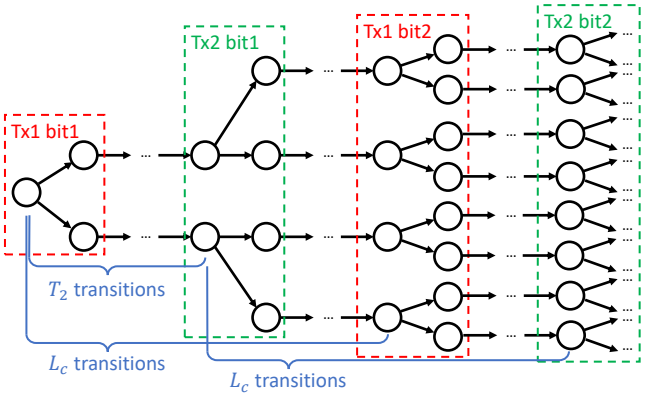


Figure 4: Chip-based Viterbi state transitions for two transmitters. L_c is the length of the coding, and T_2 is the time offset (in the unit of chips) between the two transmitters. The omitted transitions only have one outgoing state.

MoMA solves the above optimization through an adaptive filtering algorithm using iterative gradient descent. The adaptive filter is initialized with the Least Squares solution to Eq. 8 which can also be used to initialize q_i in Eq. 11. Finally, once the adaptive filter converges, we can also compute noise power n by comparing the reconstructed Xh and actual value y of the received signal. We can then use the noise power for decoding in the following section.

5.3 Viterbi Decoder

We use a Viterbi algorithm to decode the data bits which we modify to account for multiple transmitters, the CDMA codes, and the MC channel. The Viterbi algorithm [61] is a dynamic programming algorithm for obtaining the maximum a posteriori probability estimate of the most likely sequence of hidden states—-called the Viterbi path-—that results in a sequence of observed events, typically in the context of Hidden Markov Model (HMM). In general, the HMM for MC transmission has the following components:

- Hidden Markov Chain: The state of the Markov Chain is a sequence of consecutively transmitted bits whose particles remain in the channel and influence the received signal. The length of the sequenced is decided by the length of ISI, and the transition between states happens when a new data bit is sent by the transmitter. All the possible transitions construct a large-scale Viterbi trellis shown in Fig. 4, and the algorithm aims to find the path with the highest probability.
- Observed events: The received signals serve as the observations of the HMM. First, an expectation is computed from the current state of the HMM and the CIR estimated from the preamble. Then, the probability of the actual observation is computed with its divergence from the expectation and the noise power in the system. Such probability accumulates as the Viterbi path prolongs as more receiver signals are collected with time.

For the case with a single transmitter and a single receiver without coding or oversampling (e.g. [63]), the state of the HMM is a sequence of data bits each with one sample of the receiver signal as observation. With CDMA codes and multiple transmitters, we modify the model by changing the number of observations for each state to the length of the code. However, this HMM based on data

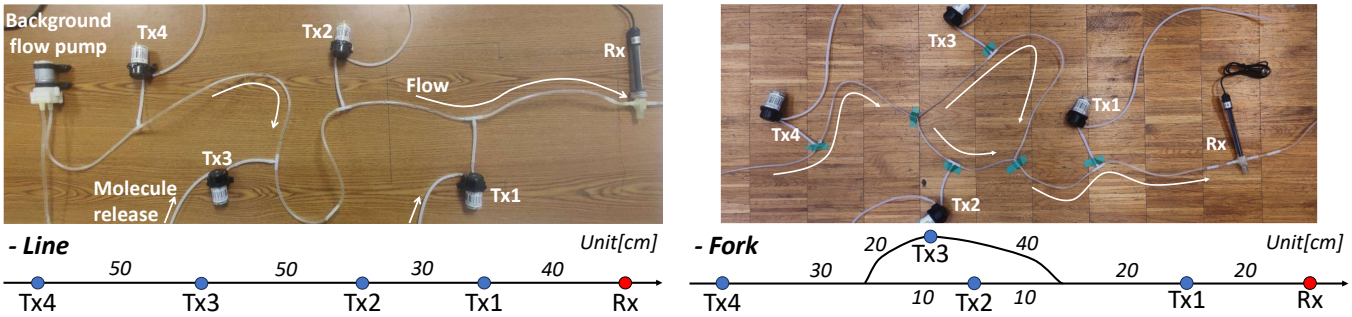


Figure 5: A synthetic liquid testbed with four transmitters and one receiver. Left is a line topology, while right is a fork topology. (Circuitry of the testbed is not shown for clarity.)

bits is not accurate to describe the MoMA, due to the lack of synchronization among multiple transmitters. Considering that the packets could arrive at the receiver with offsets, MoMA uses one or more sequences of chips (instead of data bits) to represent the HMM state, where each sequence represents one of the detected packets. With chip-rate sampling, each state still has one receiver sample as the observation. The major difference from single transmitter without CDMA is that not all HMM states will transition to two different states. From the perspective of each transmitter, such transition only happens when the first chip of the data symbol comes into the state sequence, while for the other states the transition is deterministic according to the CDMA code. From the perspective of all transmitters, there is the possibility that one HMM state could transit to more than 2 states (i.e. power of 2) when multiple transmitters are coincidentally synchronized at symbol level. In general, the multiplication of Viterbi states for each transmitter has a period equal to the CDMA code length, but with random delays.

6 TESTBED

We built a synthetic experimental testbed, as shown in Fig. 5, to evaluate MoMA. This testbed is centered around channel with either a single path or a fork in the middle, where on the one side a background flow pump continuously pumps water through the tube to the receiver on the other side. In addition, four other tubes are interconnected with the mainstream at various distances from the receiver, each of which has a pump that can inject bursts of information molecule solution into the mainstream. These four pumps are controlled electronically with transistor circuits and serve as the communication transmitters in the molecular network. Two different channels are evaluated (the line channel and the fork channel, which are presented in Fig. 5), while most of the results in Sec. 7 are under the line channel. The fork-channel figures, as a support, will be marked out in the title.

In this testbed, we use *NaCl* as the information molecule, and we use an Electric Conductivity (EC) reader that indicates the *NaCl* concentration in the solution as the communication receiver. The Tx pumps and the Rx EC reader are controlled by an Arduino Mega 2560 REV3. Since most of the common soluble molecules will change the electric conductivity with high possibility, which means that introducing them to will infer the measurement of *NaCl*, we cannot conduct real-world experiments with two types of molecules in the current stage. However, we will discuss testbeds using other molecules as well as possible methods to implement a multi-molecule testbed in Sec. 9. Alternatively, we use emulation to evaluate multiple molecules. For each data point reported in the paper, we repeated the experiment of one molecule 40 times with different data streams and code assignments. To emulate two molecules, we randomly pick two experiments of the same transmitters and concurrently process them, which assumes that the two molecules are not interfering. Each data point of the two molecules include 500 such emulations.

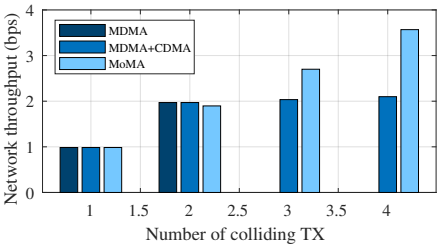
7 RESULTS

7.1 Main Results

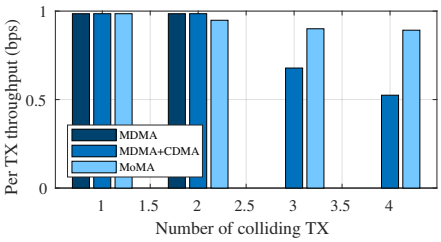
We evaluate MoMA in a network with 4 transmitters, and each transmitter can emit 2 types of molecules. To demonstrate the scalability of MoMA, we compare the total network throughput and per transmitter throughput as we intentionally cause different number of transmitters to collide. We compare MoMA with the following two baselines.

- MDMA (Molecule-Division Multiple-Access): Each transmitter only uses a distinct type of molecule, and we assume that the receiver can independently measure each type of molecule. Under this scheme, OOK (On-Off Keying) can be used to encode data symbols and pseudo-random sequences as the preambles. Note that MDMA requires the number of usable molecules to be greater than or equal to the number of transmitters.
- MDMA+CDMA: When there are more transmitters than the available types of molecules, one has to combine MDMA with CDMA. We first evenly divide all transmitters among the molecule categories and then assign distinct CDMA codes to different transmitters using the same molecule.

Since these two baselines can be viewed as special cases of MoMA, we use the same decoder, and we assume that the receiver drops packets with BERs greater than 0.1. For the fairness of throughput comparison, we normalize the transmission data rate to 2/1.75 bps for all schemes. In particular, (1) for MDMA, each transmitter uses 1 molecule and the symbol interval is 875 ms, (2) for MDMA+CDMA, each transmitter uses 1 molecule and the CDMA code length is 7 with a chip interval of 125 ms, (3) for MoMA, each transmitter uses 2 molecules and code length is 14 with a chip interval of 125 ms. Besides, we also ensure that the preambles introduce the same overhead in all schemes. To do so, we ensure that the length of the preamble is equal to 16 times the length of the data symbol and each packet encodes 100 data bits. One thing to note is that the reported



(a) Total network throughput.



(b) Per Tx throughput.

Figure 6: Throughput of molecular network with different number of transmitters and multiple access protocols. The throughput is the lower bound for each transmission, as all transmitters are intentionally made to collide with different offsets.

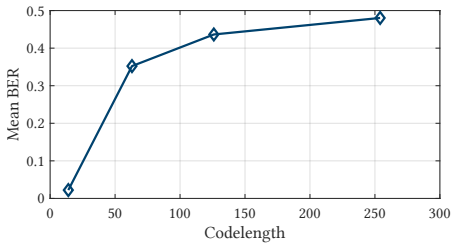


Figure 7: BER with different code lengths but the same data rate.

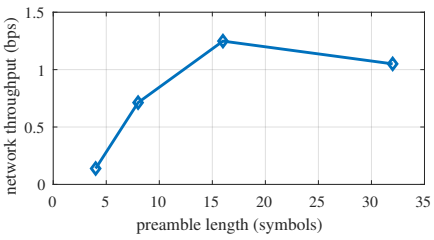


Figure 8: Network throughput with different preamble lengths.

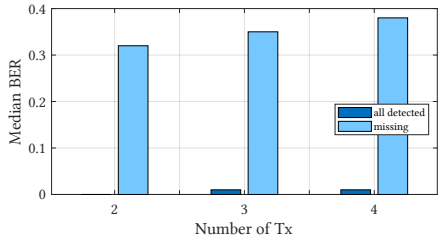


Figure 9: BER comparison with and without miss-detected packets.

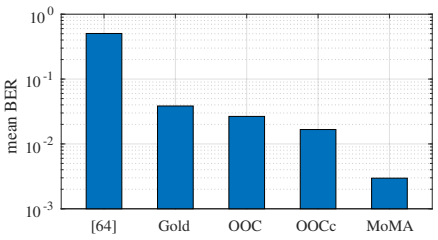


Figure 10: BER comparison against baseline coding schemes.

throughput is NOT the capacity of the network, since the number of transmitters in the network does not reach the maximum size of the codebook.

Fig. 6 shows the overall and per transmitter throughput of the network as the number of actively transmitting and colliding transmitters varies from 1 to 4. When there are no more than two active transmitters, MDMA provides the highest throughput of 0.99 bps per transmitter. However, it cannot support more than 2 transmitters since the system is restricted to 2 different molecules. Even though MDMA+CDMA can support up to 4 transmitters, the per transmitter throughput drastically decreases when two transmitters share the same molecule. This is because MDMA+CDMA fails to detect colliding packets carried by the same molecules, which causes packet loss. On the contrary, the throughput of MoMA is able to scale with the number of transmitters with much less loss, and it still achieves 0.89 bps per transmission with 4 colliding transmitters, which is 1.7x the throughput of MDMA+CDMA.

7.2 Micro benchmarks

7.2.1 Impact of code length on decoding BER.
Fig. 7 shows the average BER when the code length changes but the data rate is fixed. The BER increases along with the length of the code since longer code results in more inter-symbol interference. Taking this into consideration, MoMA uses the shortest code possible when the codebook is large enough to support the required number of transmitters.

7.2.2 Impact of preamble length on throughput.
Fig. 8 shows the total network throughput achieved by MoMA with different preamble lengths. All four TXs transmit and collide using only one molecule at a data rate of 1/1.75 bps. The network throughput first increases along with the preamble length until the preamble reaches 16x the symbol length. This is because the additional preamble sequences improve packet detection and channel estimation accuracy. However, these benefits saturate at 16x symbol length and the more overhead starts to harm the throughput.

7.2.3 Importance of detecting colliding packets.
In Fig. 9, we compare the BER when all colliding packets are correctly detected and when some of them are missing, which uses the same experiments as MoMA with 2/3/4 transmitters in Fig. 6. The median BER only considers the transmissions that are still correctly detected. It is obvious that incorrect detection of any colliding packets results in a disastrous BER in the decoding of the other detected packets. As a consequence, almost all packets are dropped because the BERs are above 0.3. Based on this result, we prioritize packet detection when designing MoMA.

7.2.4 Comparison of coding schemes.
We compare five decoding schemes of two categories: (1) individual threshold decoder on directly correlating the receiver signal with the code [64] and (2) MoMA's joint decoder. For MoMA's joint decoder, there are four possible combinations of two coding choices, i.e., OOC and MoMA's code, and two representations of bit -1, that are transmitting nothing and complementary code. The test case had increasing number of colliding packets, each with 100 data bits, with code length 14 and chip interval 125 ms. In this comparison, we adopt a set of (14,4,2)-OOC codes as specified in [9], which also has code length 14. However, each code in this set has four +1s and the maximum cross-correlation is 2. To isolate the impacts of different coding schemes from all the other aspects, e.g., packet detection and channel estimation, we assume that the receiver knows the exact packet arrival time and the exact CIR of every packet. To obtain the ground truth CIR, we assume we know all the transmitted bits and use all of them to estimate the CIR as opposed to only using the preambles to estimate CIR. Fig. 10 shows the BER averaged over 40

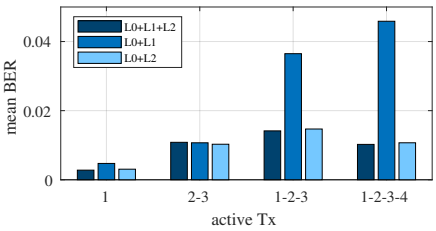
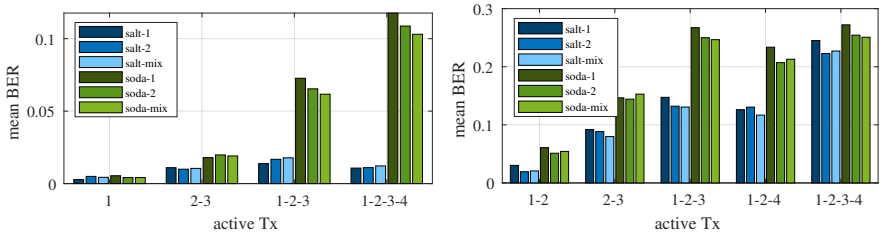


Figure 11: BER of using different combined loss terms for channel estimation. Single-molecule results are presented, so similarity loss $\mathcal{L}_3$ does not apply. ToA is assumed known.



(a) Line channel **(b) Fork channel**

Figure 12: BER of single-molecule experiments versus double-molecule emulations. ToA is assumed known.

different traces. Based on the first bar in the figure, the chosen OOC code cannot provide good correlation properties to address ISI and MAI (multiple access interference) issues. Using a longer OOC code set might handle the issue, but it cannot achieve a comparable data rate as MoMA. From the other bars in the figure, we can see that MoMA achieves the best possible decoding among all four coding schemes. Besides, using complementary code for bit -1 can also improve the performance of OOC.

7.2.5 *Benefits of empirical loss in channel estimation.*

We also validate the benefits introduced by the empirical loss terms in the adaptive filtering channel estimation algorithm in MoMA. The test case had increasing number of colliding packets, each with 100 data bits, with code length 14 and chip interval 125 ms. First, we evaluate the improvement in channel estimation by assuming the ground truth time of arrival (ToA) and comparing the BER of different channel estimation losses. Fig. 11 shows the BER averaged over 40 different traces for different number of colliding TXs. These results are for one molecule, so $\mathcal{L}_3$ is not applicable and will be evaluated later. We compare the performance of the full loss with the cases of missing either $\mathcal{L}_1$ or $\mathcal{L}_2$. In Fig. 11, it can be observed that $\mathcal{L}_2$ greatly improves the mean BER as the second bar is much higher than the other two. We also note that although L1 offers some improvement, (first and last bars), its contributions are not significant.

7.2.6 *Benefits of multiple molecules in channel estimation.*

In Sec. 4, we explain the advantages of multiple molecules. On one hand, more molecules provide more samples to conduct better packet detection. On the other hand, more molecules make it possible for the decoder to retrieve the common information about the channel, which improves channel estimation accuracy and thus reduces decoding error.

In Fig. 12a (similar to Fig. 11), we extend our results to two molecules. Fig. 12a shows the improvement in channel estimation with two molecules due to $\mathcal{L}_3$. The "salt-1" bar denotes the original one molecule data with $NaCl$, and the "salt-2" bar denotes the two-molecule emulation with both molecules as $NaCl$. Comparing these two bars reveals no improvement for $NaCl$ for decoding collisions. This is because estimating $NaCl$ CIR is good enough with one molecule under the current preamble length. Thus, we also conducted the same experiments with baking soda $NaHCO_3$. Although we used different concentration of TX solution to achieve roughly equal number of molecules per volume (20 grams per liter for $NaCl$ and 40 grams per liter for $NaHCO_3$), we can still see that $NaHCO_3$ has worse performance by comparing "salt-1" with "soda-1". So emulating two molecules with $NaHCO_3$ reveals higher improvement from "soda-1" to "soda-2". More intriguing results are revealed by the "salt-mix" and "soda-mix" bars, which represents the emulation of combining one $NaCl$ experiment with one $NaHCO_3$ experiment. The performance of each molecule are presented separately considering that these are anywhere different molecules. By comparing "soda-1" with "soda-mix", we can clearly see the improvement in channel estimation for the worse molecule even under the case of combing two different molecules, while the better one is not much influenced ("salt-1" with "salt-mix"). This validates the idea of using the commonness between CIRs to improve both of them based on their similarity. By comparing "soda-2" with "soda-mix", we can see that such improvement could be better replicating each one of the molecules and more improvements are foreseen with a smarter channel estimation algorithm.

Fig. 12b presents the similar experiments under the fork channel, which is more complicated than the line one. A preliminary conclusion can be achieved from Eq. 3 that slower background flow is equivalent to longer propagation distance. Assuming the flow splits equally into the forked tubes, which roughly doubles their length if under the same flow, the TX2 and Tx3 in the fork channel are at a equivalent distance of 60cm and 120cm in the line channel. However, the BER for these two TXs are much higher. Although the benefits of multiple molecules can still be observed under such case, the fork topology actually introduces more factors to the molecular channel, which will be considered in future.

A more extreme case is presented in Fig. 13. In this experiment, two colliding TXs used different codes on molecule A but the same code on molecule B. Further, the packets are intentionally made to collide in the preamble, which is the worst case scenario for channel estimation. Assuming ground truth ToA, we compare the mean BER when applying two different channel estimation losses, i.e. with and without the similarity loss $\mathcal{L}_3$. From Fig. 13, we can see that $\mathcal{L}_3$ merely affects decoding in molecule A since the two transmitters are distinguishable in coding, but the improvement is obvious when it comes to molecule B, where BER is reduced by more than half and becomes comparable to molecule A. This again validates the benefits of connecting signals from multiple molecules with $\mathcal{L}_3$.

7.2.7 *Benefits of multiple molecules in packet detection.*

In the end of this part, the general improvement in packet detection is evaluated. Fig. 14 shows percentage of detecting all 4

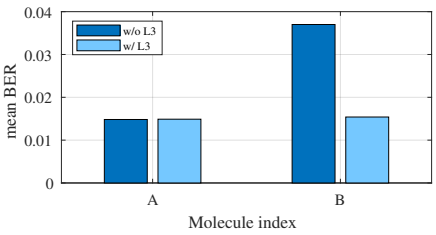


Figure 13: BER of decoding two colliding TXs under two-molecule emulations. Moreover, the two TXs share the same code on molecule B, but use different codes on molecule A, so $\mathcal{L}_3$ plays the role. ToA is assumed known.

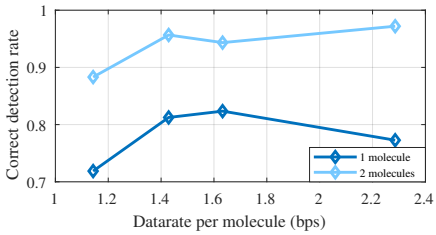


Figure 14: Benefits of multiple molecules in packet detection. Percentage of detecting all 4 colliding TXs correctly at different datarate.

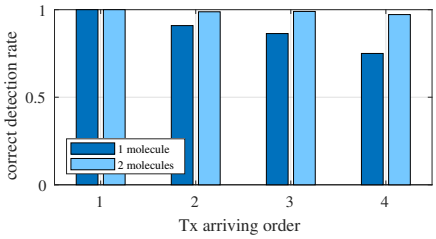


Figure 15: Benefits of multiple molecules in packet detection. Percentage of detecting each of the 4 colliding TXs correctly at a datarate of 2.29 bps per molecule.

colliding TXs correctly at different datarate. One can see that when using two information molecules, the correct detection rates are 10% higher than those with only one molecule and such improvement is consistent at all data rates. To better understand the improvements, we further examine whether each of the 4 packets is detected correctly, which is shown in Fig. 15. First, it can be seen that the later packets are more likely to be miss-detected because the detection happens concurrently with the decoding of all previous packets. The increased signal-dependent noise would lead to more decoding errors, making it harder to detect the newly arrived packets. Second, we can also see the huge improvement in using multiple molecules in packet detection, especially for the last arriving packet.

8 RELATED WORK

Enabling multiple access in molecular networks has been studied by past work. Except for a few [30, 64], most work has been theoretical or simulation-based:

- **TDMA:** [55] proposes a TDMA (Time Division Multiple Access) scheme. However, TDMA requires synchronization which is difficult to implement in practice due to (a) the long propagation delay of molecular signals and (b) the added complexity of requiring a receiver on the micro-implant.
- **MDMA:** [8, 18] introduce MDMA (Molecular Division Multiple Access) where different molecules are used by different transmitters to avoid interference altogether. This scheme is similar to FDMA (Frequency Division Multiple Access) in wireless networks which avoids synchronization and interference. However, it is difficult to scale this approach to many transmitters as most practical biomolecular systems are limited to 2-3 molecules [29, 53, 57].
- **ADMA:** [30] presents ADMA (Amplitude Division Multiple Access) that assigns the amplitude of the transmitted signal as the address of the source. This method reduces the error in resolving the source addresses under collisions, but the required signal power scales exponentially with the number of transmitters and the performance is sensitive to changes in the channel response.
- **CDMA:** [64, 68] propose a CDMA (Code Division Multiple Access) scheme that adopts Optical Orthogonal Codes (OOC) from fiber optic networks. [64] further tests the OOC codes in a gaseous molecular testbed with two transmitters. Despite the fact that, similar to molecular networks, the signal in fiber optic networks is also non-negative since it represents the intensity of light [67], OOC is not suitable for molecular networks for the following reasons: (a) Unlike molecular networks, fiber optic networks can deal with high ISI in hardware by compensating for the dispersion of light pulses. Techniques such as DCF (Dispersion Compensation Fiber) and FBG (Fiber Bragg Grating) [25] can change the physical properties of the channel to reduce ISI. However, in the presence of high ISI, one cannot simply correlate with the OOC code of each transmitter and decode them independently as proposed in [64, 68]; [4] (b) Commercial fiber optic systems that use OOC typically use long OOC codes which can significantly reduce the data rate [10]. Unlike optical networks that operate at link budgets of 100 Gbps [21], molecular networks operate in the few bits per second regime and cannot afford a significant reduction in data rate; (c) We show in our results in section 7 that using OOC codes of short length (high data rate) leads to high Bit Error Rates (BER). Moreover, [64] itself reports poor performance where the error rate is low only at data rates below 0.051 bits per second. MoMA, on the other hand, can achieve low error rate at data rates of 1 bit per second per transmitter. [28] also proposes a CDMA scheme that uses two molecules but is not based on OOC. However, the two molecules are used for modulation to encode different bits, i.e., releasing molecule A represents a "1" bit while releasing molecule B represents a "0". Since the two molecules cannot be released at the same time, the network capacity is cut by half. In contrast, MoMA uses the two molecules to send two data streams while at the same time improving packet detection and channel estimation.
- **Broadcast (No Multiple Access Scheme):** [14] computes the theoretical BER of a multi-transmitter network where the transmitters broadcast their data packets that collide at the receiver. However, in this work, each transmitter is decoded independently resulting in a drastic increase in BER as the number of transmitters increase. [4] also computes the theoretical capacity of a multiple-Tx single-Rx network when the network uses one molecule and transmitters broadcast their packets. However, the work does not present any approach for decoding the colliding packets.

Finally, designing practical transceivers that can process multiple molecules is a requirement for MoMA. While this is outside the scope of this paper, it is important to note that there exist studies on both natural and artificial molecular transmitters that can leverage multiple information particles [3]. Researchers have created cells

[4]Despite their name, OOC codes are not perfectly orthogonal but have good correlation properties when the code length is large.

that can respond to and integrate multiple molecules. For example, [62] constructs an Escherichia coli consortium-based biosensor that can detect and integrate three environmental signals (arsenic, mercury, and copper ion levels). [65] presents a robust, general, and scalable system (named BLADE) to engineer genetic circuits with multiple inputs and outputs in mammalian cells, enabling sophisticated cellular computation.

9 DISCUSSIONS

In this section, we shall explore the constraints of the study and the possible avenues for future research.

9.1 Alternative molecular testbed

In this work, we have used $NaCl$ to emulate the propagation of molecules in a liquid medium. In addition to $NaCl$, researchers have used a variety of particles to emulate molecular communication. This includes hydrogen ions with pH probe [16], superparamagnetic iron oxide nano-particles with susceptometer [59], glucose with electrolyte-gated FET [27] and color pigments with color sensor [42]. A detailed survey on existing molecular testbeds can be found at [36]. It must be noted that our current testbed can be seamlessly combined with other existing particles as well as biological systems to potentially implement a real-world molecular network system.

Although this paper highlights the challenges to address the non-negative signal, some of the above mentioned testbeds may have negative measurements. For example, in a pH-based testbed [16], acid (pH<7) can be treated as negative signal while base (pH>7) as positive. However, pH signal introduces its own challenge for CDMA. The variation in pH is in the log relation, instead of linear, which affects the cancellation of CDMA code because CDMA code assumes a linear relation between the measured signal and the transmitted code. But taking a step back, all these molecular signals reveal the variation in the concentration of information particles. Studying the relation between these two, such various signals can all be translated into the general concentration signal, so the MoMA decoder is generally applicable.

9.2 Multiple-molecule testbed

One major limitation of this work is that the testbed does not support concurrent transmission of multiple molecules. Due to the fact that our first measurement is electric conductivity indicating the concentration of $NaCl$, it is hard to introduce another molecule to the system such that it does not affect the EC value of the solution and at the same time can be measured with other methods. A direct improvement is to combine two existing testbeds (like the ones reported in [36]) with independent molecules and measurements. But targeting a more general solution, if we loosen the constraints on the independence, there will be a much wider choice. For example, if we use some acid as the second information molecule and pH value as the second measurement, the two molecules can be distinguished easily since $NaCl$ only changes the EC value while the acid changes both. Such testbed requires the knowledge in the influence of the acid concentration on the EC and pH value respectively, so it can recover the concentration changes of each molecule from the received signal. Moreover, these two measurements could even support more than two molecules. The preliminary idea is to understand how one molecule changes the two measurements at difference ratio. HCl dissolves in water and becomes H^+ and Cl^-, so EC and pH should change at a ratio of 1:1. Similarly, $NaCl$ is at a ratio of 1:0 and $NaOH$ of 1:-1. With such relation, the decoder is able to separate the signals of each molecule using a modified Viterbi algorithm that jointly decode signals from all molecules.

9.3 Current and future MC research

Majority of research works on molecular communication focuses on studying and modeling the molecular channel, transceiver designs of various particles, and sensing mechanisms using a theoretical framework. This was followed by a shift towards experiments to evaluate molecular communication spanning a range of few millimeters all the way to few meters. Our work falls in the longer range. A survey of the state of the art can be found in [35, 36]. The community seems to diverge into two directions. For the short range, experiments are typically involved with micro-scale transmitters that requires a strong basis in chemistry or biology. The long range experimental testbeds can emulate biological systems. Although such a testbed has similarities to wireless networks, due to slow propagation and particle diffusion, molecular signals endure high ISI that prohibits comparable datarate as existing wireless techniques in the same range.

In order to leverage the advancements in wireless and wired communication techniques in molecular systems, interdisciplinary collaboration is necessary. In this work, we have taken a step towards this collaboration by asking the question "what CAN a biological transmitter do" instead of “what SHOULD a transmitter do”. Unlike existing works, our design incorporates the hardware constraints of a bio-transmitter, a bottleneck posed by bio-engineering. We then offload processing to the receiver, which is more computationally resourceful than the transmitters. With this offloading, MoMA eliminates the needs for synchronization between the transmitter and the receiver. This is our attempt to bridge the gap between communication and biochemistry. Yet the community is not limited to these two fields. Based on our limited medical knowledge, we envision that such system can be used to monitor the spread of inflammation or the metastasis of cancer cells. Nonetheless, the value as well as the practical implementation of such an in-body system should be evaluated from the medical perspective, which invites more interdisciplinary researchers to participate and work together towards a innovative community.

10 CONCLUSION

In this paper, we introduce an innovative multiple access protocol, MoMA, for the molecular network with multiple transmitters and a single receiver. We highlight the unique properties of the molecular channel including non-negative signal, lack of synchronization and high ISI and explain the encoding and decoding schemes of data packets as well as the packet detection and channel estimation processes under our protocol. Our results demonstrate that using multiple transmitters along with multiple molecules with MoMA brings us closer to enabling more practical and scalable molecular networks.

REFERENCES

[1] Ozgur B Akan, Hamideh Ramezani, Tooba Khan, Naveed A Abbasi, and Murat Kuscu. 2016. Fundamentals of molecular information and communication science. *Proc. IEEE* 105, 2 (2016), 306–318.

[2] Ian F Akyildiz, Max Pierobon, Sasi Balasubramaniam, and Y Koucheryavy. 2015. The internet of bio-nano things. *IEEE Communications Magazine* 53, 3 (2015), 32–40.

[3] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter, et al. 2003. Molecular biology of the cell. *Scandinavian Journal of Rheumatology* 32, 2 (2003), 125–125.

[4] Baris Atakan and Ozgur B Akan. 2008. On molecular multiple-access, broadcast, and relay channels in nanonetworks. In *Proceedings of the 3rd International Conference on Bio-Inspired Models of Network, Information and Computing Sytems*. 1–8.

[5] Youssef Chahibi, Ian F Akyildiz, Sasitharan Balasubramaniam, and Yevgeni Koucheryavy. 2015. Molecular communication modeling of antibody-mediated drug delivery systems. *IEEE Transactions on Biomedical Engineering* 62, 7 (2015), 1683–1695.

[6] Youssef Chahibi, Massimiliano Pierobon, and Ian F Akyildiz. 2015. Pharmacokinetic modeling and biodistribution estimation through the molecular communication paradigm. *IEEE Transactions on Biomedical Engineering* 62, 10 (2015), 2410–2420.

[7] Youssef Chahibi, Massimiliano Pierobon, Sang Ok Song, and Ian F Akyildiz. 2013. A molecular communication system model for particulate drug delivery systems. *IEEE Transactions on biomedical engineering* 60, 12 (2013), 3468–3483.

[8] Xuan Chen, Miaowen Wen, Chan-Byoung Chae, Lie-Liang Yang, Fei Ji, and Kostromitin Konstantin Igorevich. 2021. Resource Allocation for Multiuser Molecular Communication Systems Oriented to the Internet of Medical Things. *IEEE Internet of Things Journal* 8, 21 (2021), 15939–15952.

[9] Wensong Chu and Charles J Colbourn. 2004. Optimal (n, 4, 2)-OOC of small orders. *Discrete Mathematics* 279, 1-3 (2004), 163–172.

[10] Fan RK Chung, Jawad A Salehi, and Victor K Wei. 1989. Optical orthogonal codes: design, analysis and applications. *IEEE Transactions on Information theory* 35, 3 (1989), 595–604.

[11] James E Dahlman, Kevin J Kauffman, Yiping Xing, Taylor E Shaw, Faryal F Mir, Chloe C Dlott, Robert Langer, Daniel G Anderson, and Eric T Wang. 2017. Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics. *Proceedings of the National Academy of Sciences* 114, 8 (2017), 2060–2065.

[12] Ramiz Daniel, Jacob R Rubens, Rahul Sarpeshkar, and Timothy K Lu. 2013. Synthetic analog computation in living cells. *Nature* 497, 7451 (2013), 619–623.

[13] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. 2010. A synchronized quorum of genetic clocks. *Nature* 463, 7279 (2010), 326–330.

[14] Yansha Deng, Adam Noel, Weisi Guo, Arumugam Nallanathan, and Maged Elkashlan. 2017. Analyzing large-scale multiuser molecular communication via 3-D stochastic geometry. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 3, 2 (2017), 118–133.

[15] Md Fakruddin, Zakir Hossain, and Hafsa Afroz. 2012. Prospects and applications of nanobiotechnology: a medical perspective. *Journal of nanobiotechnology* 10, 1 (2012), 1–8.

[16] Nariman Farsad, David Pan, and Andrea Goldsmith. 2017. A novel experimental platform for in-vessel multi-chemical molecular communications. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 1–6.

[17] Nariman Farsad, H Birkan Yilmaz, Andrew Eckford, Chan-Byoung Chae, and Weisi Guo. 2016. A comprehensive survey of recent advancements in molecular communication. *IEEE Communications Surveys & Tutorials* 18, 3 (2016), 1887–1919.

[18] Lluís Parcerisa Giné and Ian F Akyildiz. 2009. Molecular communication options for long range nanonetworks. *Computer Networks* 53, 16 (2009), 2753–2766.

[19] Robert Gold. 1967. Optimal binary sequences for spread spectrum multiplexing (Corresp.). *IEEE Transactions on information theory* 13, 4 (1967), 619–621.

[20] Andrea Goldsmith. 2005. *Wireless communications*. Cambridge university press.

[21] Lars Grüner-Nielsen, Stig Nissen Knudsen, Bent Edvold, Torben Veng, Dorte Magnussen, C Christian Larsen, and Hans Damsgaard. 2000. Dispersion compensating fibers. *Optical fiber technology* 6, 2 (2000), 164–180.

[22] Jack Kenneth Holmes. 2007. *Spread spectrum systems for GNSS and wireless communications*. Artech House Norwood.

[23] Vahid Jamali, Arman Ahmadzadeh, Wayan Wicke, Adam Noel, and Robert Schober. 2019. Channel modeling for diffusive molecular communication—A tutorial review. *Proc. IEEE* 107, 7 (2019), 1256–1301.

[24] Oliver Jonas, Heather M Landry, Jason E Fuller, John T Santini Jr, Jose Baselga, Robert I Tepper, Michael J Cima, and Robert Langer. 2015. An implantable microdevice to perform high-throughput in vivo drug sensitivity testing in tumors. *Science translational medicine* 7, 284 (2015), 284ra57–284ra57.

[25] NK Kahlon and G Kaur. 2014. Various dispersion compensation techniques for optical system: A survey. *Open journal of communications and software* 1, 1 (2014), 64–73.

[26] A Oguz Kislal, Bayram Cevdet Akdeniz, Changmin Lee, Ali E Pusane, Tuna Tugcu, and Chan-Byoung Chae. 2020. ISI-mitigating channel codes for molecular communication via diffusion. *IEEE Access* 8 (2020), 24588–24599.

[27] Bon-Hong Koo, Ho Joong Kim, Jang-Yeon Kwon, and Chan-Byoung Chae. 2020. Deep learning-based human implantable nano molecular communications. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 1–7.

[28] Sebastian Korte, Martin Damrath, and Peter Adam Hoeher. 2017. Multiple channel access techniques for diffusion-based molecular communications. In *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*. VDE, 1–6.

[29] Bhuvana Krishnaswamy. 2018. *Algorithms for molecular communication networks*. Ph. D. Dissertation. Georgia Institute of Technology.

[30] Bhuvana Krishnaswamy, Yubing Jian, Caitlin M Austin, Jorge E Perdomo, Sagar C Patel, Brian K Hammer, Craig R Forest, and Raghupathy Sivakumar. 2017. ADMA: Amplitude-division multiple access for bacterial communication networks. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 3, 3 (2017), 134–149.

[31] Mehmet S Kuran, Huseyin Birkan Yilmaz, Tuna Tugcu, and Ian F Akyildiz. 2011. Modulation techniques for communication via diffusion in nanonetworks. In *2011 IEEE international conference on communications (ICC)*. IEEE, 1–5.

[32] Murat Kuscu, Ergin Dinc, Bilgesu A Bilgin, Hamideh Ramezani, and Ozgur B Akan. 2019. Transmitter and receiver architectures for molecular communications: A survey on physical design with modulation, coding, and detection techniques. *Proc. IEEE* 107, 7 (2019), 1302–1341.

[33] Lin Lin, Qian Wu, Fuqiang Liu, and Hao Yan. 2018. Mutual information and maximum achievable rate for mobile molecular communication systems. *IEEE transactions on nanobioscience* 17, 4 (2018), 507–517.

[34] Ignacio Llatser, Eduard Alarcón, and Massimiliano Pierobony. 2011. Diffusion-based channel characterization in molecular nanonetworks. In *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 467–472.

[35] Sebastian Lotter, Lukas Brand, Vahid Jamali, Maximilian Schäfer, Helene M Loos, Harald Unterweger, Sandra Greiner, Jens Kirchner, Christoph Alexiou, Dietmar Drummer, et al. 2023. Experimental Research in Synthetic Molecular Communications–Part I: Overview and Short-Range Systems. *arXiv preprint arXiv:2301.06417* (2023).

[36] Sebastian Lotter, Lukas Brand, Vahid Jamali, Maximilian Schäfer, Helene M Loos, Harald Unterweger, Sandra Greiner, Jens Kirchner, Christoph Alexiou, Dietmar Drummer, et al. 2023. Experimental research in synthetic molecular communications-part ii. *IEEE Nanotechnology Magazine* (2023).

[37] Mohab A Mangoud, Marios Lestas, and Taqwa Saeed. 2018. Molecular motors MIMO communications for nanonetworks applications. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 1–5.

[38] James S McLean. 1996. A re-examination of the fundamental limits on the radiation Q of electrically small antennas. *IEEE Transactions on antennas and propagation* 44, 5 (1996), 672.

[39] Mark Mimee, Phillip Nadeau, Alison Hayward, Sean Carim, Sarah Flanagan, Logan Jerger, Joy Collins, Shane McDonnell, Richard Swartwout, Robert J Citorik, et al. 2018. An ingestible bacterial-electronic system to monitor gastrointestinal health. *Science* 360, 6391 (2018), 915–918.

[40] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C Stanton, and Christopher A Voigt. 2012. Genetic programs constructed from layered logic gates in single cells. *Nature* 491, 7423 (2012), 249–253.

[41] Phillip Nadeau, Mark Mimee, Sean Carim, Timothy K Lu, and Anantha P Chandrakasan. 2017. 21.1 Nanowatt circuit interface to whole-cell bacterial sensors. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*. IEEE, 352–353.

[42] Wenxin Pan, Xiaokang Chen, Xiaodong Yang, Nan Zhao, Lingguo Meng, and Fiaz Hussain Shah. 2022. A molecular communication platform based on body area nanonetwork. *Nanomaterials* 12, 4 (2022), 722.

[43] Yashar Rajavi, Mazhareddin Taghivand, Kamal Aggarwal, Andrew Ma, and Ada SY Poon. 2017. An RF-powered FDD radio for neural microimplants. *IEEE Journal of Solid-State Circuits* 52, 5 (2017), 1221–1229.

[44] Giordano Rampioni, Livia Leoni, and Pasquale Stano. 2018. Molecular communications in the context of "synthetic cells" research. *IEEE Transactions on NanoBioscience* 18, 1 (2018), 43–50.

[45] Shirin Salehi, Naghmeh S Moayedian, Shaghayegh Haghjooy Javanmard, and Eduard Alarcón. 2018. Lifetime improvement of a multiple transmitter local drug delivery system based on diffusive molecular communication. *IEEE transactions on nanobioscience* 17, 3 (2018), 352–360.

[46] Ehsan Samiei, Maryam Tabrizian, and Mina Hoorfar. 2016. A review of digital microfluidics as portable platforms for lab-on a-chip applications. *Lab on a Chip* 16, 13 (2016), 2376–2396.

[47] Daria M Shcherbakova, Anton A Shemetov, Andrii A Kaberniuk, and Vladislav V Verkhusha. 2015. Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annual review of biochemistry* 84 (2015), 519–550.

[48] Huanhuan Shi, Kaixuan Nie, Bo Dong, Mengqiu Long, Hui Xu, and Zhengchun Liu. 2019. Recent progress of microfluidic reactors for biomedical applications.

Chemical Engineering Journal 361 (2019), 635–650.
[49] Daniel F Sievenpiper, David C Dawson, Minu M Jacob, Tumay Kanar, Sanghoon Kim, Jiang Long, and Ryan G Quarfoth. 2011. Experimental validation of performance limits and design guidelines for small antennas. *IEEE Transactions on Antennas and Propagation* 60, 1 (2011), 8–19.
[50] Piro Siuti, John Yazbek, and Timothy K Lu. 2013. Synthetic circuits integrating logic and memory in living cells. *Nature biotechnology* 31, 5 (2013), 448–452.
[51] Shimyn Slomovic, Keith Pardee, and James J Collins. 2015. Synthetic biology devices for in vitro and in vivo diagnostics. *Proceedings of the National Academy of Sciences* 112, 47 (2015), 14429–14435.
[52] Christoph Steiger, Alex Abramson, Phillip Nadeau, Anantha P Chandrakasan, Robert Langer, and Giovanni Traverso. 2019. Ingestible electronics for diagnostics and therapy. *Nature Reviews Materials* 4, 2 (2019), 83–98.
[53] Kristina Stephens, Maria Pozo, Chen-Yu Tsao, Pricila Hauk, and William E Bentley. 2019. Bacterial co-culture with cell signaling translator and growth controller modules for autonomously regulated culture composition. *Nature communications* 10, 1 (2019), 4129.
[54] Andrew Stok and Edward H Sargent. 2002. The role of optical CDMA in access networks. *IEEE Communications Magazine* 40, 9 (2002), 83–87.
[55] Junichi Suzuki, Sasitharan Balasubramaniam, and Adriele Prina-Mello. 2012. Multiobjective TDMA optimization for neuron-based molecular communication. In *7th International Conference on Body Area Networks*.
[56] Alvin Tamsir, Jeffrey J Tabor, and Christopher A Voigt. 2011. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469, 7329 (2011), 212–215.
[57] Jessica L Terrell, Tanya Tschirhart, Justin P Jahnke, Kristina Stephens, Yi Liu, Hong Dong, Margaret M Hurley, Maria Pozo, Ryan McKay, Chen Yu Tsao, et al. 2021. Bioelectronic control of a microbial community using surface-assembled electrogenetic cells to route signals. *Nature Nanotechnology* 16, 6 (2021), 688–697.
[58] Bige D Unluturk, A Ozan Bicen, and Ian F Akyildiz. 2015. Genetically engineered bacteria-based biotransceivers for molecular communication. *IEEE Transactions on Communications* 63, 4 (2015), 1271–1281.
[59] Harald Unterweger, Jens Kirchner, Wayan Wicke, Arman Ahmadzadeh, Doaa Ahmed, Vahid Jamali, Christoph Alexiou, Georg Fischer, and Robert Schober. 2018. Experimental molecular communication testbed based on magnetic nanoparticles in duct flow. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 1–5.
[60] Alexander Van Reenen, Arthur M de Jong, Jaap MJ den Toonder, and Menno WJ Prins. 2014. Integrated lab-on-chip biosensing systems based on magnetic particle actuation–a comprehensive review. *Lab on a Chip* 14, 12 (2014), 1966–1986.
[61] Andrew Viterbi. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory* 13, 2 (1967), 260–269.
[62] Baojun Wang, Mauricio Barahona, and Martin Buck. 2013. A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosensors and Bioelectronics* 40, 1 (2013), 368–376.
[63] Jiaming Wang, Dongyin Hu, Chirag Shetty, and Haitham Hassanieh. 2020. Understanding and embracing the complexities of the molecular communication channel in liquids. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–15.
[64] Linchen Wang and Andrew W Eckford. 2017. Nonnegative code division multiple access techniques in molecular communication. In *2017 15th Canadian Workshop on Information Theory (CWIT)*. IEEE, 1–5.
[65] Benjamin H Weinberg, NT Pham, Leidy D Caraballo, Thomas Lozanoski, Adrien Engel, Swapnil Bhatia, and Wilson W Wong. 2017. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nature biotechnology* 35, 5 (2017), 453–462.
[66] Tomohiro Yamada, Takumi Uezono, Kenichi Okada, Kazuya Masu, Akio Oki, and Yasuhiro Horiike. 2005. RF attenuation characteristics for in vivo wireless healthcare chip. *Japanese journal of applied physics* 44, 7R (2005), 5275.
[67] Guu-Chang Yang and Wing C Kwong. 1997. Performance comparison of multiwavelength CDMA and WDMA+ CDMA for fiber-optic networks. *IEEE Transactions on Communications* 45, 11 (1997), 1426–1434.
[68] Yeganeh Zamiri-Jafarian, Saeed Gazor, and Hossein Zamiri-Jafarian. 2016. Molecular code division multiple access in nano communication systems. In *2016 IEEE Wireless Communications and Networking Conference*. IEEE, 1–6.
[69] Mohammad Zoofaghari and Hamidreza Arjmandi. 2018. Diffusive molecular communication in biological cylindrical environment. *IEEE transactions on nanobioscience* 18, 1 (2018), 74–83.

Appendices are supporting material that has not been peer-reviewed.

A DECODER PSEUDO CODE

Algorithm 1 is the pseudo code of the whole decoding process of MoMA.

Algorithm 1: MoMA Sliding Decoder

$S \leftarrow$ set of all transmitters;
$S_d \leftarrow \emptyset$;
for *each sliding window* **do**
  $y \leftarrow$ receiver signal in the current window;
  $S_c \leftarrow \emptyset$;
  while $|S_d| + |S_c| < |S|$ **do**
    if $|S_d| + |S_c| > 0$ **then**
      $b \leftarrow$ decoded data for transmitters in $S_d \cup S_c$ with y;
      $y_d \leftarrow \sum_{i \in S_d \cup S_c} b_i * h_i$;
      $y_r \leftarrow y - y_d$;
    else
      $y_r \leftarrow y$;
    end
    $t \leftarrow$ peak location of correlating the preamble of each transmitter $i \in S/(S_d \cup S_c)$ with y_r;
    $tx\_added \leftarrow false$;
    for $i \in S/(S_d \cup S_c)$ *in the increasing order of* t **do**
      $b \leftarrow$ decoded data for transmitters in $S_d \cup S_c \cup \{i\}$, which iterates between decoding and channel estimation until convergence;
      $h_1 \leftarrow$ CIR of all transmitters in $S_c \cup \{i\}$, estimated with b and the samples of y which overlap with the 1st half of transmitter i's preamble;
      $h_2 \leftarrow$ CIR of all transmitters in $S_c \cup \{i\}$, estimated with b and the samples of y which overlap with the 2nd half of transmitter i's preamble;
      if h_{1i} *is similar to* h_{2i} **then**
        $tx\_added \leftarrow true$;
        $S_c \leftarrow S_c \cup \{i\}$;
        break for;
      end
    end
    if tx_added *is false* **then**
      break while;
    end
  end
  $S_d \leftarrow S_d \cup S_c$;
  $h \leftarrow$ CIR estimation of all transmitters in S_d;
  Decode all transmitters in S_d using h;
  Remove all transmitters from S_d at end of packet;
  Advance the sliding window;
end

B FURTHER SCALING WITH MULTIPLE MOLECULES

In this section, we will discuss the possibility of further scaling up the number of transmitters in an MC network. The methods are named code tuple and delayed transmission, which are both supported by the usage of multiple molecules.

B.1 Code Tuple

First, we propose the concept of code tuple, which refers to the code assignments of one transmitter on all molecules. It, allowing multiple transmitters to share the same code on some but not all molecules, greatly scales up the number of transmitters in an MC network. Theoretically this scheme is also applicable in conventional networks when the receiver jointly decode all colliding packets with the knowledge of the CIRs. However, the benefits of decoding collisions do not beat the cost in hardware when the majority of collisions can be avoided by the synchronization techniques. The more critical issue is that channel estimation is almost impossible when the two packets with the same code collide in the preamble.

As illustrated in Fig. 13, using the relation between CIRs is able to address collisions of multiple transmitters using the same code, as long as there is different coding on other molecules. Suppose the size of the codebook on each molecule is G, the total number of different code tuples for a system with M molecules is G^M. However, how to find the suitable number of transmitters in network is a complicated problem relating to multiple factors, such as the packet collision frequency, the various propagation loss, as well as the increasing signal dependent noise with more colliding packets. Besides, how to select the suitable ones among all the code tuples is also an interesting topic.

B.2 Delayed transmission

Another possible improvement of MoMA is to introduce delayed transmission. This delay does not refer to packets from different transmitters, which should be a common case without synchronization in MoMA. **Instead, one transmitter does not have to send packets on all molecules in the same time slot, but with specifically designed offsets.** For example, packet on the second molecule starts one symbol later than the first, and packet on the third molecule starts two symbols later than the second. With different delays introduced for transmitters using the same code tuple, the network can accommodate more transmitters. Even when the two transmitters share the same codes on all molecules, they can distinguish in the packet transmission order on molecules, such as the earliest packet of one transmitter is on the first molecule while another transmitter is on the second molecule. Besides, delayed transmission is expected with another advantage, which reduces the influence of sudden change in the channel especially for the preamble. The quality of packet detection and channel estimation could be greatly compromised due to the bursting noise when the packet arrives. By separating the preambles, the influence of such burst error could be reduced with the receiver signal on other molecules.