

Low-latency Visible Light Backscatter Networking with RETROMUMIMO

Kenuo Xu[†], Chen Gong[†], Bo Liang[†], Yue Wu[†], Boya Di[†], Lingyang Song[†], Chenren Xu[†]★✉*

[†]Peking University ★Key Laboratory of High Confidence Software Technologies, Ministry of Education (PKU)

ABSTRACT

Visible Light Backscatter Communication (VLBC) presents an emerging ultra-low-power IoT connectivity solution with high spatial-spectral efficiency and intrinsic human-perceivable privacy advantages. However, research progress on enhanced data rate and sophisticated device coordination of state-of-the-art VLBC systems still cannot meet the low-latency requirement (sub-second level for an IoT network) for massive connections.

In this paper, we present the design, implementation and evaluation of RETROMUMIMO, the first visible light backscatter network that enables physical-layer concurrency to minimize network latency. We propose a pulse feature extraction scheme to enable concurrent transmissions with the unexplored pulse diversity of tags. We design a low-latency demodulation algorithm assisted by such features to demodulate the concurrent symbols efficiently and effectively. Based on the physical-layer concurrency, we further design a MAC layer with contention-based transmission and minimal protocol overhead for low-latency networking. Our evaluation shows that the prototype system achieves up to 8 concurrent VLBC uplinks with real-time demodulation and up to 92.0% latency reduction compared with the state-of-the-art VLBC systems.

CCS CONCEPTS

• **Hardware** → **Wireless devices**; • **Computer systems organization** → *Embedded systems*.

KEYWORDS

Visible Light Backscatter Networking; Backscatter Communication; Multi-User MIMO; Pulse-division Concurrent Transmission

ACM Reference Format:

Kenuo Xu, Chen Gong, Bo Liang, Yue Wu, Boya Di, Lingyang Song, Chenren Xu. 2022. Low-latency Visible Light Backscatter Networking with RETROMUMIMO. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22)*, November 6–9, 2022, Boston, MA, USA. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3560905.3568507>

1 INTRODUCTION

Backscatter communication and networking [1], as an emerging sub-mW connectivity solution for IoT networks, has attracted significant research interest over the past few years. Flourishing research

✉: chenren@pku.edu.cn

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](https://permissions.acm.org).

SenSys '22, November 6–9, 2022, Boston, MA, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9886-2/22/11...\$15.00

<https://doi.org/10.1145/3560905.3568507>

results in this area have covered nearly all communication mediums, including microwaves and shortwaves [2], millimeter waves [3], magnetic field [4], and acoustics [5]. Along with this technological trend, visible light backscatter communication (VLBC) [6], based on the idea of modulating retroreflection of visible light, offers unique advantages from the directionality nature of visible light, such as high spectral efficiency and human-perceivable privacy benefits. The combination of low-power connectivity and visible light potentially enables a wide range of applications such as navigation information services [7], indoor positioning [8], behavioral sensing [9], and human-computer interaction [10].

State-of-the-art practical VLBC systems yield considerable end-to-end latency (*i.e.*, from the time when a new message packet is to be sent until the packet is received and decoded), especially in most IoT scenarios with sparse traffic and massive connectivity. For example, it takes 2.2 seconds for one reader to scan 5 tags separately with a CSMA-like MAC [11], but the time increases to 70 seconds with 8 tags. Apart from the inefficient MAC protocol, the latency bottleneck is rooted in the low refreshing/switch rate (*i.e.*, 100 – 240 Hz) of the TFT liquid crystal modulator (LCM) for optical modulation, which leads to a low-bit-rate physical layer (less than 1 kbps). RetroTurbo [12] pushes the bit rate limit to 8 kbps with an advanced time-polarization orchestrated modulation on a multi-pixel LCM. However, the per packet channel estimation overhead of 130 ms (to utilize the multiple LCM pixels) dilutes the effort in rate boosting. In a nutshell, the latency of all the concurrent VLBC networks grows to hundreds of seconds or even higher when the network size scales to tens of nodes. Therefore, practical VLBC networking craves a systematic low-latency design that minimizes the networking latency originating in both PHY and MAC layers.

In this paper, we present RETROMUMIMO, a visible light backscatter system dedicated to low-latency networking. RETROMUMIMO is inspired by radio-frequency (RF) MU-MIMO that coordinates multiple users (tags in this paper) to transmit signals concurrently over the same channel via spatial multiplexing. It enables and utilizes the physical-layer concurrency to minimize the delay of a user waiting for other users to exit the channel. We identify three key design challenges in building the RETROMUMIMO system.

Diversity Challenge. We identify the primary challenge as *promoting the diversity in signals* of multiple channels from different tags, *i.e.*, the transmitting signals of multiple tags must be different enough to be distinguished from each other. Almost all the concurrent RF backscatter communication techniques explicitly or implicitly examine and enhance the channel diversity, such as the orthogonality (perfect diversity) in the frequency domain [13, 14], frequency-selective channel diversity [15], spatial-frequency diversity [16], time-frequency diversity [17–19], time-IQ diversity [20–24], diversity in the transmitted information [25, 26], and coding diversity [27, 28]. Unfortunately, all these techniques do not

apply to VLBC. LCMs can only modulate two states, ON and OFF (e.g., “1” and “0”) with their unique non-linear pulse responses, whereas the RF techniques typically require meticulous modulation of phase, frequency, or amplitude of sine waves. Therefore, it is a challenge to find and promote channel diversity for LCMs with insufficient modulation capability in order to enable practical concurrent transmission.

- *Solution:* Although the diversity techniques are not applicable to VLBC, we find that there is unexplored diversity in the temporal sequences sent by LCMs. We first take a deep inspection of the LCM pulse response (*pulse* in short) and discover the heterogeneity of the pulses from different tags. The heterogeneity (*pulse diversity*), which is specific to the *tag context* such as its placement and manufacturing imperfection, can be leveraged to distinguish the signal source from multiple tags during their concurrent transmissions. Pulse diversity can also be integrated with status-quo modulation schemes such as pulse amplitude modulation. We further design a data-driven method that profiles and extracts *p-features* (the unique features of a pulse) as a low-dimensional but precise fingerprint of each tag during transmission. The extraction maps the intuitively-similar temporal sequences of pulses into a *p-feature* domain where a pulse can be clearly distinguished from others. Once the tags are coordinated by the MAC protocol (to be discussed later), we can modulate and superimpose the pulses from multiple tags for *pulse-division concurrent transmission*.

Demodulation Challenge. The pulse-division concurrent transmission shares the same mathematical formalization as RF (MU-)MIMO. However, the channel matrix constructed from LCM pulses is rank-deficient because of the strong correlation between the pulses (although diversity exists). The rank deficiency poses a challenge for MIMO detectors to decode the transmitted message bits from the received symbols. The widely-used linear detectors (e.g., zero-forcing) [29] are inherently unable to cope with the non-orthogonal channel with a large *condition number* (typically more than 40) of the channel matrix [30]. In contrast, a maximum likelihood detector (MLD) traverses the entire search space to find the optimal solution at the cost of exponential computational complexity, leading to prolonged demodulation latency. Status-quo acceleration techniques for MLD algorithms (e.g., sphere decoders [31]) still fail to handle the rank-deficient channels [32]. Therefore, designing a low-latency demodulation algorithm for our rank-deficient pulse-division concurrent channels remains a challenge.

- *Solution:* The reason why existing detection optimization methods fail is that their *instable operations* (such as matrix inversion and decomposition) have large relative errors when applied to the rank-deficient matrices. Our demodulation algorithm avoids such operations and optimizes the basic MLD design in two aspects, the *search space* and the *objective functions*. Both optimizations are based on the insight that the *p-features* provide a precise low-dimensional fingerprint of pulse diversity. The algorithm employs different numbers of *p-features* in different stages to enhance efficiency and ensure searching accuracy.

MAC Challenge. The pulse-division concurrent transmission design takes an essential step towards low-latency networking, yet

RETROMUMIMO still needs a MAC layer to efficiently manage the full communication process, including channel estimation, data transmission, and control message transmission. For channel estimation, it is required because the channel matrix, as an input to the demodulator, may vary with time, especially when the tags are mobile and may dynamically enter/exit the coverage of the network. In practice, the reader probes the channel state information with a stand-alone channel estimation process where the tags send channel probe signals sequentially. A mixture of probe signals leads to estimation failures and should be avoided. The MAC protocol should thoroughly analyze, manage, and coordinate the *sequential channel estimation* and *concurrent transmission* processes. Control messages such as ACKs are necessary to ensure the reliability of RETROMUMIMO. RETROMUMIMO needs to minimize the control overhead because of the limited physical-layer data rate (and end-to-end latency). Therefore, the MAC protocol carefully analyzes the use of different control messages and minimizes the overall overhead while ensuring reliability.

- *Solution:* We design a centralized MAC sublayer that leverages the broadcasting nature (within the field of view, or FoV) of the reader’s downlink to synchronize and control all the tags. The tags contend for channel access to minimize queuing delay and network statuses for flexible tag (re)grouping under dynamic channel conditions. The contention process reserves several slots for sequential channel estimation. The classic time slotting and binary exponential backoff strategies are adopted to guarantee efficiency in collision detection and recovery. Transmission failures (from collisions or other reasons) are further handled with a simplified ACK mechanism design of both the reader and tags to minimize the protocol state space but still ensure reliability.

We build a prototype system of RETROMUMIMO and our experimental results show RETROMUMIMO supports up to 8 concurrent links in various indoor settings and achieves real-time demodulation. The links are robust to different types of interfering factors, including ambient position arrangements, illumination, and tag heterogeneity. The demodulation algorithm achieves superior accuracy and reduces demodulation time by 60.1% compared with the baseline (sphere decoder). We further verify our MAC protocol that achieves up to 92.0% latency reduction over the state-of-the-art solutions with tens of tags.

Contributions.

- We investigate the pulse diversity of tags and design a pulse-division concurrent communication channel that enables multiple tags to transmit simultaneously (§4).
- We propose a low-latency demodulation algorithm that addresses the challenge from rank-deficient channels in the pulse-division concurrent transmission design (§5).
- We design a MAC protocol that coordinates the concurrency-based contentions and efficiently turns PHY-layer concurrency into low-latency multiple access (§6).
- We build the RETROMUMIMO system (§7) that supports up to 8 concurrent VLBC links for the first time and our evaluation shows it can achieve more than 90% in latency reduction over the status-quo solutions (§8).

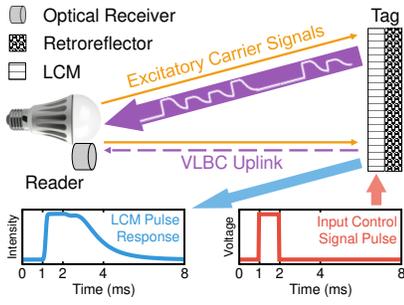


Figure 1: Illustration of VLBC uplink and LCM pulse.

2 BACKGROUND

2.1 VLBC Primer

Visible light backscatter communication (VLBC) is an emerging visible light communication technology that enables low-power tags to backscatter the information to the optical receivers. The unique and core design of VLBC, *i.e.*, the VLBC uplink (tag-to-reader link), is shown in Fig. 1. A reader, similar to the reader of the RFID systems, provides excitatory carrier signals to the tags for the uplink transmission. A tag exploits retroreflective fabric to bounce the incident light back to the source direction. In addition, the tag also leverages a liquid-crystal display shutter as a liquid-crystal modulator (LCM) to modulate information on the light signal. If a voltage pulse (as a control signal) is applied to the LCM, it has a unique and non-linear pulse response, and we call the LCM pulse response *pulse* for simplicity in the rest of this paper. The long-tail discharging phase leads to inter-symbol interference if the next symbol starts before the previous discharging finishes. Therefore, a number of modulation schemes such as pulse-position modulation and pulse-width modulation are inefficient and require complex demodulation design. In contrast, the simple-yet-effective pulse-amplitude modulation [33] turns on and off a certain number of LCM pixels to modulate bits onto the amplitude of the retroreflected signal with a multi-pixel LCM.

2.2 MLD for MIMO Detection

A MIMO communication system employs multiple transmitting and receiving antennas. Suppose that the symbols from the transmitting antennas form a vector \mathbf{x} . Then the received signals \mathbf{y} can be expressed as $\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta}$ where \mathbf{H} is the channel matrix and $\boldsymbol{\eta}$ is the additive noise. A MIMO receiver would leverage detectors to recover the transmitted data \mathbf{x} with the received signals \mathbf{y} and the channel matrix \mathbf{H} (estimated by channel probe signals in practice).

The *Maximum-Likelihood Detector* (MLD) is the optimum detector that minimizes the probability of error. MLD is defined as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{X}} D(\mathbf{x}). \quad (1)$$

It selects the symbol vector $\hat{\mathbf{x}}$ that minimizes an *objective function* $D(\mathbf{x})$ that measures the distance between \mathbf{y} and \mathbf{x} . In principle, $D(\mathbf{x})$ calculates the Euclidean distance by $D(\mathbf{x}) = \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2$. \mathbb{X} is the *search space*, containing all the possible solutions of \mathbf{x} by default.

3 RETROMUMIMO OVERVIEW

RETROMUMIMO is a visible light backscatter network design that utilizes concurrent uplink (tag-to-reader link) transmissions to

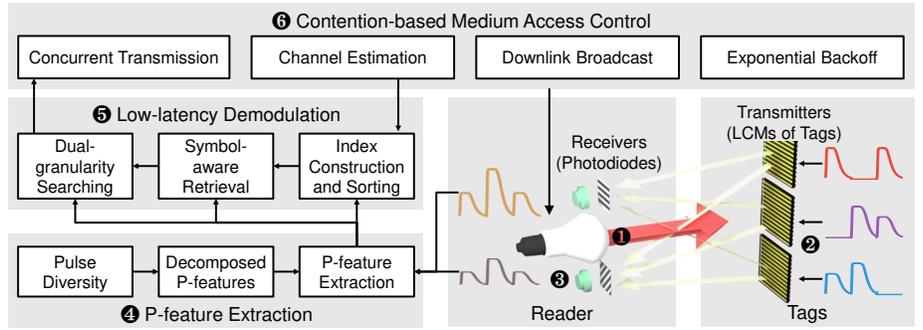


Figure 2: The architecture and workflow of RETROMUMIMO.

achieve low-latency networking. RETROMUMIMO works as shown in Fig. 2. ❶ The reader initiates transmission with a downlink broadcast (with its LED same as what the conventional VLC does) and keeps transmitting carrier signals during uplink. ❷ Each tag modulates its transmitting sequences with pulses. Pulse amplitude modulation is used to enhance the data rate. The following processes show the unique features of low-latency networking:

Pulse-division Concurrent Transmission. The shapes of pulses transmitted by the LCMs are diverse from each other, thus having the potential for concurrent transmissions (§4.1). Therefore, the simultaneously-transmitted pulses can be distinguished from each other. ❸ A reader can receive the superimposition of the pulses for concurrent transmission. ❹ The concurrent symbols are further processed with a p-feature extraction technique. The p-features function as an effective low-dimensional fingerprint of a pulse. Different pulses can be distinguished with a few p-features (§4.2). The pulse diversity further enables a pulse-division concurrent transmission design (§4.3).

Low-latency Demodulation. ❺ The pulse-division concurrent transmission requires an efficient demodulation design to extract messages from multiple simultaneous transmissions without additional computational latency. The demodulation algorithm is based on the MLD algorithm to ensure the bit-error-rate (BER) performance. In order to reduce the search space, an index for each possible solution is constructed with the principal p-feature of each tag during channel estimation (§5.1). Each concurrent symbol also corresponds to an index during transmission. The possible solutions whose indices are close to the index of the received symbol are more likely to be the optimal solution, and we retrieve a small set of such solutions for further searching (§5.2). A dual-granularity search over the retrieved solutions finally determines the optimal demodulation result efficiently and accurately (§5.3).

Contention-based MAC. ❻ A medium access control (MAC) protocol handles the PHY-layer concurrency to achieve low-latency networking. The MAC protocol features several design highlights including a contention-based design to support multiple access even when the link number is beyond the PHY concurrency limit, a centralized and slotted protocol for channel estimation to minimize collisions, and minimization of control messages such as ACKs to reduce the overhead (§6.1). Specifically, the MAC layer divides the time resource into query-response cycles to coordinate the reader and tags. The tags are synchronized by the reader and perform

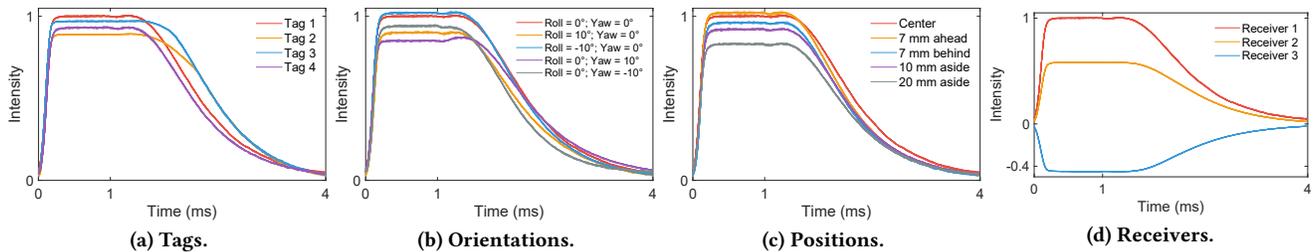


Figure 3: Illustrations of the LCM pulse shapes with different setups. The LCMs are charged in the first millisecond and discharged afterward. Each category in one figure is composed of 10 recorded pulses from two environments.

contention-based medium access to reduce networking latency. The reader notifies all the tags in the next downlink broadcast once collisions of the tags occur (§6.2). We also design a message structure for the query-response cycles (§6.3).

4 PULSE-DIVISION CONCURRENT TRANSMISSION

Concurrent transmissions require the diversity of the simultaneous signals, *i.e.*, the signals from different devices can be distinguished from each other. RF communication typically benefits from different frequencies or phases, which is not applicable to VLBC because LCMs can only modulate two states ON and OFF (*e.g.*, “1” and “0”). In this section, we explore that there is retro-transmitting diversity that can be leveraged for concurrent transmission. The diversity lies in the nature of LCMs and manifests with no need for additional active control. Therefore, it is not restricted by the insufficient modulation capability of LCMs. We further design a data-driven method to convert the intuitively-fragile pulse diversity into a vectorized and reliable fingerprint in the p-feature domain. Therefore, multiple tags can transmit pulses simultaneously. We formalize the concurrent transmission as a MIMO-like problem.

4.1 Pulse Diversity of LCMs

We mount the VLBC reader and tags to an optical platform, which guarantees a sub-millimeter level reproducibility of placement as shown in Fig. 14, to explore the diversity of LCM pulses. We collect LCM pulses (the process of LCM switches from “0” to “1” and then returns) multiple times with different setups. Fig. 3 shows the intensity (amplitude) of pulses with time. The different shapes (*i.e.*, temporal sequences) of pulses contain *pulse diversity*, which not only lies in the absolute intensity (at 1 ms) but also in the discharging phase (from 1 ms to 4 ms). The pulse diversity can be observed in different tags or the same tag with different orientations or positions. We refer to such diversity sources from tags as *tag context*. Besides, one pulse received by different receivers also shows diversity. The details are illustrated as follows.

Tag Diversity. We mount 4 tags in turn and record the pulse shapes respectively. As shown in Fig. 3a, the discharging phases of tags show different changing rates, and each phase has a unique gradient pattern of the phase. Besides, the moment the received signal starts to change its strength also differs. Therefore, different tags exhibit pulse diversity from the manufacturing and assembly tolerance.

Orientation Diversity. We vary the orientations of one tag and plot the pulses in Fig. 3b. The results show that the shape of the pulse still varies with slight orientation changes (10°). Orientation diversity comes from the optical property of liquid crystals. An LCM rotates the polarization direction of the incident light by

90° or not according to its internal molecular distribution. The conversion of two stable states (especially from “1” to “0”) is complex because the refractive index for different wavelengths changes differently [34]. Therefore, the total intensity of visible light of different wavelengths differs in different orientations, especially in the disordered discharging phase where the liquid crystal molecules are driven by intermolecular force. As a result, the orientation-relevant pulse shapes also bring orientation diversity.

Position Diversity. Different tag positions also bring diversity, as shown in Fig. 3c. Even a few millimeters of position change can result in diversity in pulse shapes. We attribute the position diversity to the change in the relative positions between the receiver and the tag because the light signals travel in different paths and orientations. Besides, the uneven lighting (from the illumination patterns of LEDs and the optical design of the light source) increases the degree of diversity.

Receiver Diversity. The aforementioned properties are all about transmitter diversity, whereas multiple receivers of the reader, which are naturally placed in different positions and orientations, also promote pulse diversity. We use 3 pairs of receivers with 120° between each other to verify receiver diversity, and Fig. 3d shows the diversity in the received pulse shapes¹ from different receivers. This diversity indicates that it is possible to leverage multiple receivers (from different polarizations and/or positions) to further promote the diversity of the retro-transmitting channel.

Summary of Pulse Diversity. Our experiments show that pulse diversity is ubiquitous in practical VLBC setups where the tags naturally have different tag contexts because they are different tags and placed in different locations. The diversity comes from the heterogeneity of different LCMs, orientations, positions, and receivers. The transmitted pulse shapes in one tag context are stable over time and in different nearby environments (*e.g.*, different rooms separated by opaque walls). The pulse shapes in multiple tag contexts differ from each other. This observation indicates that we can use pulse diversity as the foundation of concurrent transmission diversity, without the need for actively-controlled diversity like frequency shifting [19]. It is likely (but with low probability) that the pulse shapes from two certain tag contexts are exactly the same by coincidence, and thus cannot be distinguished. This case, of which the probability and influence are to be evaluated as shown in Fig. 15f, leads to a concurrency failure but can be further handled with the MAC protocol (§6).

¹The negative intensity comes from the reception design that differentiates signals from different polarities (§7).

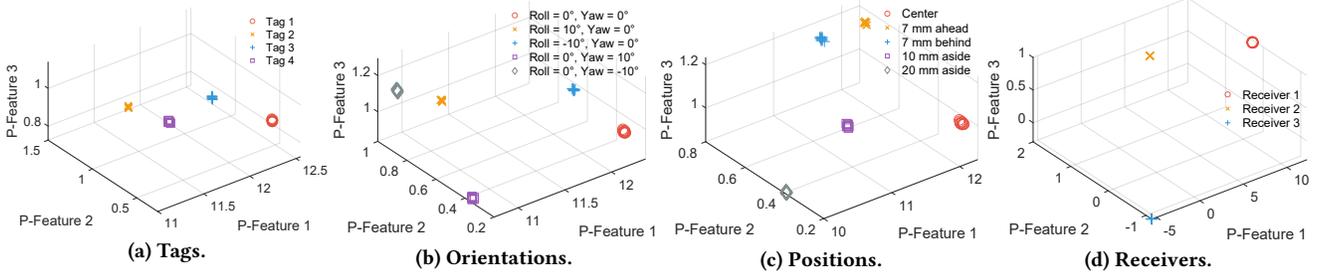


Figure 4: Illustrations of the first 3 p-features corresponding to the pulses of Fig. 3.

4.2 Feature Decomposition and Extraction

As shown in Fig. 3, the difference between the pulses is subsistent but slight compared with the pulse itself. In order to utilize and leverage the pulse diversity against stronger random noise, we need a method to represent the pulses efficiently and filter out the random noise. We propose a method based on singular value decomposition (SVD), a mathematical tool applied widely for matrix approximation and compression. Different from the conventional wireless communication systems where SVD is applied for MIMO precoding [35], this method uses SVD for feature extraction of pulses. Our method explores the fundamental physical properties of liquid crystals and prepares such properties for concurrent transmission. We name the singular vectors from SVD as *p-features* (short for pulse features) for the rest of this paper.

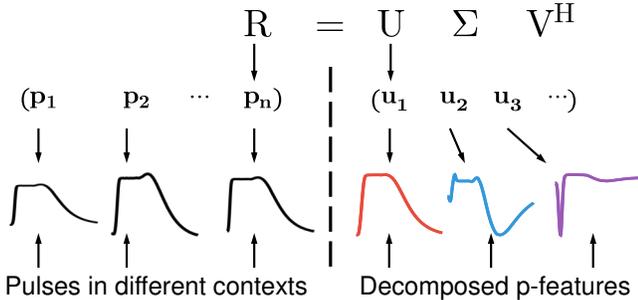


Figure 5: An illustration of p-feature decomposition.

Decomposing P-features. The process of SVD-based p-feature decomposition is shown in Fig. 5. We first construct a training set $\mathbf{R} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n)$ with offline-recorded pulses from the tags in different contexts, including various tags, orientations, positions, and receivers. In other words, the pulse diversity is abundantly collected in \mathbf{R} . We further perform SVD of \mathbf{R} as $\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Each column vector \mathbf{u}_i of \mathbf{U} are perpendicular to each other and form a set of orthogonal bases of any column of \mathbf{R} . Therefore, \mathbf{u}_i represents the features of the pulses (*i.e.*, *p-features*) in the raw pulse matrix. In addition, a few leftmost p-features (corresponding to the largest singular values) can be leveraged to approximate a pulse, just like how SVD is used for low-rank matrix approximation.

Extracting P-features. A pulse (not required to be included in \mathbf{R}) can be represented with the p-features by calculating its coefficient vector \mathbf{c} as shown in Fig. 6. Due to the orthogonality of \mathbf{U} (*i.e.*, $\mathbf{U}^H\mathbf{U} = \mathbf{I}$), \mathbf{c} can be calculated by $\mathbf{c} = \mathbf{U}^H\mathbf{p}$. To verify the effectiveness

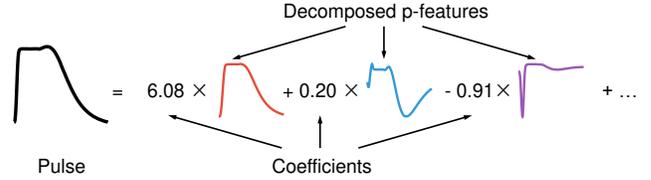


Figure 6: An example of p-feature extraction.

of our p-feature extraction method, we extract coefficient vectors of the pulses shown in Fig. 3 and scatter the coefficients of the top-3 p-features accordingly. The results in Fig. 4 indicate that (1) the contributions to signal recovery of the p-features decline quickly as the index grows, as the coefficient of the first p-feature (which we refer to as *principal p-feature*) is 10x larger than the other two; (2) the p-features from the same category cluster and different clusters can be clearly distinguished. The properties show that although the pulse diversity in Fig. 3 may seem fragile and easily disrupted by noise, the space constructed from p-features provides an explanatory and robust fingerprint for the pulses to be distinguished from each other. Therefore, the p-feature is an effective and efficient approach to represent a pulse and will be further leveraged for demodulation in §5 and evaluated in §8.3.

4.3 Concurrency with Pulse Diversity

When multiple tags transmit pulses simultaneously, the superimposed pulse received from a receiver is additive, *i.e.*, $\mathbf{p}_r = \sum_{j=1}^n \mathbf{p}_j$. The coefficients of p-features satisfy $\mathbf{c}_r = \mathbf{U}^H\mathbf{p}_r = \mathbf{U}^H \sum_{j=1}^n \mathbf{p}_j = \sum_{j=1}^n \mathbf{c}_j$, where \mathbf{p}_j is the pulse from the j^{th} tag out of the n tags in total, *i.e.*, *the coefficients of p-features are additive* for superimposed pulses. Therefore, we can use the p-features to enable concurrent transmission with pulse diversity.

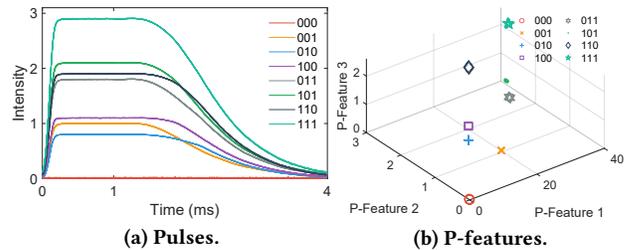


Figure 7: Illustrations of concurrent symbols of 3 tags. “101” means tag 1, 2, 3 transmit “1”, “0”, “1” respectively.

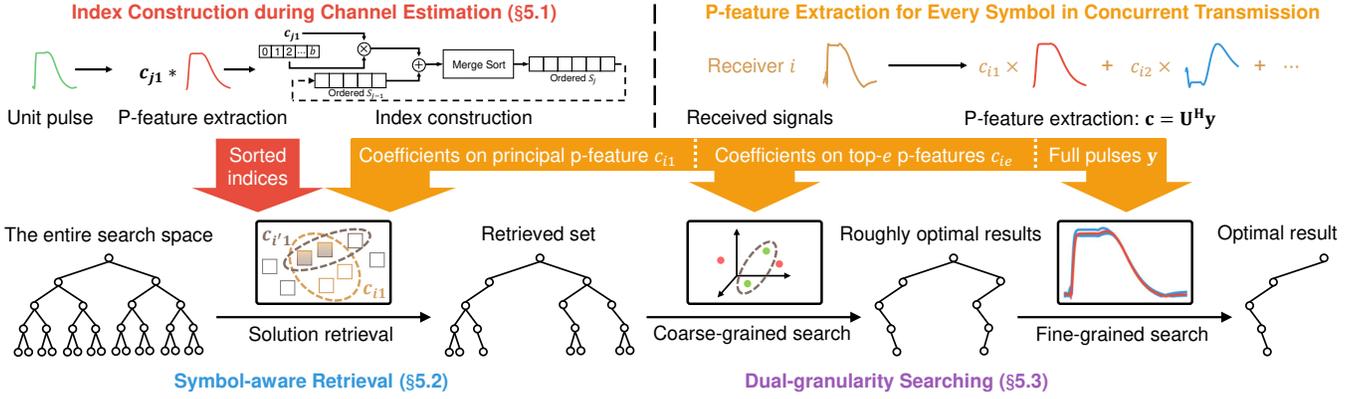


Figure 8: Overview of the demodulation algorithm.

Pulses for Concurrent Transmission. Pulse diversity enables the simultaneously-transmitted pulses to be distinguished. Given the pulse shapes (which can be obtained in the channel estimation process where each tag sequentially sends a pre-defined sequence of symbols; the details will be illustrated in §6) of the transmitting tags, all the possible received symbols can be calculated via the addition of the individual p-features, and thus the symbols for concurrent transmission can be calculated. Fig. 7 shows an example of pulse-superimposed symbols from 3 different tags. The symbols can be distinguished from each other with either pulse shapes (Fig. 7a) or the extracted p-features (Fig. 7b) so that the modulated bit of each tag is detected. Therefore, it is feasible to demodulate bits from the pulses of concurrent transmission with pulse diversity.

MIMO-like Formalization. The concurrent transmission design with pulses allows multiple tags to transmit their bits simultaneously. Meanwhile, multiple receivers can also receive superimposed pulses with receiver diversity. Therefore, we can formalize the communication system into a MIMO-like expression:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} \quad (2)$$

where each y_i is the superimposed pulse received by the i^{th} receiver, and x_j is the symbol transmitted at the j^{th} tag. The value of x_j is restricted to a fixed number of integers, i.e., from 0 to b (the number of independent LCM pixels), according to pulse amplitude modulation. p_{ij} is the *unit pulse*, defined as the pulse sent from one pixel of the j^{th} tag to the i^{th} receiver. We refer to Eqn. 2 as $\mathbf{y} = \mathbf{P}\mathbf{x}$, which is actually the formulation of MIMO illustrated in §2.2, for simplicity in the rest of this paper. Given the received signal \mathbf{y} and the pulse channel matrix \mathbf{P} , a demodulator can determine the transmitting vector \mathbf{x} as illustrated in §5.

5 LOW-LATENCY DEMODULATION

The brute-force MLD (Eqn. 1) fails to meet the low-latency requirements because: (1) it computes for every possible solution in the search space, and the space size grows exponentially with the degree of concurrency; (2) the default objective function that computes the Euclidean distance of the full pulses requires extensive computation when \mathbf{x} is of high dimension. In this section, we

describe our design that leverages p-features to accelerate MLD, as shown in Fig. 8. We first reduce the search space by constructing an *index* (i.e., the coefficient corresponding to the principal p-feature) for each possible solution during channel estimation. The indexing enables a symbol-aware solution retrieval that constructs a small search space (i.e., $< 5\%$ compared with the original one) for each received symbol. The next step is to refine the objective function with p-features for dual-granularity effective-and-efficient searching.

5.1 Index Construction with Principal P-feature

The size of the entire search space for the optimal solution (\mathbf{x}_{opt}) increases exponentially with the number of tags, leading to a prohibitive complexity. One important property of the p-feature extraction method is that the principal p-feature (i.e., the leftmost column of \mathbf{U} in §4.2, \mathbf{U}_1) contributes most to a pulse. Therefore, we can take the coefficient of the principal p-feature in $\mathbf{P}\mathbf{x}$ as the index² for each \mathbf{x} , i.e., $c_1 = \mathbf{U}_1^H \mathbf{P}\mathbf{x}$. The indices represent the possible solutions \mathbf{x} to a large extent and enable a faster retrieval than the full pulses (of high dimensions) if the one-dimensional indices are sorted.

The indices consist of c_1 for every possible \mathbf{x} , defined as a sequence \mathcal{S}_n given n tags. Each tag j can only transmit a fixed number of amplitudes, i.e., $x_j = 0, 1, \dots, b$. Therefore, the indices for tag j can be represented as $c_{j1} = \{x_j d_j\}$, where d_j is the index of the unit pulse, i.e., $d_j = \mathbf{U}_1^H \mathbf{p}_j$. As a result, $\mathcal{S}_n = \{\sum_{j=1}^n c_{j1}\} = \{\sum_{j=1}^n x_j d_j\}$ where x_j is an arbitrary integer from 0 to b .

We design an efficient index construction algorithm with sequential merging to construct a sorted \mathcal{S}_n . Our algorithm takes the advantage that each tag transmits its channel estimation symbols exclusively and sequentially (will be illustrated in §6). Once the first tag finishes its channel estimation, we can calculate the initial subsequence $\mathcal{S}_1 = \{\sum_{j=1}^1 x_j d_j\} = \{0d_1, 1d_1, \dots, bd_1\}$, which is inherently sorted. When the k^{th} tag takes its turn, the subsequence \mathcal{S}_{k-1} is ready and sorted. The next step is to add c_{k1} to the items in \mathcal{S}_{k-1} . We can obtain the subsequences $\{0d_k + \mathcal{S}_{k-1}\}, \{1d_k + \mathcal{S}_{k-1}\}, \dots, \{bd_k + \mathcal{S}_{k-1}\}$, each of which is sorted. In order to merge the subsequences to a sorted \mathcal{S}_k , we use the loser tree data structure that ensures the complexity of the index construction to be $O(m)$, given m is the number of possible solutions that grows

²The $\mathbf{P}\mathbf{x}$ from each receiver has a separate index. §5.2 will discuss how to jointly use the indices from multiple receivers.

exponentially to the concurrency. In contrast, the complexity of the straightforward compute-and-sort design is $O(m \log m)$.

5.2 Symbol-aware Retrieval of Possible Solutions

We design a symbol-aware retrieval mechanism to retrieve the possible solutions from the sorted indices. It dynamically decides the index range of retrieved possible solutions according to the received symbol. We first retrieve a set \mathcal{N}_i for each independent receiver i . Specifically, for $\hat{\mathbf{x}}$ whose index is closest to $\mathbf{U}_1^H \mathbf{y}_i$, we first compute its distance to $\mathbf{U}^H \mathbf{y}_i$ on all the p-features by $\hat{d} = \|\mathbf{U}^H \mathbf{P} \hat{\mathbf{x}} - \mathbf{U}^H \mathbf{y}_i\|_2$. \hat{d} is the upper bound of the distance between \mathbf{x}_{opt} and $\mathbf{U}_1^H \mathbf{y}_i$ on the index, i.e., $\|\mathbf{U}_1^H \mathbf{P} \mathbf{x}_{\text{opt}} - \mathbf{U}_1^H \mathbf{y}_i\|_2 \leq \hat{d}$. Therefore, we can retrieve a set \mathcal{N}_i that includes all the \mathbf{x} satisfying $\|\mathbf{U}_1^H \mathbf{P} \mathbf{x} - \mathbf{U}_1^H \mathbf{y}_i\|_2 \leq \hat{d}$.

Multiple receivers can be leveraged to promote diversity and further narrow the retrieval range. Receiver diversity (§4.1) enables multiple independent receivers to have different sets of possible solutions. With the adjacent index set \mathcal{N}_i for each \mathbf{y}_i from multiple receivers, the optimal \mathbf{x}_{opt} should coexist in all these sets, whereas the majority of distractors exist in fewer sets. Therefore, we can take the intersection $\cap_i \mathcal{N}_i$ as the retrieval result. The retrieved solutions are taken as the reduced search space that is empirically much smaller than the original search space (will be evaluated in §8.3). The \mathbf{x}_{opt} is to be exhaustively searched from the space with the objective function(s).

5.3 Dual-granularity Search

We take two objective functions to search the \mathbf{x}_{opt} with Eqn. 1: $D_1(\mathbf{x})$ as a coarse-grained one and $D_2(\mathbf{x})$ as a fine-grained one. $D_1(\mathbf{x})$ calculates the distance between the coefficient vectors of the received signal and a possible solution with a few p-features, which is low-dimensional and relatively fast to compute. It provides tolerable accuracy due to the effectiveness of p-feature extraction. $D_2(\mathbf{x})$ deals with the full pulses and thus provides higher accuracy. Our dual-granularity search algorithm incorporates the two functions to keep a balance between efficiency and accuracy.

Coarse-grained Search with P-feature Coefficients. The coarse-grained objective function $D_1(\mathbf{x}) = \|\mathbf{U}_e^H \mathbf{P} \mathbf{x} - \mathbf{U}_e^H \mathbf{y}\|_2$ is constructed upon the observation that p-features can approximate full pulses efficiently. As illustrated in Fig. 4, $\mathbf{U}_e^H \mathbf{p}$ is an explanatory and efficient fingerprint of \mathbf{p} , where e is the number of the most significant p-features involved (to be evaluated in §8.3). With $D_1(\mathbf{x})$, the coarse-grained searching selects a small number (typically 5 to 10) of solutions and leaves them for fine-grained searching.

Fine-grained Search with Full Pulses. The fine-grained objective function $D_2(\mathbf{x}) = \|\mathbf{U}_e \mathbf{U}_e^H \mathbf{P} \mathbf{x} - \mathbf{y}\|_2$ calculates the Euclidean distance between a possible symbol and the received full pulse \mathbf{y} . The expected pulse signals of transmitted symbols \mathbf{x} are reconstructed using the e most significant p-features \mathbf{U}_e . The insight here is that \mathbf{U}_e is decomposed from multiple pulses with diversity and thus alleviates random noise. With a proper value of e , $\mathbf{U}_e \mathbf{U}_e^H \mathbf{P} \mathbf{x}$ can alleviate the noise effect as well as maintain a precise resolution for the details of the pulse diversity. With $D_2(\mathbf{x})$, the fine-grained searching finally outputs the demodulation result \mathbf{x}_{opt} .

6 MAC PROTOCOL

This section presents our MAC protocol design that turns PHY concurrency into low-latency networking. We first raise several

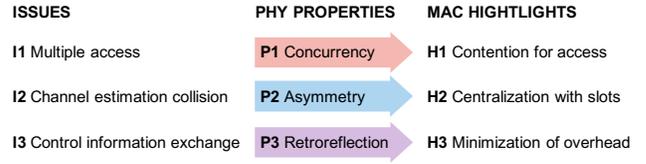


Figure 9: The outline of MAC design.

Issues to realize the MAC protocol in an IoT network. Next, we revisit the related PHY Properties and reveal our MAC Highlights, i.e., how the MAC layer utilizes these PHY properties to solve issues. Fig. 9 summarizes the design outline.

Assuming there is a dynamic number of tags. It is necessary to satisfy the multiple access because the potentially-tremendous networking requirements (e.g., 64 tags or even more) may exceed PHY concurrency capacity (e.g., 8 concurrent links) (Issue 1). The reader has to take channel estimation every time querying the tag due to the possible channel shift from mobility. The channel estimation before communication is sequential to obtain the pulse shape of every tag respectively. It is challenging to manage this sequential channel estimation with minimum collision overhead (Issue 2). Reliable multiple access requires additional control information exchange, e.g., Automatic Repeat reQuest (ARQ). Our MAC protocol should minimize its side effect on latency (Issue 3).

6.1 Design Highlights

Our MAC addresses the above issues and achieves low-latency networking from a PHY-property-directed design. We revisit three key PHY properties of RETROMUMIMO: *Property 1: Concurrency*. Multiple tags can transmit uplinks simultaneously with the pulse-division concurrent transmission, which is a key property for the MAC layer to leverage. *Property 2: Asymmetry*. As the master device, the reader can broadcast messages to all the tags within its FoV, which is convenient for centralized synchronization and scheduling. *Property 3: Retroreflection*. The retroreflection property guarantees that all the valid links are between the reader and tags and that no inter-tag communication exists. Therefore, the hidden terminal problem does not harm the network performance. Besides, an uplink transmission, which reflects the light carrier signals from the reader, is a natural ACK of the downlink.

Highlight 1: Contention-based Multiple Access. RETROMUMIMO exploits a contention-based protocol to support multiple access even when the number of tags exceeds the concurrency limit. (Issue 1). Each tag sends its message as soon as possible without queueing in a cyclic order (like Token Ring) or proactively attempting to avoid collisions (like CSMA/CA). As a result, latency is reduced compared with protocols based on polling or carrier sensing. Collision is the primary factor to degrade the performance of a contention-based protocol, but the PHY concurrency (*Property 1*) reduces the probability of destructive collisions.

Highlight 2: Centralized and Slotted Protocol. We design a centralized and slotted protocol to coordinate the tags during channel estimation (Issue 2). We divide the time resource into slots to reduce the probability of collisions, at a cost of precise time synchronization across the reader and tags. However, it is naturally supported by the asymmetric characteristic (*Property 2*). Therefore,

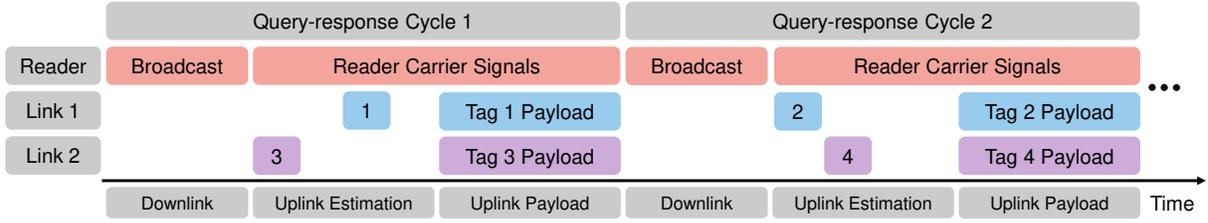


Figure 10: An example of MAC query-response cycles. There are two concurrent uplinks but four tags. Tag 1 and tag 3 have transmission requirements during cycle 1 so they send channel estimation sequences at a random estimation slot and then send payload simultaneously. Tag 2 and tag 4 have transmission requirements during cycle 2 and execute a similar process to cycle 1.

the reader can centrally schedule the slotted communication with its downlink broadcast.

Highlight 3: Overhead Minimization. Our protocol is built upon minimum control information exchange to minimize protocol overhead (*Issue 3*) and still guarantees reliability. We reuse the downlink broadcast as beacons for synchronization. Our error-control method does not require tag-to-reader ACKs because the tags are triggered only by the reader, so a message from a tag indicates a successful downlink broadcast. The reader-to-tag ACKs are piggybacked into the broadcast. The merging does not influence reliability because a tag can only receive the reader’s broadcast that contains the ACK of the last message (*Property 3*).

6.2 Workflow and Protocol State Machine

Following the design highlights, we specify a centralized, slotted, and contention-based MAC protocol. It divides the time resource into a series of *query-response cycles*, each with three stages, *Downlink Broadcast*, *Uplink Channel Estimation*, and *Uplink Payload*. We give an expository example of the workflow with two MAC cycles in Fig. 10. The protocol, including query-response cycles and collision handler, is internally driven by a state machine, which is shown in Fig. 11. We will illustrate how it works as follows.

Downlink Broadcast (Red). The reader broadcasts to trigger a query-response cycle. All the tags are centrally managed by this broadcast according to *Highlight 2*. According to *Highlight 3*, the reader also broadcasts other information, including protocol parameters (will be discussed soon) and the ACK of the last cycle. If the expected ACK is not received, *i.e.*, the last transmission of a certain

tag failed, the tag will autonomously retransmit in the forthcoming stages or prepare to retransmit in the following cycles following the standard binary exponential backoff mechanism.

Uplink Channel Estimation (Blue). The channel estimation is contention-based according to *Highlight 1*. The reader reserves estimation slots and keeps sending carrier signals. Each tag that decides to transmit in this cycle randomly selects a slot to reflect the carrier signals to send its channel estimation sequences. There is no explicit ACK in this stage according to *Highlight 3*.

Uplink Payload (Purple). The reader still keeps sending carrier signals. The tags transmit their packets simultaneously, following the pulse-division concurrent transmission design.

Handling Collisions. Estimation sequence collision is the primary factor of decoding failure. If no collisions occur, the concurrent payload can be decoded successfully (if we do not consider the random PHY layer failures). If collisions occur, the collided estimation sequences disrupt the decoding process. A CRC failure indicates the collision (or other factors that interrupt/interfere with the link), and the reader will notify this failure to all the tags in the next downlink broadcast.

Analysis of Parameters. Two major parameters influence MAC performance. (1) The number of reserved slots for channel estimation N_{trn} influences the probability of collisions, which is further to be evaluated in §8.4.1. (2) The backoff mechanism requires the tag to retransmit with a random delay from the uniform distribution in $[1, 2^{N_{retx}}]$, where N_{retx} is the number of past retransmissions of the packet. To avoid congestion, a packet will be discarded if N_{retx} exceeds a threshold.

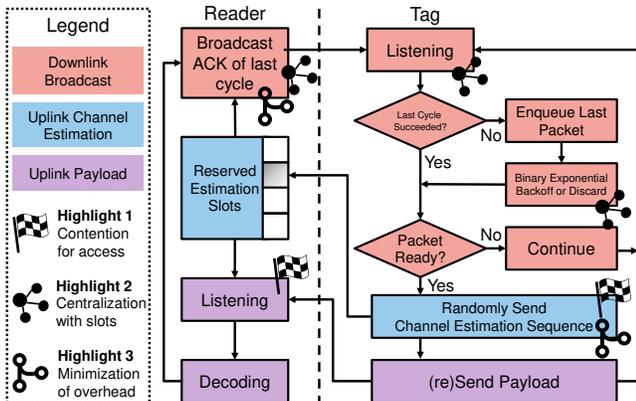


Figure 11: Protocol state machine of the workflow.

6.3 Message Structures

Fig. 12 shows the message structures of every stage in one query-response cycle. (1) Downlink Broadcast. The reader broadcasts information, including preamble, packet length, MAC parameters (*i.e.*, 6 bits of the number of estimation slots N_{trn} , 2 bits of maximum retransmission N_{retx}), ACK of the last cycle, broadcast payload,

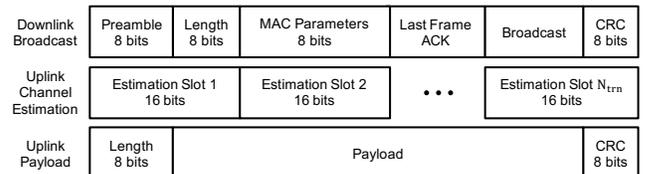


Figure 12: Message structures.

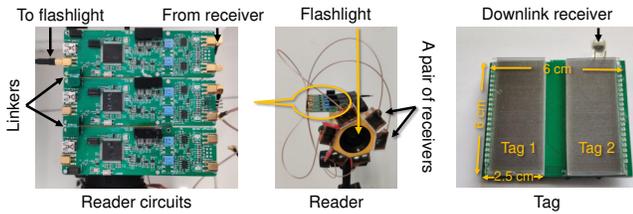


Figure 13: RETROMUMIMO prototype.

and CRC. (2) Uplink Channel Estimation. There are N_{trn} reserved channel estimation slots, which are already synchronized with tags in the downlink broadcast. Each tag with a packet to send will pick a random estimation slot and send its channel estimation sequence. (3) Uplink Payload. The tags simultaneously send their packets (*i.e.*, packet length, payloads, and CRC) to the reader. The payload length is variable, so multiple packets can be aggregated in one payload to handle a large amount of traffic.

7 IMPLEMENTATION

As shown in Fig. 13, RETROMUMIMO has two types of devices: reader and tag. The reader is composed of a commercial LED flashlight (power of 4 W and FoV of 20°), optical receivers, and a few customized circuit boards connected to an external computer. The tag is an MCU-equipped circuit board that periodically receives downlink broadcasts and transmits uplink packets with LCMs.

Reader. The optical receiver incorporates polarization reception design [12] to receive arbitrary-polarized signals and filter out the unpolarized ambient light noise. Three pairs of receivers form a receiver ring with 120° between each other; all the receivers are stuck to a 3D-printed case surrounding the flashlight. The uplink reception circuits include switching carrier, passband receiver, gain control, down-conversion, and decimation. We design a board-to-board linker that synchronizes the trigger and clock of multiple boards in a modular and elastic approach. The reader circuits stream the received samples to a laptop computer with an Intel i5-8265U@3.9GHz CPU and 8 GB RAM for demodulation.

Tag. Each tag is equipped with a 4-pixel LCM for modulation measuring 6 cm by 2.5 cm. The power consumption of one tag is 0.2 mW during transmission. Two tags are connected to one backend circuit board to reduce circuit complexity and simplify the experimental setup. They share the same downlink receiver (an S6036 photodiode) and an STM32-L4 MCU for decoding downlink broadcast, but they transmit uplinks under the control of two independent sets of shift registers connected to the MCU.

8 EVALUATION

8.1 Experimental Setup

We conduct experiments to validate the effectiveness and efficiency of our prototype communication system, as shown in Fig. 14. We first collect the training set \mathbf{R} for p-feature decomposition. For the evaluation, we only collect \mathbf{R} once and use it in all the experiments. The matrix \mathbf{R} is composed of a total number of 43200 pulses, consisting of pulse diversity from 3 receivers, 6 tags, 6 positions, and 400 pulses for each case. Multiple pixels of the same tag transmit pulses simultaneously. We believe this dataset is not an enormous one, considering each pulse is only 4 ms long. However, \mathbf{R} shows

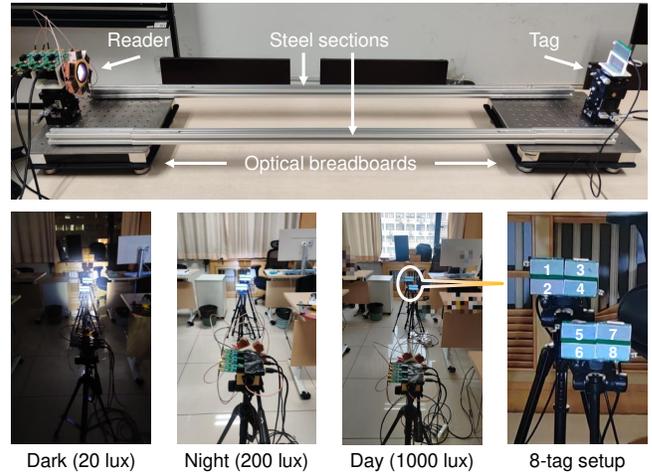


Figure 14: Optical platform and experimental setup.

adequate generality when dealing with pulses from different tags, positions, lighting conditions, and nearby environments that are not included in the training set.

During the experiments, each tag transmits at a 500-bps data rate, composed of 250 pulses per second and 2 bits per pulse. By default, 8 closely spaced tags are located on two tripods without distinct angle differences (but not precisely calibrated) in an illuminated office at night (around 200 lux). All experiments are set up in typical indoor environments including common objects with reflective surfaces such as glass. People are free to move during the experiments as long as the light path is not blocked. Each point includes 25 packets of 8 bytes (1,600 bits) from each tag, and the average bit error rate (BER) of all tags is a core indicator for link quality. We define the reliable working range to be the communication distance with less than 1% BER, as has been proved when the classic Reed-Solomon error correction code is applied [12].

8.2 Concurrency Performance

In this part of the evaluation, we mainly focus on the physical layer properties of RETROMUMIMO. We disable the MAC layer and manually arrange the tags for concurrent transmission, *i.e.*, all the transmissions are concurrent.

Number of Tags. As shown in Fig. 15a, RETROMUMIMO supports simultaneous uplink transmission within 2.50 m for 8 tags, 3.25 m for 6 tags, and 3.75 m for 4 tags. More tags incur shorter inter-symbol distance in the pulse domain and thus require a higher signal-to-noise ratio (SNR) to demodulate. The pulse-division concurrent transmission with our p-feature extraction and demodulation is effective to support up to 8 concurrent links.

Number of Receivers. More receivers will promote receiver diversity and we evaluate the impact of this diversity in Fig. 15b where 6 concurrent tags are demodulated with different numbers of receivers. The results show that the BER less than 1% is almost infeasible with 1 receiver, but is achievable for both 2 and 3 receivers within the 3.25 m working range. Therefore, RETROMUMIMO would benefit from a higher integration level of receiver design, which is currently restricted by the size of photodiodes.

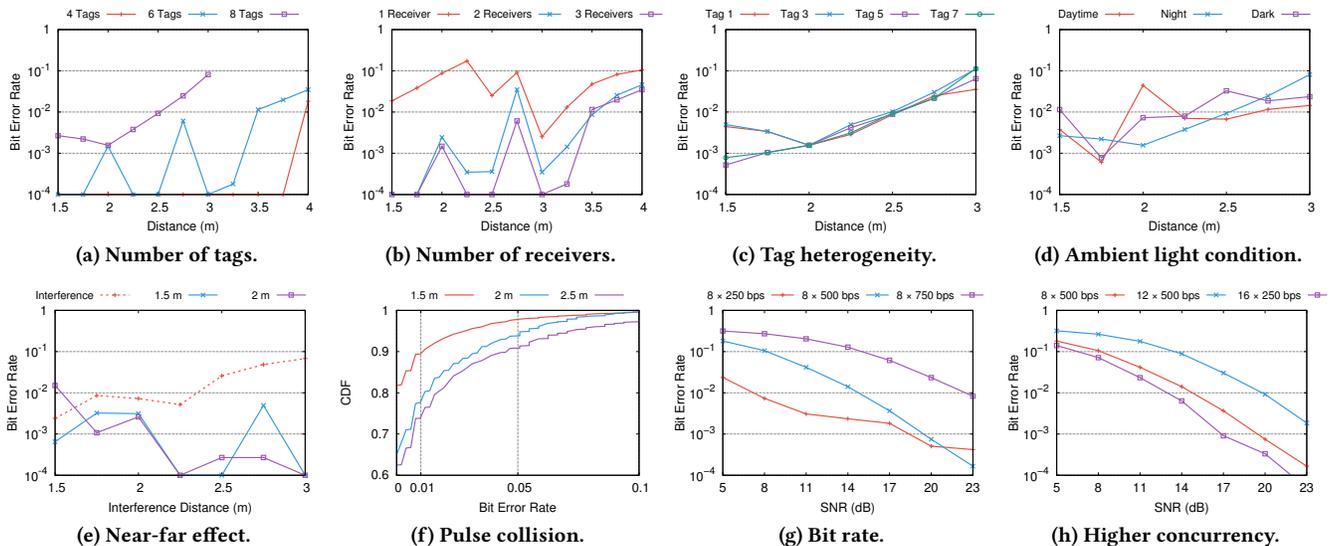


Figure 15: Concurrency performance results.

Tag Heterogeneity. Different tags participating in a concurrent transmission might have different BER performances. During the 8-concurrent transmission with the same arrangement as Fig. 14, we inspect the BER performance of 4 tags out of the 8 concurrent tags (tag 1, 3, 5, and 7 shown in Fig. 14, which are scattered in the FoV). Fig. 15c shows that the BER performances are close to each other in general, yet a slight difference can be observed when the distance is less than 2 m where the BER is below 1%. Therefore, we believe our p-feature extraction approach is effective among the tags with different tag contexts.

Ambient Light Condition. We conduct experiments in different indoor lighting environments, and the results shown in Fig. 15d indicate that RETROMUMIMO is robust to provide reliable links within 2.5 m in various indoor illumination environments including daytime, night, and dark. Ambient light as background noise may influence common visible light communication but it has little effect on RETROMUMIMO because only the carrier signal frequency is allowed to pass our optical receivers. Besides, the p-feature extraction method provides additional resistance against noise.

Near-far Effect. The near-far effect, which means the signals from a nearby tag are stronger than a farther tag, may harm the performance of concurrent communication systems because the far tags are likely to be overpowered [18, 28]. To evaluate how this effect influences RETROMUMIMO, we regard four tags as interference and vary their distances, and fix the other four tags to test the BER performance. As shown in Fig. 15e, when fixed at 1.5 m or 2 m, the tags mostly provide reliable links with BER less than 1% regardless of the distance of the interference. The BER of the fixed tags drops as the interference becomes weaker (*i.e.*, the distance of interference increases). This result indicates that RETROMUMIMO is robust against the near-far effect and provides robust links to both the nearby and the farther tags in the reliable working range.

Pulse Collision. As illustrated in §4.1, a slight orientation or position movement of a tag may incur different pulse shapes. One

may have noted that there are abrupt BER increases and falls in the evaluation results. We attribute the phenomenon to the fact that a certain group of tags can not transmit concurrently after slight movement because a certain combination of symbols pulses may collide with others. To evaluate the probability of this effect, we set the tags to different orientations (including yaw and roll angles) and positions in the field of view at 3 distances (1.5 m, 2 m, 2.5 m) to the reader, and collect 200 sequences of channel estimation sequences and modulated payloads. The minimum position difference of the tags is smaller than the size of a tag. We generate and demodulate 2000 groups of the transmissions that are superimposed of 8 randomly-chosen stand-alone sequences for each distance. Fig. 15f shows the CDF of the BER results. The majority of tag arrangements (89% for 1.5 m, 77% for 2m, and 73% for 2.5 m) allow a reliable link with BER less than 1%, whereas 5% position arrangements incur the BER significantly larger than the average. This result demonstrates that RETROMUMIMO provides reliable concurrent transmissions for most pulses in different contexts of arbitrary tags. The concurrency failures from pulse collisions can be further handled by the random backoff of the MAC protocol.

Bit Rate. The bit rate of the concurrent links, which is determined by the modulation scheme, may have an impact on the overall BER and concurrency. To evaluate this effect, we carry out an emulation-based evaluation to study the BER-SNR relationship with different bit rates (by varying the amplitude levels in a pulse). Fig. 15g shows that compared with our default setup (eight 500-bps concurrent links), a lower bit rate requires lower SNR for demodulation, but the discount is mild when SNR is sufficient (*e.g.*, larger than 20 dB). Meanwhile, a higher bit rate requires higher SNR for successful demodulation, *e.g.*, extra 8 dB to achieve 1% BER for 750-bps links. Our current hardware prototype cannot support such a high SNR requirement. Besides, it is feasible for 1-kbps links to achieve 4-concurrency given the 20 dB SNR requirement.

Higher Concurrency. We further extend our emulation to explore the maximum number of tags RETROMUMIMO can support. As

	BER	Time (ms)
Our method	0.0021	56.07
Performance breakdown		
w/o fine-grained searching	0.0133	50.41
w/o coarse-grained searching	0.0027	464.49
w/o index	0.0022	1542.88
w/o index w/o fine-grained searching	0.0136	1403.76
w/o index w/o coarse-grained searching	0.0026	23030.61
Number of p-features e		
1	0.1852	46.42
2	0.0949	47.07
4	0.0048	49.56
8 (selected)	0.0021	56.07
16	0.0031	65.43
32	0.0029	91.06
320	0.0032	580.16
Existing methods		
V-BLAST with 8 p-features	0.3799	4.81
V-BLAST with full pulses	0.3767	5823.43
Sphere decoder with 8 p-features	0.0090	116.96
Sphere decoder with full pulses	0.0090	140.50

Table 1: Demodulation performance results.

shown in Fig. 15h, when maintaining the data rate of 500 bps for each tag, 5dB additional SNR is required to increase the tag number from 8 to 12, which is beyond the capabilities of our prototype. However, up to 16 tags can transmit concurrently if the data rate per tag is halved to 250 bps, with a 1.5 dB lower SNR requirement. Therefore, the pulse-division concurrent transmission could scale to tens of tags or even more with proper settings.

8.3 Demodulation Performance

We evaluate the performance of our demodulation algorithm extensively and show the results in Tab. 1. We use different algorithms to demodulate the 8-tag concurrent communication with the distance of 1.5 m. Our algorithm costs 56.07 ms to demodulate a 64-bit 8-concurrency packet, which requires 128 ms for payload transmission. Therefore, the algorithm achieves real-time demodulation for our prototype. We further analyze the performance in depth and compare the method with existing methods.

Performance Breakdown. We evaluate the performance of our demodulation algorithm by selectively disabling the optimizing schemes. The symbol-aware retrieval with constructed indices reduces demodulation time by more than 96%. In addition, the coarse-grained searching reduces time consumption by 88%. Finally, the fine-grained searching guarantees an accuracy level similar to the brute-force MLD (*i.e.*, without indexing and coarse-grained searching for acceleration), at a cost of 11% time overhead compared with only coarse-grained searching.

Number of P-features e . The dual-granularity searching utilizes multiple p-features to reduce the searching space and alleviate the noise. The best BER performance is achieved with 8 p-features, indicating that $e = 8$ is the optimal parameter selection in §5.3. A fingerprint constructed from insufficient p-features represents a

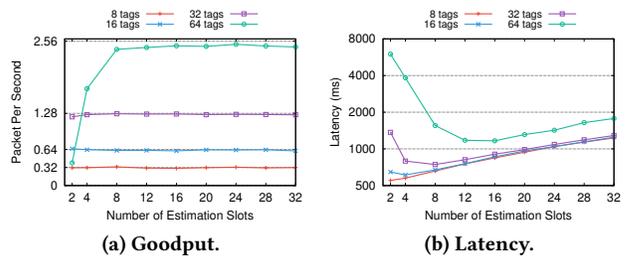


Figure 16: Microbenchmark of parameter N_{trn} .

pulse with fewer dimensions (*i.e.*, less information), thus increasing the probability of pulse collisions. In contrast, the coefficients of insignificant p-features are influenced more severely by the disturbance of random noise and reduce the demodulation accuracy.

Comparison with Existing Methods. We implement a linear detector (V-BLAST [36]) and a low-complexity MLD (sphere decoder [31]) to compare their performances in RETROMUMIMO with our algorithm³. A linear detector can not deal with the strongly correlated channel matrix (either from only a few p-features or from the full pulses), and thus amplifies noise and mixes up different symbols. Meanwhile, to prune the symbol-wise tree traversal, a sphere decoder uses QR factorization of the channel matrix, which is rank-deficient in our case and leads to unstable results. To our knowledge, this problem can be readily solved either when the number of transmitters is more than the receivers [38, 39] or with constant modulus constellation [32], which are not applicable in our case. Therefore, the two algorithms achieve worse BER performance with higher time consumption. Our algorithm turns the rank deficiency into the optimizations from p-features to achieve superior performance.

8.4 Latency Performance

We further complement our experimental evaluation with extensive simulation to study the latency performance with the MAC layer. The simulator implements the MAC protocol and follows the parameters of the PHY experimental results. Specifically, the reader supports up to 8 concurrent links, and the bit rate for each link is 500 bps. The payload length is 16 bytes by default. The downlink broadcast costs 50 ms and one uplink channel estimation slot is 16 ms. Therefore, the time of one query-response cycle is $50 + 16N_{trn} + 16 \times 8/0.5$ ms. Demodulation latency is not included because our algorithm achieves real-time demodulation. A packet will be discarded if it has been retransmitted more than 4 times.

We set up the network with n tags and one reader. Every tag produces packets with the probability P_s in one second, which leads to a Poisson distribution of transmission attempt number k in the whole network. Network traffic (or carrier load) G is the mathematical expectation of k , *i.e.*, $G = E(k) = P_s \times n$. We test the impact of the parameter N_{trn} , evaluate the goodput (the number of packets arriving at the reader per second) and average latency performance with different carrier loads and payload lengths, and compare RETROMUMIMO with the state-of-the-art design.

³V-BLAST associates interference cancellation with a linear detector, and a sphere decoder accelerates MLD with tree searching and pruning. The taxonomy of MIMO detectors is more complex than what is covered in this paper [37]. The classifications of our baselines may change depending on the context, but we believe they are competent representatives of non-ML and ML detectors for MIMO.

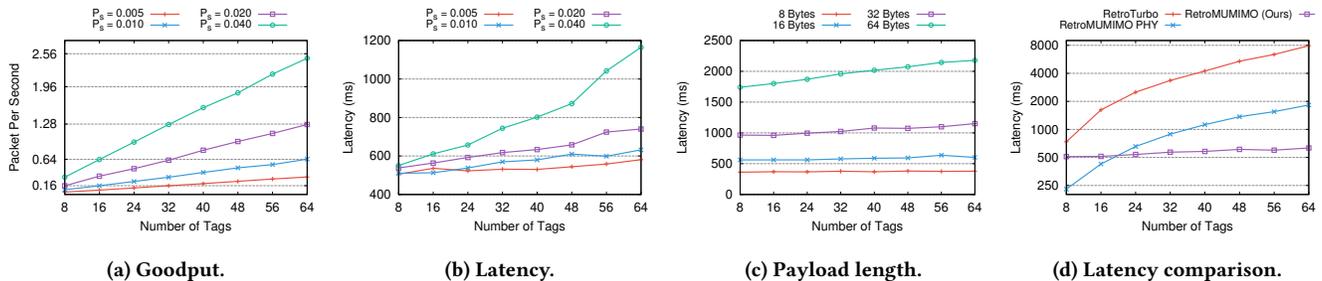


Figure 17: Latency performance results.

8.4.1 Microbenchmark of the Parameter N_{trn}

We evaluate how the number of uplink channel estimation slots N_{trn} influences the goodput and latency performance. We set P_s to 0.04 during the microbenchmark.

Goodput. We take packet per second (PPS) as the metric to evaluate the goodput of the network. When there are less than 32 tags in the network, the goodput is close to the mathematical expectations of carrier load G , which are marked as y-axis scales in Fig. 16a. In this case, the packet loss rate is close to zero even with arbitrary N_{trn} . When there are more tags in the network (e.g., 64 tags), the goodput can be significantly less than G if N_{trn} is insufficient. As a result, an adequate N_{trn} is necessary to get the largest possible goodput and ensure a low packet loss rate.

Latency. We evaluate the latency performance by analyzing the latency of the goodput packets. As shown in Fig. 16b, the latency performance changes non-linearly with N_{trn} increasing. When there is only a small number of tags (e.g., 8 tags) in the network, channel estimation collisions are infrequent so a small N_{trn} works well. As the network size grows larger, a larger N_{trn} is needed to reduce collisions. An excess of reserved estimation slots may reduce the collision probability to near zero, but the idling slots increase the latency linearly. Therefore, a suitable N_{trn} is important to ensure the lowest possible latency.

8.4.2 MAC Performance

We evaluate the performance of our MAC protocol with different carrier loads and payload lengths. Based on the microbenchmark, we fix $N_{trn} = 8$, which works well for the networks with different numbers of tags from 8 to 64.

Carrier Load. We evaluate the goodput and latency performance with different carrier loads. The goodput results in Fig. 17a are close to the corresponding carrier loads in most cases, i.e., the loss rates are near zero. As shown in Fig. 17b, the latency almost keeps constant as the number of tags increases when $P_s = 0.005$. Latency increases almost linearly as the carrier load grows larger and the collisions occur more frequently. The results indicate our MAC protocol consistently utilizes the PHY concurrency to achieve low-latency networking.

Payload Length. We also evaluate the latency performance with different payload lengths. As shown in Fig. 17c, the rising tendencies of latency with different payload lengths are similar, only with a near-constant latency increase due to the additional transmission time for a longer payload. Therefore, the MAC protocol allows a flexible length of the payload.

8.4.3 Comparison with State-of-the-art

We further compare RETROMUMIMO with the status-quo designs. The decentralized contention-based protocol [11] is only applicable to small-scale networks (e.g., 70 seconds to scan 8 tags), so we do not consider it as a benchmark. We select RetroTurbo [12] as the baseline, which has a faster PHY link (8 kbps, higher than the maximum sum rate of RETROMUMIMO, i.e., 500 bps \times 8) and arranges the links with a polling-style protocol. We also implement the polling protocol upon the pulse-division concurrent links, i.e., RETROMUMIMO PHY, as another baseline. We set P_s to 0.01 to make the goodput results all close to the carrier load with near-zero packet loss. Fig. 17d shows the latency performances and we make the following comparisons.

Comparison of PHY. We compare the latency performance of RETROMUMIMO PHY and RetroTurbo, both equipped with a polling-style protocol. Without the MAC protocol, RETROMUMIMO still reduces network latency by 68.9% (8 tags) – 76.9% (64 tags) even with lower PHY data rates. This result indicates that concurrency is the key to enabling low-latency networks for VLBC by minimizing the delay in waiting for other tags to exit the channel.

Comparison of PHY+MAC. When incorporated with the MAC, RETROMUMIMO further achieves 30.6% (8 tags) – 92.0% (64 tags) latency reduction. As the network scales up, the latency of our MAC protocol grows slowly from 510.2 ms (8 tags) to 631.9 ms (64 tags). In contrast, the latency of RetroTurbo grows approximately linearly from 735.6 ms (8 tags) to 7903.6 ms (64 tags). Therefore, our MAC design is especially suitable for large-scale networks (e.g., more than 16 tags) in IoT scenarios.

9 DISCUSSION

Robustness of LCM Pulse Diversity. As illustrated in §8.1, the p-feature extraction method does not require an exhaustive collection of pulses to construct the training set \mathbf{R} . During our experiments, the training set composed of 43200 pulses is compatible with new tags, receivers, relative positions, and environments. The pulse shapes from one tag do not change over time with our modulation for a few minutes. Temperatures may change the response time of liquid crystals [40], but we do not find any noticeable effect of ambient temperatures during the experiments in a typical office. We believe our method still works in more practical and challenging setups, such as with various temperatures and hardware devices, yet more pulses might be required to construct a more comprehensive training set \mathbf{R} in these complex setups.

Mobility of Devices. The pulse-division concurrent transmission assumes the channel is quasi-static in one query-response cycle (for

tens of milliseconds). Our MAC protocol provides a coarse-grained solution for mobility with per-cycle channel estimation. The estimation allows temporary movement and blockage out of the cycles, as is common in IoT networks. One potential solution to support continuous mobility (e.g., for vehicular or drone networks) is to construct a learning-based Bayes model of LCM pulse diversity, such that the channel information can be estimated online and updated along with the payload. A fundamental approach is to develop an analytical model that captures the basic physical characteristics of the pulses to analyze and investigate the pulse. Such a model is still missing in today’s literature to our knowledge.

Alternative Diversity Techniques. The pulse diversity leveraged by RETROMUMIMO is essentially a time-amplitude diversity of symbols. A few other techniques may also be selected, but are less suitable. The CDMA-like concurrency, which assigns orthogonal codes to the tags, has been experimentally evaluated to be inefficient for backscatter because the long coding sequence hurts throughput significantly, especially for the LCMs with even slower data rates compared with RF backscatters [25]. Another type of alternative diversity technique originates from visible light. An advanced optical modulator may have the capability of modulating with polarization, color, light path, or spatial diversity to enable concurrent transmissions, yet such approaches require dedicated hardware modification of the VLBC systems and are still to be researched. However, RETROMUMIMO can be potentially integrated with such diversity techniques (once available) for higher concurrency.

Multiple Readers. Currently, RETROMUMIMO employs one reader to build an experimental network with multiple tags. Our MAC layer is designed to achieve a higher performance gain as the network size grows larger (e.g., 64 tags), which may exceed the coverage of one reader. In a real-world network, multiple readers can increase the network coverage (and illumination) for an extensive working range. RETROMUMIMO can be straightforwardly extended to multi-reader scenarios by connecting all the readers in parallel to make them work like one reader. Further, readers with proper coordination may enable multiple concurrent query-response cycles to enhance network concurrency and reduce latency. RETROMUMIMO can be potentially fused with existing multi-reader coordination solutions [41] to enable a larger-scale low-latency VLC networking.

10 RELATED WORK

Visible Light Backscatter Link and Networking. Visible light backscatter utilizes the different properties of LCMs to create asymmetric links. The modulation scheme has been developed from time-domain OOK variants [6, 42], frequency-domain modulation [43], and color-based modulation [34] on a single pixel, to pulse amplitude modulation [33] and time-polarization collaborative modulation [12] through multi-pixel orchestration. PhotoLink [44] replaces LCMs with sophisticated digital micro-mirror devices to achieve the highest ever 80-kbps data rate, whereas the narrow $\pm 1^\circ$ working angle prevents multiple-access networking in the real world. The performance improvement on the PHY layer (e.g., communication distance and data rate) pushes the development of visible light backscatter networking. Retrol2V [11] presents a contention-based decentralized MAC to enable multiple access. RETROMUMIMO focuses on the latency issue and is the first system that

supports low-latency VLBC networking with dedicated PHY and MAC layer designs including pulse-division concurrent transmission, low-latency demodulation, and a centralized MAC protocol.

Parallel Backscatter Communication. Recent work has been proposed to provide support for parallel decoding of RF-based backscatter communication systems, including the systems separating concurrent transmissions in the IQ domain [20–24], with dedicated sparse codes [25, 26, 28], in the time and frequency domain [17], using chirp spread spectrum modulation and ON-OFF keying [18], with multiple frequency channels [15], by modifying OFDMA with digital frequency synthesis [13, 14], by building spatial streams by antenna array and frequency agnostic property [16], and with specially-designed FFT bins and window function [19]. All these methods require diversity from modulation approaches beyond LCM’s capability. We construct and leverage the diversity in the pulse domain, which is essentially the diversity in temporal amplitude sequences based on the unique LCM properties.

MIMO Decoding Systems with MLD. An MLD is the optimal detector to decode MIMO symbols but suffers from considerable (exponential) computational complexity. Sphere decoders [30, 45–47] and simulated annealing heuristics [48] are designed to reduce complexity, yet require powerful GPUs or FPGAs for parallel optimization. Quantum annealing [49, 50] has been leveraged to speed up MLD, but only achieves remarkable performance enhancement on expensive and scarce quantum annealers. Our demodulation algorithm, which is still serial and has the potential for parallel programming, leverages the unique characteristics of p-features for acceleration and achieves superior accuracy and efficiency compared with the baseline decoders.

11 CONCLUSION

In this paper, we have considered how to investigate and leverage the “inaccessible” imperfections of hardware manufacturing and physical properties to realize system-level performance gains. At the first sight, the pulse diversity might be a suboptimal property because the communication link of LCM pulses could be disrupted by a sight disturbance. In general, the negative effects should be eliminated with well-designed algorithms [12]. In contrast, our design has successfully turned such imperfections into a concurrent transmission design and achieved system-level latency reduction. Our design chimes with a series of loosely-related research efforts, such as improving VLC reception with rolling-shutter effect [51], leveraging carrier frequency offset for BLE signal detection [52], and using out-of-band RFID fingerprints for identification [53]. We believe our design complements and refreshes this research trend to utilize physical-layer suboptimality for performance enhancement, and the trend may benefit future networking systems.

ACKNOWLEDGMENTS

We are grateful to the anonymous shepherd and reviewers for their constructive and insightful critique, which has helped us greatly improve this paper. We also thank Purui Wang for his insightful comments on the work. We are grateful to Zizheng Guo for his valuable suggestions for developing the demodulation algorithm. This work is supported in part by National Natural Science Foundation of China (Grant No. 62022005, 62272010, and 62061146001). Chenren Xu is the corresponding author.

REFERENCES

- [1] Chenren Xu, Lei Yang, and Pengyu Zhang. Practical backscatter communication systems for battery-free internet of things: A tutorial and survey of recent research. *IEEE Signal Processing Magazine*, 2018.
- [2] Vamsi Talla, Joshua Smith, and Shyamnath Gollakota. Advances and open problems in backscatter networking. *ACM GetMobile*, 2021.
- [3] Mohammad Hossein Mazaheri, Alex Chen, and Omid Abari. mmtag: a millimeter wave backscatter network. In *ACM SIGCOMM*, 2021.
- [4] Renjie Zhao, Purui Wang, Yunfei Ma, Pengyu Zhang, Hongqiang Harry Liu, Xianshang Lin, Xinyu Zhang, Chenren Xu, and Ming Zhang. Nfc+ breaking nfc networking limits through resonance engineering. In *ACM SIGCOMM*, 2020.
- [5] Reza Ghaffarivardavagh, Sayed Saad Afzal, Osmy Rodriguez, and Fadel Adib. Ultra-wideband underwater backscatter via piezoelectric metamaterials. In *ACM SIGCOMM*, 2020.
- [6] Xieyang Xu, Yang Shen, Junrui Yang, Chenren Xu, Guobin Shen, Guojun Chen, and Yunzhe Ni. Passivevlc: Enabling practical visible light backscatter communication for battery-free iot applications. In *ACM MobiCom*, 2017.
- [7] Miguel Chávez Tapia, Talia Xu, Zehang Wu, and Marco Zúñiga Zamalloa. Sunbox: Screen-to-camera communication with ambient light. *ACM IMWUT*, 2022.
- [8] Weizheng Wang, Qing Wang, Junwei Zhang, and Marco Zuniga. Passivevlp: Leveraging smart lights for passive positioning. *ACM Transactions on Internet of Things*, 2020.
- [9] Tianxing Li, Qiang Liu, and Xia Zhou. Practical human sensing in the light. In *ACM MobiSys*, 2016.
- [10] Yichen Li, Tianxing Li, Ruchir A Patel, Xing-Dong Yang, and Xia Zhou. Self-powered gesture recognition with ambient light. In *ACM UIST*, 2018.
- [11] Purui Wang, Lilei Feng, Guojun Chen, Chenren Xu, Yue Wu, Kenuo Xu, Guobin Shen, Kuntai Du, Gang Huang, and Xuanzhe Liu. Renovating road signs for infrastructure-to-vehicle networking: A visible light backscatter communication and networking approach. In *ACM MobiCom*, 2020.
- [12] Yue Wu, Purui Wang, Kenuo Xu, Lilei Feng, and Chenren Xu. Turboboosting visible light backscatter communication. In *ACM SIGCOMM*, 2020.
- [13] Renjie Zhao, Fengyuan Zhu, Yuda Feng, Siyuan Peng, Xiaohua Tian, Hui Yu, and Xinbing Wang. Ofdma-enabled wi-fi backscatter. In *ACM MobiCom*, 2019.
- [14] Fengyuan Zhu, Yuda Feng, Qianru Li, Xiaohua Tian, and Xinbing Wang. Digiscatter: efficiently prototyping large-scale ofdma backscatter networks. In *ACM MobiSys*, 2020.
- [15] Tanmoy Das and Prasun Sinha. Ads: accurate decoding of rfid tags at scale. In *ACM CoNEXT*, 2019.
- [16] Qianyi Huang, Guochao Song, Wei Wang, Huixin Dong, Jin Zhang, and Qian Zhang. Freescatter: Enabling concurrent backscatter communication using antenna arrays. *IEEE Internet of Things Journal*, 2020.
- [17] Xianjin Xia, Yuanqing Zheng, and Tao Gu. Ftrack: Parallel decoding for lora transmissions. In *ACM SenSys*, 2019.
- [18] Mehrdad Hesar, Ali Najafi, and Shyamnath Gollakota. Netscatter: Enabling large-scale backscatter networks. In *USENIX NSDI*, 2019.
- [19] Jinyan Jiang, Zhenqiang Xu, Fan Dang, and Jiliang Wang. Long-range ambient lora backscatter with parallel decoding. In *ACM MobiCom*, 2021.
- [20] Pan Hu, Pengyu Zhang, and Deepak Ganesan. Laissez-faire: Fully asymmetric backscatter communication. In *ACM SIGCOMM*, 2015.
- [21] Jiajue Ou, Mo Li, and Yuanqing Zheng. Come and be served: Parallel decoding for cots rfid tags. In *ACM MobiCom*, 2015.
- [22] Meng Jin, Yuan He, Xin Meng, Yilun Zheng, Dingyi Fang, and Xiaojiang Chen. Fliptracer: Practical parallel decoding for backscatter communication. In *ACM MobiCom*, 2017.
- [23] Chengkun Jiang, Yuan He, Meng Jin, Xiaolong Zheng, and Junchen Guo. Canon: Exploiting channel diversity for reliable parallel decoding in backscatter communication. In *IEEE ICNP*, 2018.
- [24] Meng Jin, Yuan He, Xin Meng, Dingyi Fang, and Xiaojiang Chen. Parallel backscatter in the wild: When burstiness and randomness play with you. In *ACM MobiCom*, 2018.
- [25] Jue Wang, Haitham Hassanieh, Dina Katabi, and Piotr Indyk. Efficient and reliable low-power backscatter networks. In *ACM SIGCOMM*, 2012.
- [26] Wei Sun. Towards parallel decoding with compressive sensing in multi-reader large-scale rfid system. In *IEEE RFID*, 2021.
- [27] Linghe Kong, Liang He, Yu Gu, Min-You Wu, and Tian He. A parallel identification protocol for rfid systems. In *IEEE INFOCOM*, 2014.
- [28] Nanhuan Mi, Xiaoxue Zhang, Xin He, Jie Xiong, Mingjun Xiao, Xiang-Yang Li, and Panlong Yang. Cbma: Coded-backscatter multiple access. In *IEEE ICDCS*, 2019.
- [29] Wenjie Zhou, Tanmoy Das, Lu Chen, Kannan Srinivasan, and Prasun Sinha. Basic: Backbone-assisted successive interference cancellation. In *ACM MobiCom*, 2016.
- [30] Konstantinos Nikitopoulos, Juan Zhou, Ben Congdon, and Kyle Jamieson. Geosphere: Consistently turning mimo capacity into throughput. In *ACM SIGCOMM*, 2014.
- [31] Michael Pohst. On the computation of lattice vectors of minimal length, successive minima and reduced bases with applications. *ACM SigSam Bulletin*, 1981.
- [32] Tao Cui and Chintha Tellambura. An efficient generalized sphere decoder for rank-deficient mimo systems. *IEEE Communications Letters*, 2005.
- [33] Sihua Shao, Abdallah Khreishah, and Hany Elgala. Pixelated vlc-backscattering for self-charging indoor iot devices. *IEEE Photonics Technology Letters*, 2017.
- [34] Seyed Keyarash Ghiasi, Marco A. Zúñiga Zamalloa, and Koen Langendoen. A principled design for passive light communication. In *ACM MobiCom*, 2021.
- [35] Ang Li and Christos Masouros. Hybrid precoding and combining design for millimeter-wave multi-user mimo based on svd. In *IEEE ICC*, 2017.
- [36] Peter W Wolniansky, Gerard J Foschini, Glen D Golden, and Reinaldo A Valenzuela. V-blast: An architecture for realizing very high data rates over the rich-scattering wireless channel. In *URSI international symposium on signals, systems, and electronics*, 1998.
- [37] Shaoshi Yang and Lajos Hanzo. Fifty years of mimo detection: The road to large-scale mimos. *IEEE communications surveys & tutorials*, 2015.
- [38] Mohamed Oussama Damen, Karim Abed-Meraim, and J-C Belfiore. A generalized lattice decoder for asymmetrical space-time communication architecture. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000.
- [39] Mohamed Oussama Damen, Hesham El Gamal, and Giuseppe Caire. On maximum-likelihood detection and the search for the closest lattice point. *IEEE Transactions on information theory*, 2003.
- [40] Yosuke Iwata, Mitsuhiro Murata, Kohei Tanaka, Tadashi Ohtake, Hidefumi Yoshida, and Koichi Miyachi. Novel super fast response vertical alignment-liquid crystal display with extremely wide temperature range. *Journal of the Society for Information Display*, 2014.
- [41] Jona Beysens, Ander Galisteo, Qing Wang, Diego Juara, Domenico Giustino, and Sofie Pollin. Densevlc: A cell-free massive mimo system with distributed leds. In *ACM CoNEXT*, 2018.
- [42] Jiangtao Li, Angli Liu, Guobin Shen, Liqun Li, Chao Sun, and Feng Zhao. Retrovlc: Enabling battery-free duplex visible light communication for mobile and iot applications. In *ACM HotMobile*, 2015.
- [43] Ye-Sheng Kuo, Pat Pannuto, Ko-Jen Hsiao, and Prabal Dutta. Luxapose: Indoor positioning with mobile phones and visible light. In *ACM MobiCom*, 2014.
- [44] Talia Xu, Miguel Chávez Tapia, and Marco Zúñiga. Exploiting digital micro-mirror devices for ambient light communication. In *USENIX NSDI*, 2022.
- [45] Christopher Husmann, Georgios Georgis, Konstantinos Nikitopoulos, and Kyle Jamieson. Flexcore: Massively parallel and flexible processing for large mimo access points. In *USENIX NSDI*, 2017.
- [46] Konstantinos Nikitopoulos, Georgios Georgis, Chathura Jayawardena, Daniil Chatzipanagiotis, and Rahim Tafazolli. Massively parallel tree search for high-dimensional sphere decoders. *IEEE TPDS*, 2018.
- [47] Chathura Jayawardena and Konstantinos Nikitopoulos. G-multisphere: Generalizing massively parallel detection for non-orthogonal signal transmissions. *IEEE Transactions on Communications*, 2019.
- [48] Minsung Kim, Salvatore Mandrà, Davide Venturelli, and Kyle Jamieson. Physics-inspired heuristics for soft mimo detection in 5g new radio and beyond. In *ACM MobiCom*, 2021.
- [49] Minsung Kim, Davide Venturelli, and Kyle Jamieson. Leveraging quantum annealing for large mimo processing in centralized radio access networks. In *ACM SIGCOMM*, 2019.
- [50] Minsung Kim, Davide Venturelli, John Kaewell, and Kyle Jamieson. Warm-started quantum sphere decoding via reverse annealing for massive iot connectivity. In *ACM MobiCom*, 2022.
- [51] Peng Ji, Hsin-Mu Tsai, Chao Wang, and Fuqiang Liu. Vehicular visible light communications with led taillight and rolling shutter camera. In *IEEE Vehicular Technology Conference*, 2014.
- [52] Weiping Sun, Jeongyeup Paek, and Sunghyun Choi. Cv-track: Leveraging carrier frequency offset variation for ble signal detection. In *ACM HowWireless*, 2017.
- [53] Qingrui Pan, Zhenlin An, Xuexuan Yang, Xiaopeng Zhao, and Lei Yang. Rf-dna: large-scale physical-layer identifications of rfids via dual natural attributes. In *ACM MobiCom*, 2022.