Scalable Acoustic IoT through Composable Distributed Beamforming Tags

Mohammad Rostami, Alan Liu, Karthikeyan Sundaresan
School of ECE, Georgia Institute of Technology

Abstract

With the proliferation of smart acoustic devices in everyday environments, low-power acoustic tags offer a promising choice for IoT applications. However, their highly limited operational range, throughput, and energy efficiency significantly restrict their viability for practical applications. In this work, we propose a first-of-its-kind distributed acoustic system called Disco. It consists of several low-power acoustic tags that can be flexibly composed on-demand to create an aperture array capable of distributed beamforming. The innovation of Disco lies in creating a 'virtual' distributed 2-speaker system that serves to wirelessly synchronize and enable distributed *temporal* beamforming at the tags independently. The low-power tags of Disco are prototyped with simple acoustic and analog-digital elements to create arrays comprising up to 8 tags. The beamforming performance of Disco scales with the number of tags in the array, delivering a multiple-fold increase in range, throughput, and energy efficiency. This advancement brings acoustic IoT applications closer to practical implementation.

1 Introduction

The last decade has witnessed substantial growth in wireless IoT applications in domains like smart homes, industrial automation, smart cities, and transportation, among others. This growth has spurred advances in low-power RF-based IoT solutions (e.g., RF backscatter tags [16, 31, 45, 47]) compatible with our commodity wireless devices utilizing WiFi, BLE and LoRa interfaces, enhancing the scope of their use. However, despite their interoperability advantage, RF tags' coexistence with other wireless devices in the environment has several challenges such as coexistence, low spectral efficiency, and deployment complexity (Section 2.1) that impede their practical deployment.

Acoustic interfaces for IoT: In recent years, voice has gained widespread adoption as a preferred user interaction modality [14, 29]. Coupled with advancements in natural language processing [6, 25], this has led to the integration of acoustic interfaces (microphones, speakers) in various everyday devices, including security systems [5, 17], voice assistants [33, 34], and smart appliances [20, 27]. In this work, we explore the potential of designing acoustic *tags* that can seamlessly operate with these commodity acoustic devices, providing a compelling option, besides RF, for IoT applications.

Acoustic tag designs and limitations: While acoustic tags are often used in underwater applications [7, 11, 49], where RF faces substantial attenuation, some works have examined their design for various over-the-air (OTA) applications [9, 18], that forms our interest. Existing acoustic tag designs face numerous challenges, limiting their practical use: (i) Limited bandwidth: Interoperability with commodity devices constrains usable bandwidth to 17-20 KHz to avoid overlap with the audible range of users (0.2-16 KHz [26]), leading to lower throughputs of less than 3 Kbps [23]. (ii) Limited range: Higher air attenuation restricts operation to just a few meters (e.g., 2m in [3]). (iii) Low energy efficiency: Transmitting a few Kbps over a few meters requires higher transmit power, causing low energy efficiencies (e.g. ≈ 1 mW for 3 Kbps over 2 m).

Figure 1: Disco use cases–Left: Industrial IoT. Right: Selective audio delivery in crowded spaces (e.g., airports).

Disco: To enable viable IoT applications with acoustic tags (e.g. industrial IoT, targeted audio delivery, etc. in Fig. 1), we present a novel distributed acoustic system, Disco. It includes several low-power acoustic tags that can be composed on-demand (as depicted in Fig. 2) to create an aperture array capable of distributed beamforming. This configuration increases operational bandwidth, range, and energy efficiency while ensuring transmissions stay inaudible to people. Using different numbers of tags in flexible configurations, Disco can adjust the SNR gain of its aperture to meet diverse application requirements like throughput, range, power, and constraints of the object or surface to which they are attached, without any deployment overhead.

To realize this vision in practice, Disco has to address several technical challenges, including: (i) *High variability of wavelength in the acoustic band:* Unlike RF, where wavelengths fall within a small interval due to high carrier frequency, they can vary significantly from 34 cm to 3.4 cm between 1 KHz and 10 KHz. This variation poses a challenge for phased-array system designs relying on half-wavelength element spacing. (ii) *Enabling tight OTA synchronization:* Use of extra (e.g., RF) interfaces or wires between tags for synchronization undermines the practicality and energy efficiency of their design. Hence, we aim to leverage the existing acoustic interface for synchronization as well. (iii) *Real-time data sharing and distributed beamforming:* Beyond clock synchronization, the data used for beamforming must be uniform and shared across tags using the same acoustic interface. Also, the tags must independently decide how to participate in transmissions to execute beamforming. (iv) *Maintaining inaudibility outside of the desired receive direction:* The total signal power outside of the main beamforming lobe directed at the receiver should be below the audible threshold for human ears (roughly 20 dB above a silent room's noise level) across the entire operation spectrum.

Design: DISCO is built on the novel primitive of transforming the exciter (commodity) device and one of the tags with the transmit data (a.k.a. master) into a 'virtual' distributed 2-speaker aperture, where the master tag doubles as a relay to retransmit the exciter

979-8-3503-6201-5/24/$31.00 ©2024 IEEE
DOI 10.1109/IPSN61024.2024.00008

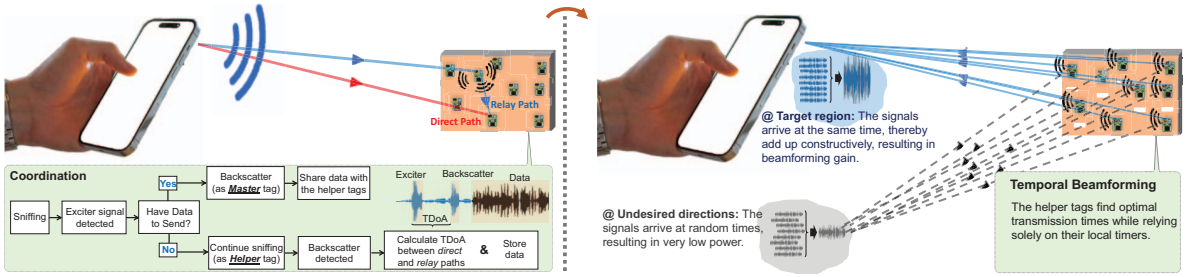

Figure 2: Disco Overview – Left: OTA synchronization triggered by the smart device. Right: Temporal beamforming.

signal. The time difference of arrival (TDoA) between the aperture's signals, enables 'helper' tags to independently achieve over-the-air synchronization, as well as time their participation in a coherent distibuted beamforming transmission as part of a larger array (Fig. 2) towards the exciter, all without explicit knowledge of the latter's location. To turn this primitive into a practical system, Disco incorporates four key design elements:

(i) *Relaying-aided synchronization:* A narrowband signal from the exciter is relayed by the master tag to synchronize the other tags in the aperture. Unlike the unknown distances from each tag to the exciter, the known separation between the master and helper tags aids synchronization to a common time reference by using the relayed signal.

(ii) *Temporal beamforming:* The master tag relays the exciter signal, enabling helper tags to measure the TDoA between the two signals, thereby determining the angle of departure (AoD) and relative phase offset with the master tag. Leveraging the lower time resolution of acoustic signals and TDoA, Disco achieves accurate time synchronization and employs time-based beamforming. Here, the distributed tags strategically delay their transmissions relative to the master tag for coherent reception at the commodity device. Temporal beamforming overcomes challenges of varying wavelengths innate to phased array beamforming.

(iii) *Data sharing through relaying:* The master tag appends its wideband data onto the relayed exciter signal, allowing helper tags to recover the data using the exciter signal as a preamble. Helper tags incorporate appropriate relative delays to compensate for data reception latency and their respective TDoA and schedule their wideband transmissions, over the entire acoustic band, at higher data rates, aligning at the exciter device.

(iv) *Spatially-selective signal whitening:* Disco utilizes beamforming to maintain inaudibility by operating at lower transmit power and minimizing energy propagation in undesired directions. Additionally, it employs signal whitening, encoding the data signal with a predetermined sequence at each tag to resemble independent white noise sequences outside the receiver. This results in an N-fold signal-to-noise ratio (SNR) gain and an $\sqrt{N}$ range gain at the device (from an N-tag aperture), along with a 10-fold interference-to-noise ratio (INR) drop elsewhere.

We built Disco using a composable, flexible array of low-power acoustic tags employing simple acoustic elements and analog-digital circuitry. Our evaluations show that Disco can enable distributed beamforming to smartphones, providing a scalable gain (with increasing tags) resulting in data rates as high as 34 Kbps (7× gain), coverage as high as 9–18 m (5× gain), and transmit power as low as 9 μW (65× reduction) for inaudibility, using a linear array of 8 acoustic tags compared to an equivalent phased-array system. Our contributions in this work are manifold:

- We propose Disco based on a novel primitive of composable, low-power acoustic tags that can wirelessly synchronize to create a first-of-its-kind time-based distributed acoustic beamforming system.
- Disco's design incorporates elements of acoustic data distribution and spatially-selective signal whitening to deliver a practical system for commodity devices while maintaining user experience through inaudibility in undesired directions.
- Disco's flexible distributed apertures yield significant gains in range, throughput, and power consumption to enable practical IoT applications.

We believe that Disco, with its practical design and resulting benefits, opens the door for several industrial and consumer IoT applications (Fig. 1) enabled by acoustic interfaces in our commodity devices (§7 for further discussions).

2 Prior Art and Motivation

2.1 Low-power RF networking

The last decade has witnessed significant progress in achieving ultra-low power and reliable communication for IoT devices using RF signals [16, 19, 22, 47]. Backscatter research [4, 30, 37, 44, 48] has played a crucial role in enabling communication with minimal very low energy footprint. Commodity backscatter [16, 31] is an important approach within this category, utilizing existing infrastructure and commodity devices to passively transfer data bits from wireless tags over the same RF spectrum used for existing communication protocols (e.g., Bluetooth, WiFi, LoRA, cellular, etc.).

2.1.1 ***Limitations:*** Despite the promising advances, these RF-based low-power solutions face several important obstacles in practical deployment.

- *Coexistence:* The need for RF tags to share the already-crowded, unlicensed ISM bands used by WiFi and Bluetooth devices makes it highly challenging to enable reliable communication from these tags in the presence of stronger, high-traffic devices pervasively available in the environment.
- *RF spectral efficiency:* Backscatter tags utilize the entire RF channel (e.g., 20 MHz WiFi channel) but have limited capacity, transmitting only tens to hundreds of Kbps of data. This hugely impacts the spectral efficiency of the RF channel, that is otherwise used by conventional devices at much higher bit rates. For instance, if 10 RF tags transmit at 100 Kbps each, it would take a total of 10 ms to transmit just 100 bits per tag, making it difficult to scale the application to multiple tags.
- *Deployment configurations:* Commodity backscatter designs require the use of two separate commodity devices (owing to their

half-duplex operation): one for transmitting the carrier signal to the tags, and another for receiving data from the tags, potentially on a different frequency to avoid interference. This makes for a challenging deployment in practice.

2.2 Acoustic sensing and communication

In contrast to the aforementionied RF limitations, acoustic interfaces provide a promising alternative for IoT applications due to their numerous advantages. The acoustic spectrum is orthogonal to and less congested than the RF spectrum. Moreover, common acoustic devices (e.g., smart devices) are already equipped with a speaker-microphone pair, facilitating full-duplex operation with acoustic tags. Further, the growing popularity of smart acoustic devices in our everyday environments has spurred innovative research in acoustic sensing and communication, which we discuss below. Smart devices have been used to illuminate targets with acoustic signals and infer their properties based on the observed reflections. For example, [35, 43] employ a smartphone's microphone to monitor human lung function and breathing. [1] enables spatial sensing for miniaturized robots using a single speaker-microphone pair, while [10] enables speech reconstruction from inaudible throat gestures.

More recently, acoustic interfaces for communication in challenging mediums have gained attention. For instance, [7] enables underwater transmission and reception of acoustic signals between smart devices at distances up to 100m. [11] minimizes tag power consumption by designing a meta-material structure for backscattering acoustic signals at sea depths. [28] enables acoustic backscatter communication using a metal surface as a waveguide.

2.2.1 ***Limitations of OTA Acoustic Solutions:*** Several recent works [2, 3, 13, 21] have tried to utilize acoustic signals for over-the-air (OTA) communication. However, these designs face challenges impacting their viability in communicating directly with smart devices for IoT applications.

Limited bandwidth while maintaining inaudibility: To ensure compatibility with consumer devices, acoustic tags must operate within the 0-20 KHz band, aligning with the device's speaker and microphone frequency response. However, frequencies below 17 KHz are audible to 99.9% of individuals, restricting the tag's use of this range to avoid interference, despite its significant portion of the spectrum. Existing approaches either operate at very low power across the entire bandwidth [3], leading to limited ranges (<2m), or rely on custom-built ultrasonic transceivers to exploit the microphone's inaudible spectrum for high throughput [24]. In contrast, Disco aims to enable inaudible communication with smartphones by effectively utilizing the entire spectrum to achieve higher bit rates without compromising range.

Limited range: To maintain inaudibility within a few centimeters of the tag, transmit power must be below 10μW (-20 dBm), corresponding to a 70 dB SPL, much lower than the recommended 85 dB SPL in earbuds. Figure 3 displays the measured acoustic SNR when the tag transmits a single-tone signal at this power level, revealing an achievable operational range of only 0.8 m for an SNR >10 dB at data rates in the tens of Kbps across all frequencies in the band. Previous methods aimed to extend range (25m) while maintaining inaudibility via chirp spread spectrum (CSS) over the inaudible portion of the audio band (16kHz-20kHz). However, this

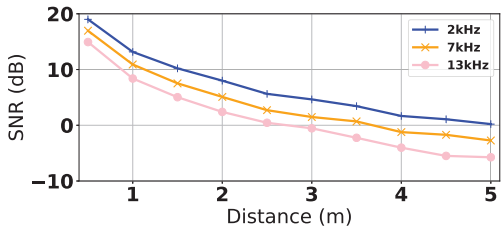

Figure 3: SNR vs distance for a single acoustic tag.

severely limits throughput to a few bps over the entire 20 KHz band, constraining practical applications.

Low energy efficiency: Current designs suffer from poor bandwidth efficiency and limited range, significantly impacting energy efficiency. Even with relaxed requirements for inaudibility, tags consume 200–1000 μW transmit power to achieve practical ranges of a few meters with throughputs of 10–20 Kbps, resulting in 10–100 bits/μJ. Reducing transmit power to 10μW limits throughput to <100 bps at practical distances, yielding approximately 10 bits/μJ. In both cases, energy efficiency is 1-2 orders of magnitude lower compared to existing ultra-low power RF counterparts achieving >1000 bits/μJ [19, 47].

3 Disco: Distributed Acoustic Beamforming

We propose Disco, a distributed acoustic beamforming system, whose significant benefits to operational range, throughput and energy efficiency, are central to enabling practical acoustic communication between low-power tags and consumer electronic devices (such as smartphones), while also simultaneously ensuring minimal to no interference to the surrounding environment (Fig 2).

Disco utilizes distributed *helper* tags around a *master* tag to transmit data to a smart device. Helper tags bolster the master tag's signal through distributed beamforming without requiring explicit knowledge of the smart device's location or wired connections. Each helper tag independently listens to the master tag's acoustic signal and relays it to facilitate constructive signal addition, enhancing signal strength at the smart device. This design ensures Disco's practicality and minimal deployment overhead in smart IoT applications, enabling easy addition of tags for extended range.

Disco goes beyond merely introducing acoustic array beamforming. As outlined in §3.1, enhancing range through beamforming presents challenges compared to RF phased-array systems. The diverse wavelengths in acoustics render conventional phased arrays impractical for consumer devices utilizing the entire audio bandwidth (3.1). Disco also enables location-based beamforming, more practical for near-field receivers, optimizing communication angles. Another key aspect is Disco's distributed approach, creating a virtual 2-speaker aperture for implicit localization. Tags use this in a completely distributed manner for wideband temporal beamforming through independent transmissions. While distributed acoustic tags have been demonstrated [15], achieving this in practical wireless communication solely based on acoustic signals, in a scalable and plug-n-play manner, constitutes the primary contribution of this work.

3.1 Practical Challenges

Realizing Disco's vision, i.e., distributed acoustic beamforming with low power tags, faces several technical challenges in practice.

Challenge 1: Designing a single aperture for varied wavelengths. In RF communication, the data signal's entire bandwidth BW is up-converted using a carrier frequency $f_c >> BW$. This results in RF signal frequencies occupying the range $(f_c - BW/2, f_c + BW/2)$, with corresponding wavelengths in a small interval $(\frac{C}{f_c+BW/2}, \frac{C}{f_c-BW/2})$ where C is the speed of light. A phased-array with elements spaced $\frac{\lambda}{2}$ $(= \frac{C}{2f_c})$ apart works well here. However, for acoustic signals transmitted at original frequencies (0, 20kHz), wavelengths vary significantly (e.g., 34 cm at 1 kHz vs 3.4 cm at 10 kHz). Thus, a single phased-array is ineffective across all frequencies and cannot utilize the entire bandwidth.

Challenge 2: Enabling tight over-the-air synchronization. Accurate beamforming needs perfectly synchronized tag clocks. Misalignments between the sent waveforms result in sub-optimal beamforming or an non-decodable aggregate signal. Also, wiring the tags or employing an alternate interface (e.g., RF) for synchronization is not practical. Thereby, it is desirable, yet challenging, to use the acoustic interface for both synchronization and communication.

Challenge 3: Sharing data and enabling beamforming from the distributed array. The helper tags don't possess the master tag's transmission data in advance. Without alternate interfaces, the same acoustic interface used for beamforming and clock synchronization must be leveraged for data sharing. Further, each tag needs to independently determine how to participate in the joint transmission to deliver the desired beamforming effect.

Challenge 4: Ensuring the signal is inaudible but decodable at the receiver device. Lastly, we need to ensure that the array's beamforming gain allows data decoding at the smart device but operates at a low enough transmit power to remain inaudible for human ears at practical distances (> 30-40 cm) from the array, enabling coexistence in practical deployments.

4 Disco's Design

4.1 Overview

Disco's innovation lies in bringing together completely distributed acoustic tags on-demand, and making them jointly perform as a coherent distributed aperture to enable beamforming towards the commodity (exciter) device. In contrast to the conventional phase-based beamforming that is achieved with wire-synchronized arrays, Disco accomplishes its objective through *temporal beamforming*, whereby each of the tags distributively (over wireless) identify their appropriate time of transmission, accounting for both the direction (AoD) to the exciter device and their relative path lengths, such that all the tags' transmissions arrive at the exciter device coherently to deliver beamforming gains.

An overview of Disco's design is captured in Fig. 2. Towards overcoming the hard challenge of distributed acoustic beamforming, Disco's novelty lies in transforming the exciter device along with one of the tags (a.k.a. master tag) into a virtual, distributed 2-speaker system through relaying. The latter is used to wirelessly synchronize, share transmission data, as well as provide the trigger needed for the other tags (a.k.a. helper tags) to locally estimate their differential ToA (time of arrival) from the 2-speaker system. The implicit estimation of the AoD to the exciter allows the tags to independently time their transmissions to enable distributed beamforming from the array.

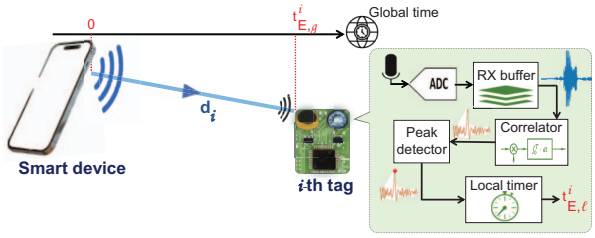

Figure 4: Detecting the ToA of the exciter signal.

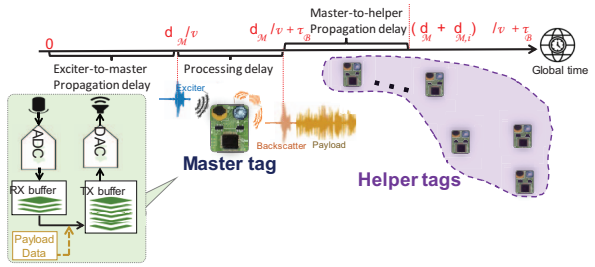

Figure 5: Enabling virtual dual speaker system via relaying.

Disco's design and operation involve the following key components: (i) *Enabling a virtual dual-speaker system:* The exciter device generates a narrowband exciter signal in the inaudible part of the spectrum, which, upon reception at the master tag, is relayed wirelessly. The relayed signal includes the narrowband signal (from the exciter) and the wideband data signal for beamforming by the array. Helper tags listen to this distributed 2-speaker system, recovering their local Time of Arrivals (ToAs) related to d_i and d_0 in Fig. 2, along with data for beamforming. (ii) *Relaying-aided synchronization:* Utilizing ToAs, each helper tag synchronizes its local timers (using the relay signal) and determines the Angle of Departure (AoD) to the exciter device from the Time Difference of Arrival (TDoA) between the two signals. (iii) *Distributed temporal beamforming and data scheduling:* Helper tags independently schedule their data transmissions in *time* based on synchronized timers for beamformed reception at the exciter, accounting for relative delays in data reception from the master tag and respective path delays to the exciter. (iv) *Data whitening:* To ensure inaudibility outside of the exciter, the data transmitted by the master tag is whitened prior to sharing, scrambled with a pre-determined sequence to resemble i.i.d white noise, utilized by all helper tags in their transmissions. Realizing this process in practice requires careful synergy between the exciter device, the master tag, and the helper tags and a distributed sequence of operations that we explain further in this section.

4.2 Distributed Virtual Two-Speaker Setup

The exciter device initiates distributed beamforming by sending an inaudible narrowband *exciter* signal as a preamble. All tags in the distributed array are in the *default* sniffing mode, continuously correlating received samples in their RX buffer with the known exciter waveform. A correlation peak indicates the arrival of the exciter signal (E) at the i-th tag, the timestamp of which is stored as $t^i_{E,l}$ (ToA of the exciter signal with respect to the local timer). If the i-th tag has no data to send, it transitions to the helper mode and continues sniffing to detect the ToA of the relayed signal. If all tags are in the helper mode (i.e., no data to send), they clear the value

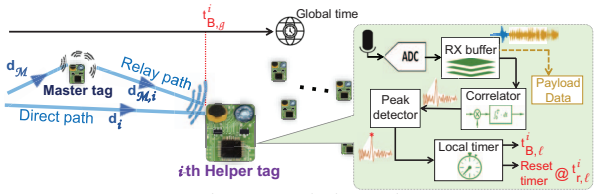

Figure 6: Relaying-aided synchronization.

of $t^i_{E,l}$ and return to the default sniffing mode, waiting for the next exciter signal. However, if a tag has data to send, it employs a simple random access (probability related to # tags in array) to decide when to become the *master* tag and relays the exciter signal, creating a virtual relay channel for the helper tags. With low-modest data rate requirements, this is sufficient to address the limited contention in the array.

The master tag introduces a fixed delay τ_B before relaying to ensure accurate detection of the ToA of both the direct exciter signal (E) and the relayed preamble (B) at the helper tags. The relayed preamble arrives at t^i_B after a delay of $\frac{d_M+d_{M,i}}{v}$ ($v \approx 340$m/s), always following the direct exciter signal arrival at t^i_E after a delay of $\frac{d_i}{v}$ (Figure 4), i.e., $\frac{d_M+d_{M,i}}{v} \geq \frac{d_i}{v}$ (due to the triangle inequality). Thus, $t^i_B \geq t^i_E$. To accurately detect the correlation peaks of both signals without overlap, the beginning of the B preamble must arrive after the end of the E preamble. Therefore, the master tag waits for $\tau_B = T$, the duration of the exciter preamble, before relaying. This ensures a minimum ToA difference equal to the duration of the exciter signal, allowing reliable detection at the helper tags through preamble correlation.

4.3 Relaying-Aided Synchronization

For effective temporal beamforming, synchronized local timers are crucial. Even a minor time offset can significantly diminish the beamforming gain at the receiver. Here, we show how Disco attains tight synchronization among distributed tags using only the relay signal from the master tag. This method allows for synchronization before each transmission, essentially resetting timer drifts between IoT transmissions.

Global vs. local time: To illustrate Disco's tag synchronization, we define global and local time notations. The moment the exciter signal is transmitted over the air from the speaker of the exciter device is regarded as the global time reference, i.e., the global time $t = 0$ (Figure 4). Each helper tag i in the distributed array has a unique local time offset δ^i relative to this global reference. Without synchronization, the δ^i values are arbitrary across tags. If we denote the local and global time at the i-th tag as t^i_ℓ and t^i_g respectively, we have $t^i_\ell = t^i_g + \delta^i$, where $t^i_g = t$ and is the same for all tags. However, these t^i_ℓ values cannot be directly used for temporal beamforming due to potential timing errors. The goal is to simultaneously reset all local timers, ensuring all tags have the same offset with respect to global time. Let $t^i_{r,g}$ denote the global time at which the local timer is reset to zero, i.e., $t^i_\ell = t - t^i_{r,g}$. The objective is for all tags to share the same $t^i_{r,g}$ value, resulting in simultaneous resetting of all t^i_ℓ values. Unlike a wired array, where timers can be shared or reset by a central unit, Disco achieves synchronization through a common over-the-air signal, using its Time of Arrival (ToA) as the timing reference for all tags.

Limitation of exciter signal as timing reference: Note that the ToA of the exciter signal at tag i is $t^i_{E,g} = \frac{d_i}{v}$, where d_i is the distance from the exciter device to the tag (Figure 4). If all the d_i's were known prior to synchronization, the tag's timer could be reset at $t^i_{r,g} = t^i_{E,g} + (C - \frac{d_i}{v})$, where C is a constant greater than the maximum $\frac{d_i}{v}$. In other words, tag i's local timer is reset at $C - \frac{d_i}{v}$ seconds after detecting the exciter signal. However, since the d_i values depend on the exciter device's location and are unknown to the tags, using the ToA of the exciter signal for resetting the timer is not viable.

Relay ToA as the timing reference: In contrast, Disco uses the relayed signal from the master tag as the timing reference (Figure 6). Unlike the direct arrival of the exciter signal at the helper tags, the relay signal undergoes three distinct delays before reaching the i-th tag (Figure 5) First, the propagation delay from the exciter device to the master tag is $\frac{d_M}{v}$. Second, master tag's delay τ_B, accounting for its processing of the exciter signal and subsequent relaying. Finally, the propagation delay from the master tag to the i-th helper tag, i.e., $\frac{d_{M,i}}{v}$, which is fixed based on the array configuration (known prior to communication) and ID of the master tag (from relayed signal). Disco is capable of self-calibrating, in the event the array configuration changes, as discussed in §4.4.1. Consequently, the ToA of the relay signal at the i-th tag is

$$t^i_{B,g} = \frac{d_M + d_{M,i}}{v} + \tau_B; \text{ where; } t^i_{B,\ell} = t^i_{B,g} + \delta^i \tag{1}$$

Thus, the i-th helper tag can reset its local timer $\frac{D-d_{M,i}}{v}$ seconds after detecting the relay signal, i.e.,

$$t^i_{r,\ell} = t^i_{B,\ell} + \frac{D - d_{M,i}}{v} = t^i_{r,g} + \delta^i \tag{2}$$

$$\implies t^i_{r,g} = t^i_{B,g} + \frac{D - d_{M,i}}{v} = \frac{d_M + D}{v} + \tau_B \tag{3}$$

which is the same across all helper tags. Note that D is a constant representing the maximum allowable distance between any two tags in the array. Thus, all the helper tags can reset their timers simultaneously, enabling distributed synchronization using local info and the relay signal's ToA.

Granularity of synchronization: While precise time synchronization is required for efficient temporal beamforming, the lower velocity of sound allows for a much coarser granularity of synchronization (order of μ seconds – three orders less than that needed for RF beamforming). This, in turn, can be easily achieved with low-power commodity clocks.

Robustness against local timer drifts: Disco synchronizes clocks per packet, initiated by the smart device's exciter signal. This preemptively compensates for drifts, averting significant synchronization errors. Tags share a common reference signal from the master tag, ensuring minimal timing differences post-synchronization, with errors smaller than audio sampling time.

4.4 Distributed Temporal Beamforming

With synchronized local timers, the helper tags must determine the appropriate time to transmit their data signals. To achieve temporal beamforming, all tag signals need to arrive at the exciter device's receiver simultaneously. Each helper tag relies on its local timer to initiate data transmission (D) at $t^i_{D,\ell}$. The global transmission time is adjusted by the timer reset, i.e., $t^i_{D,g} = t^i_{r,g} + t^i_{D,\ell}$. Consequently,

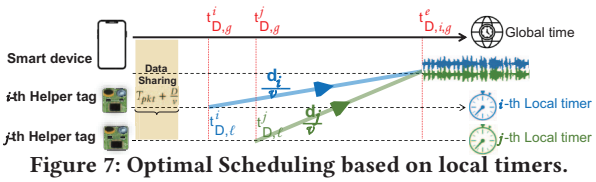

Figure 7: Optimal Scheduling based on local timers.

the Time of Arrival (ToA) of tag i's transmitted signal at the exciter e, denoted as $t^e_{D,i,g}$, depends on the transmission time and the air propagation delay (as shown in Fig 7):

$$t^e_{D,i,g} = t^i_{r,g} + t^i_{D,\ell} + \frac{d_i}{v} \tag{4}$$

Since $t^i_{r,g}$ is the same for all tags after synchronization (Section 4.3), ensuring alignment of signals at the exciter requires that $t^i_{D,\ell} + \frac{d_i}{v}$ is equal across all tags. In other words, the local transmission time must implicitly incorporate the fact that tags farther from the exciter in the array should start transmitting earlier than those closer to it, and vice versa.

Leveraging TDoA from virtual dual-speaker system for scheduling: While the direct use of the exciter signal's Time of Arrival (ToA) for timer synchronization is not feasible, Disco utilizes the virtual 2-element Distributed Speaker System formed by the exciter and master tag, along with the resulting Time Difference of Arrival (TDoA) between the exciter and relay signals at each tag, to distribute the transmission scheduling. Using Eqn. 1 and $t^i_{E,g} = \frac{d_i}{v}$, we can derive:

$$t^i_{B,\ell} - t^i_{E,\ell} = t^i_{B,g} - t^i_{E,g} = \frac{d_M + d_{M,i}}{v} + \tau_B - \frac{d_i}{v} \tag{5}$$

Since $t^i_{B,\ell} - t^i_{E,\ell}$ represents the difference in ToA between the exciter signal and the relay signals, each tag can independently determine and store this value even before resetting its local timer (Figure 7). Therefore, each tag i can set its transmission time as:

$$t^i_{D,\ell} = (t^i_{B,\ell} - t^i_{E,\ell}) + \frac{D - d_{M,i}}{v} \tag{6}$$

By substituting the values, we obtain:

$$\begin{aligned} t^i_{D,\ell} + \frac{d_i}{v} &= (t^i_{B,\ell} - t^i_{E,\ell}) + \frac{D - d_{M,i}}{v} + \frac{d_i}{v} \\ &= \left(\frac{d_M + d_{M,i}}{v} + \tau_B - \frac{d_i}{v}\right) + \frac{D - d_{M,i}}{v} + \frac{d_i}{v} \\ &= \frac{d_M + D}{v} + \tau_B \end{aligned}$$

$$\implies t^e_{D,i,g} = t^i_{r,g} + t^i_{D,\ell} + \frac{d_i}{v} = 2\left(\frac{d_M + D}{v} + \tau_B\right) \tag{7}$$

$t^e_{D,i,g}$ is identical for all helper tags, ensuring concurrent arrival of transmitted signals at the exciter device's receiver. To achieve this, the i-th helper tag plans its transmission at $t^i_{D,\ell}$ based on the local TDoA between the exciter and relay signals, considering its relative path length to the exciter in the array (Eqn. 6). This independent scheduling facilitates temporal beamforming at the exciter device.

4.4.1 ***Self-calibration to configuration change:*** As discussed earlier, knowing $d_{M,i}$'s is necessary for helper tags to both synchronize and find their optimal transmission times based on the virtual two-speaker aperture. Thus, if the array's physical configuration changes (e.g. when attached to a new object), the helper tags need to update their $d_{M,i}$. Disco employs a mapping F that maps tag IDs and array configuration to inter-tag distances. At a high level, for each helper tag i and master tag M,

$$d_{M,i} = F\{config, i, M, d\} \tag{8}$$

where $config$ represents the shape of the array (linear, rectangular, star, etc.) and d is the distance between two adjacent tags in the array. Note that, tags are assigned IDs based on their position in the array (e.g., tags are numbered from left to right in a linear array with $d_{M,i} = |M - i| \cdot d$).

When a user changes the array configuration, the smart device is used to send a calibration signal, containing the parameter $config$. Next, the tag with ID = 1 transmits an exciter signal (akin to that from the smart device), which is relayed back by the last tag N (N being the total number of tags in array) on reception. All the tags (except N) accurately estimate the TDoA between the exciter and relay signals, and leverage it along with the geometry of the array to estimate d and hence all desired $d_{M,i}$. Tags 1 and N can switch roles in a subsequent calibration slot to allow the N-th tag to estimate d as well.

4.5 Data Distribution + Scheduling

As the helper tags lack prior access to the data signal of the master tag, they need to obtain it before participating in temporal beamforming. Disco tackles this by having the master tag broadcast the wideband data signal immediately after its narrowband relay signal, as depicted in Figure 5. It takes no more than $T_{pkt} + \frac{d_{M,i}}{v}$ for the i-th helper tag to fully receive the data packet, where T_{pkt} is the data packet transmission time (at a fixed Modulation and Coding Scheme), and $\frac{d_{M,i}}{v}$ is the propagation time. By setting a guard time of $T_{pkt} + \frac{D}{v}$ at each tag, it ensures that the last tag in the array can fully receive and store the data signal. Thus, factoring in data sharing, the updated local transmission time at each tag is:

$$t^i_{D,\ell} = (t^i_{B,\ell} - t^i_{E,\ell}) + \frac{D - d_{M,i}}{v} + T_{pkt} + \frac{D}{v} \tag{9}$$

Communication overhead: Disco's intricate role assignments and payload sharing introduce complexities and communication overhead. Despite challenges, they enhance composability and meet diverse application requirements. Specifically, Disco accommodates dynamic sensor data transmission by integrating payload sharing. This enables any tag to embed its sensor, facilitating distributed temporal beamforming. While payload sharing introduces communication overhead (evaluated in Section 6.3), it fosters flexibility and adaptability, enabling seamless cooperation among tags to optimize transmission performance.

4.6 Preserving Inaudibility via Whitening

One of Disco's goals is to ensure that the acoustic signal is inaudible in most locations but decodable at the exciter device. Disco achieves this through a two-pronged approach:

(i) *Power scaling:* One key advantage of Disco is its ability to achieve practical operational ranges through beamforming from multiple tags, surpassing what a single tag can accomplish. If the transmit power of a single tag, P_m, is sufficiently low (e.g., -20 dBm), it remains inaudible beyond a short range (>30-40 cm). Disco scales the transmit power at each tag as P_m/N, where N is the number of tags in the array. This ensures that the total power outside the beamformed direction remains equivalent to that of a single tag, namely $N \cdot \frac{P_m}{N} = P_m$, which is below the audible threshold. Within the beamformed direction, the Signal-to-Noise Ratio (SNR) gain

extends the audible range by $\sqrt{N}$, which, for an 8-array, is still under 1 meter.
(ii) *Signal whitening:* While beamforming and the transmit beam pattern are directed at the exciter device, optimization of nulls in other directions is not performed. Consequently, signals from the N-tag array may occasionally constructively add outside the main lobe, resulting in an audible range > 30-40 cm. To mitigate this, Disco applies signal whitening, where transmitted data bits are scrambled with a predefined sequence. This process ensures modulated signal resembles independent and identically distributed (i.i.d.) white noise, exhibiting low correlation and statistical independence. In Disco, signals from different tags are essentially time-shifted versions of same signal, behaving like N independent noise signals outside main lobe. Thus, with signal whitening, power outside beamformed direction is limited to P_m, preserving inaudibility. Conversely, within beamformed direction, signals arrive simultaneously, resulting in aggregate signal that is sum of N identical signals. This aggregation leads to total peak power of $N^2 \cdot \frac{P_m}{N} = N \cdot P_m$, extending range by $\sqrt{N}$. For instance, using four inaudible tags at $P_m/4$ doubles range compared to single inaudible tag at P_m, while maintaining same total transmit power.

4.7 Interference Cancellation

Beamforming enhances SNR at the receiver in Disco, while also addressing occasional strong interference from environmental audio sources like speech and music. Tags append the exciter waveform to the beamforming signal, enabling the smart device to accurately detect TDoA (δt_S) of the beamformed signal ($S(t)$) at its two microphones, using correlation, and utilize it to estimate and compensate for interference. Let $R_1(t)$ and $R_2(t)$ be received signals at the two microphones: $R_1(t) = S(t) + I(t)$ and $R_2(t) = S(t - \delta t_S) + I(t - \delta t_I)$, where $S(t)$ and $I(t)$ are beamforming and interference signals respectively, with different arrival times due to varying angles. By time-shifting $R_2(t)$ and subtracting it from $R_1(t)$, we obtain: $R_1(t) - R_2(t + \delta t_S) = \cancel{S(t)} + I(t) - \cancel{S(t)} - I(t - \delta t_I + \delta t_S) = I(t) - I(t - \delta t_I + \delta t_S)$. The second term ($I(t - \delta t_I + \delta t_S)$) represents an echoed version of $I(t)$. We employ the echo cancellation pipeline proposed in [36] to filter out the second term. Consequently, the output of the cancellation pipeline, $\tilde{I}(t)$, is a close approximation of $I(t)$. Finally, $S(t)$ is recovered by subtracting $\tilde{I}(t)$ from $R_1(t)$.

5 Implementation

The key innovation of Disco lies in delivering the equivalent benefits of a large wired array, but with simple distributed elements. To achieve this, we have built Disco using a composable array of low-power acoustic tags that employ simple acoustic elements along with analog-digital circuitry.

5.1 Acoustic Tag

PCB prototype: Figure 8 shows the PCB prototype of Disco tags with the following key components:
Ultra low-power micro-controller (MCU): The MCU stores parameters, functions as the master or helper tag, and uses Texas Instruments MSP430FR5969 [40] due to its low energy profile. Onboard integration of the MCU is planned for the next prototype version.

Figure 8: Array configurations in different environments. A: Close-up. B: [Irregular, cafeteria]. C: [Rectangular, office]. D&E: [Linear & star, conference room].

Ultra low-power programmable logic (CPLD): The Xilinx XC2C64A [42] CPLD interfaces between the MCU and the audio codec, performs real-time correlation using a 4MHz EPSON's SG-8018CB crystal oscillator [32] as the baseband clock, and implements data whitening in the master tag.
Audio CODEC: Texas Instrument's ADS7887 [39] 10-bit ADC converts received audio signals to digital samples, while DAC101S101 [38] 10-bit DAC converts digital samples to analog for transmission. Both operate at 100 Ksps and communicate with the CPLD through SPI.
Microphone/Speaker: CUI Devices' CMA-4544PF [8] microphone captures audio signals, amplified by TI's OPA1671 [41] op-amp for improved ADC input. For transmission, a 32Ω speaker, coupled with a unity gain op-amp stage (OPA1671), ensures impedance compatibility with the DAC.

5.2 Exciter Device

Commodity phone as exciter device: We implement the exciter device on a Samsung Galaxy S21 phone without modifying its hardware. Our experiments are conducted using this model, but we do not rely on device-specific features that would restrict our applicability to other models. Other phones in the market with similar specifications (e.g., 24-bit ADC, 1W speaker) can also be used as an exciter device.
Phone Audio API: We utilize the *Dimowner* open-source project [12], providing an Android API for recording, playing, and processing raw audio samples. This API offers adjustable sampling rate and audio mode (mono/stereo), with a chosen rate of 96kHz for our purposes. It facilitates tasks including measuring signal-to-noise ratio (SNR) at phone's microphone, interference-to-noise ratio (INR) at unwanted locations, decoding data bits in received beamformed signal for calculating bit error rate (BER), and transmitting inaudible narrow-band exciter signal through phone's speaker.

6 Evaluation

We extensively evaluate various aspects of Disco's design and performance across three categories:
(i) **Validation**: We validate the effectiveness of Disco's temporal approach in distributed and efficient beamforming towards the exciter device.
(ii) **Performance**: We assess its overall performance gains and scalability compared to baseline approaches, including (a) phase-based beamforming with a wired array, which serves as an upper bound, (b) time-based beamforming with a wired array, serving as a direct upper bound for Disco, and (c) Disco with a single tag, which serves as a benchmark for conventional acoustic tag systems.
(iii) **Co-existence**: We examine its performance not only on the exciter but also its impact on audibility in undesired locations and directions.

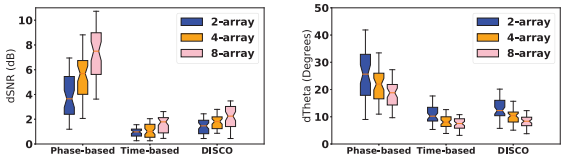

Figure 9: Beamforming Validation (d = 24 mm)

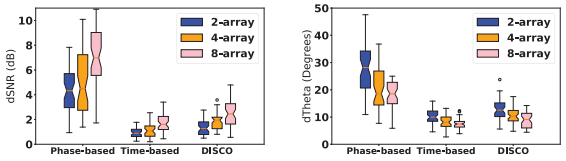

Figure 10: Beamforming Validation (d = 85 mm).

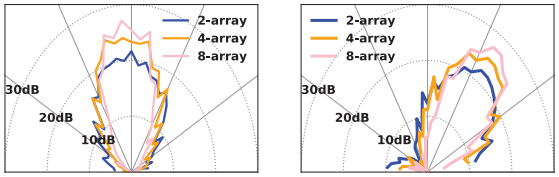

Figure 11: Beam patterns generated: 0° and 50° AoDs.

6.1 Experimental Setup

Our evaluations are conducted in diverse enterprise building locations, including open spaces (conference rooms, labs, and cafeterias) and cluttered areas (offices and corridors). Figure 8 illustrates array arrangement in different locations using various shapes (linear, star, etc.). Each experiment is averaged over multiple runs and locations for both tag array and receiver/exciter device (smartphone). Unless mentioned otherwise, we use linear arrays (Fig. 8-D) in our experiments across different environments to characterize various aspects of Disco's performance and validate design choices. Parameters studied include operation range, bandwidth, number of tags, array size (tag spacing), and transmit power. Metrics of interest are exciter's SNR, unwanted INR, BER, throughput, SIR, and energy efficiency.

6.2 Validation of Temporal Beamforming

Beamforming accuracy: We evaluate Disco's beamforming using two parameters: δ_θ and δ_{SNR}, signifying the angular and SNR difference between the ground truth AoD and the AoD at which Disco's beamforming lobe has maximum SNR. Figures 9 and 10 depict the measured δ_θ and δ_{SNR} of Disco against baseline schemes (Phased BF - wired, Timed BF - wired) for inter-tag distances d of 24 mm and 85 mm. Different numbers of tags (2, 4, 8) transmit a wideband (16 KHz bandwidth: [0.5, 16.5] KHz) signal to an exciter at 2 m. The box plots average over various exciter locations and angles. While the phased-array shows larger AoD deviations and resulting SNR drop (2-array: 28°, 4 dB; 8-array: 19°, 7 dB), Disco significantly minimizes the deviation in beamforming relative to the optimum (2-array: 10°, 1 dB; 8-array: 8°, 2 dB). Its performance nearly matches the upper bound on temporal beamforming provided by the wired array, demonstrating *Disco's relaying-aided virtual dual speaker system achieving accurate AoD identification and efficient temporal beamforming for wideband signals.*

Beam pattern and side lobes: Figure 11 demonstrates Disco's beamforming accuracy using anecdotal beam patterns for 2-array, 4-array, and 8-array configurations. These patterns are generated when beamforming a 16 KHz wideband signal to an exciter device placed 2 meters away at two ground truth AoDs: 0° and 50°. The SNRs at different AoDs are captured by a measurement device while the exciter device maintains a fixed AoD. Increasing the number of tags narrows the beam and boosts the SNR gain towards the ground truth AoD, closely aligning with Disco's peak SNR direction. Side lobe peaks stay at least 10 dB below the main lobe peak (e.g., 15 dB below for the 8-array), suggesting *no significant signal leakage in undesired directions.*

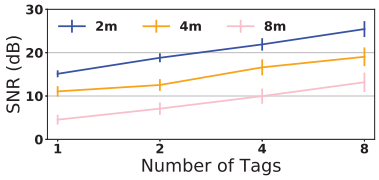

Figure 12: Scalable gain with larger arrays.

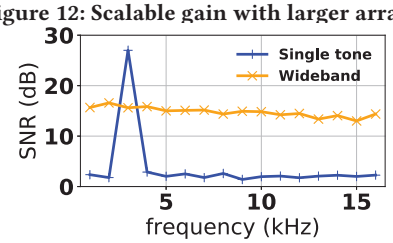

Figure 13: Spectral Efficiency.

Scalability: Next, we examine Disco's scalability in beamforming gain by varying the tag count in the array from 1 to 8. The SNR of the wideband beamforming signal is measured at various distances (2, 4, and 8 meters), with the measurements averaged over multiple AoDs for each distance. Figure 12 shows a steady 3 dB SNR boost with each tag count doubling. The 8-array displays the most beamforming gain, providing an 8–10 dB SNR enhancement compared to a single tag at various distances, thus validating *Disco's distributed temporal beamforming's scalability to larger array sizes.*

Wideband stability: We evaluate Disco's beamforming ability to uniformly enhance SNR across wideband spectrum. Figure 13 illustrates received SNR of an 8-array Disco in two scenarios: (1) narrowband transmit signal with 1 kHz bandwidth and (2) wideband transmit signal with 16 kHz bandwidth. Exciter device remains fixed at 1 meter from array. Key observations include: (i) Disco achieves total SNR gain of 27 dB for wideband signal, closely matching gain with single-tone signal (<1 dB loss); (ii) importantly, *Disco provides uniform SNR gain across entire frequency range when beamforming wideband signal.* This allows data-carrying subcarriers to fully utilize bandwidth, resulting in significantly higher throughput compared to phased-array systems. Utilizing simple binary modulation (e.g., OOK) with 1 ms symbol duration and sixteen 1 kHz subcarriers, Disco achieves throughput of 16 KBps (1 bps/Hz spectral efficiency).

Robustness to aperture size: To examine the impact of inter-tag separation d, we measure the received SNR from a 4-array and 8-array Disco with the exciter positioned 2 meters away, considering three aperture sizes: d = 13, 24, 85 mm. Figure 14 shows that *Disco's temporal beamforming approach remains resilient to changes in the aperture size*, exhibiting a dynamic SNR range of less than 1.9 dB and 2.1 dB for the 4-array and 8-array configurations, respectively. As a result, Disco adopts a default d of 24 mm for the array due to: (i) a higher SNR gain, and (ii) compact array size for easy deployment.

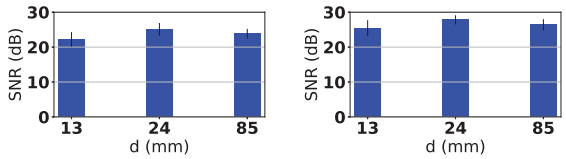

Figure 14: Robustness to aperture size (Left: 4-array; Right: 8-array).

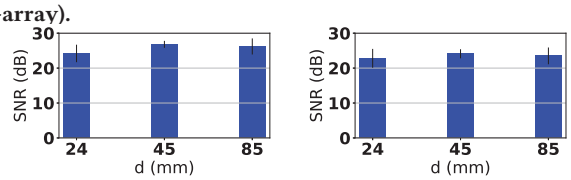

Figure 15: Robustness to shape (Left: □ , Right: ⋆).

Robustness to aperture shape: In this experiment, we evaluate Disco's performance and beamforming gain with different array configurations. We use two array setups: a rectangular 8-array (Figure 8–C) with a distance d between horizontally adjacent tags, and a star-shaped 8-array (Figure 8–E) with d equal to half the diagonal length of the star shape. Our results (Figure 15) demonstrate that Disco's temporal beamforming approach remains robust regardless of array shape, with a dynamic SNR range of under 2 dB for the rectangular array and 1.5 dB for the star-shaped array. These findings underscore the versatility and flexibility of Disco for deployment in various layout configurations.

6.3 Performance Gains

SNR: We evaluate received SNR at the exciter, positioned 2 meters from the linear array to receive the beamforming signal. Figure 16 shows results for two power management schemes in Disco: (1) power normalization for inaudibility, dividing total transmit power of -20 dBm among helper tags, and (2) tags transmitting at original fixed power of -20 dBm regardless of array size. These represent scenarios where tags are dynamically added to extend operational range without considering audibility. Beamforming gain increases by 3 dB when array size doubles with Disco's power normalization scheme. With power scaling, gain increases by steps of 6 dB, benefiting from power pooling. Upper bound scheme (time-based wired array) slightly outperforms Disco (1-2 dB) in both power management models. However, phase-based beamforming (wired) array lags, and gap with Disco widens as array size increases, exceeding *10 dB for an 8-array.*

Raw throughput: Figure 17 analyzes SNR gains for three subcarrier frequencies at different aperture sizes for an 8-array. Phased array sensitivity to frequency selection is high, with SNR varying over 10 dB within 11 kHz. Significant gain occurs only with aperture sizes aligning with half-wavelength. Phased array fails to enhance SNR for all subcarriers of wideband signals. In contrast, Disco's temporal beamforming is frequency-independent, resulting in minimal SNR variation for a given aperture size, enabling wideband operation and substantial throughput increase. With Disco currently using binary modulation (OOK), we estimate the expected throughput based on received SNR measurements using an SNR-MCS mapping table. The phased array's frequency sensitivity limits its spectrum utilization, resulting in a throughput of less than 5 Kbps. In contrast, Disco is estimated to achieve throughputs of 34 Kbps and 30 Kbps for aperture sizes of 24 mm and 85 mm, respectively, *surpassing the phase-based approach by 6–7 ×*. This enhancement can be achieved by employing adaptive, higher-order modulations on the wideband transmission subcarriers.

Component	Synchronization		Data sharing		Beamforming	
MCU	1.8	48.6	21.6	21.6	0.8	21.6
CPLD	216	126	41.4	41.4	2.3	48.6
CODEC	331.2	325.8	7.2	68.4	4.1	7.2
Total	549	500.4	70.2	131.4	7.2	77.4

Table 1: Disco's breakdown of power consumption (in μW).

Average throughput: Despite Disco's capability of up to 34Kbps with a linear 8-array, practical data rates are affected by substantial time allocated to synchronization and data sharing. To effectively monitor smart device movement, Disco repeats exciter-driven synchronization every 10ms, with 1ms for relay-aided synchronization, 1ms guard time, and the remaining 8ms for data sharing and temporal beamforming. Thus, Disco achieves an average throughput of 0.4×34Kbps = 13.6Kbps in real-world scenarios.

Range: Using an 8-array with 24 mm spacing, we evaluate bit error rate (BER) across various distances. Helper tags employ OOK modulation with power normalization on 16, 1 kHz subcarriers for beamforming. BER is computed over 16e5 bits (1e5 bits per subcarrier) transmitted at 1Kbps/kHz. Figure 18 depicts that phased array fails beyond 2.5 meters due to frequency-sensitivity, worse than a single tag. In contrast, Disco achieves BER below 1e-3 up to ≈ 9 meters, resulting in a *3.6 × increase in operational range.*

Moreover, Figure 19 illustrates Disco's range gain for an 8-array under both power management schemes. For a desired SNR of over 10 dB, *Disco extends coverage to 7.5 and 18 meters for the power-normalized and scaled models, respectively, compared to a range of just 1.5 meters with a single tag.* Thus, when inaudibility is important, Disco enhances the range by 5×, while this can be further increased to 12× in environments without humans.

Power efficiency of the audio front-end: To show Disco's advantages against a phased-array system, we examine the operational range of various schemes (for BER < 1e-3) at different transmit powers. Figure 20 reveals the baseline phased-array requires over 600 μW transmit power to achieve a 5 meter range with an 8-array for a 16 KHz signal. Conversely, Disco reaches the same range with about 70 μW using a 2-array, around 35 μW with a 4-array, and just ≈ 9 μW with an 8-array. Thus, *Disco consumes nearly 65× less power compared to a phased-array.* Including more tags notably improves the power efficiency of Disco's audio front-end.

Overall power profile: To assess DISCO's power consumption, we measure tag building blocks' power during relay-based synchronization, data sharing, and temporal beamforming. This includes MCU, CPLD, and audio CODEC (ADC + DAC + analog front-end), each consuming energy differently throughout the communication cycle. Table 1 details power consumption for master and helper tags at different stages. During relay-based synchronization, the master tag consumes 549μW, while each helper tag consumes 500.4μW. During data sharing, the master tag consumes 70.2μW, and the helper tag consumes 131.4μW. During temporal beamforming, the master tag consumes 7.2μW (quiescent power), and each helper tag consumes 77.4μW.

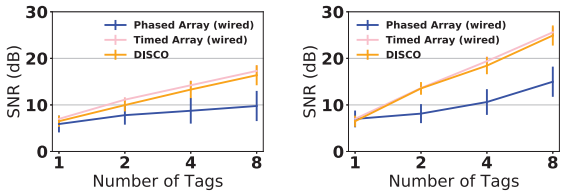

Figure 16: SNR gain (Left: normalized, Right: same power).

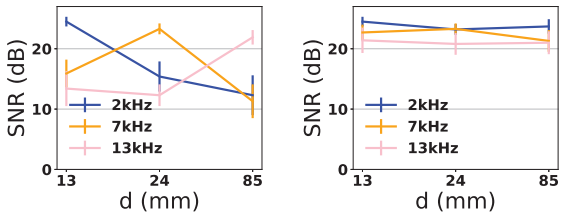

Figure 17: SNR justification. Left: Phased array; Right: Disco.

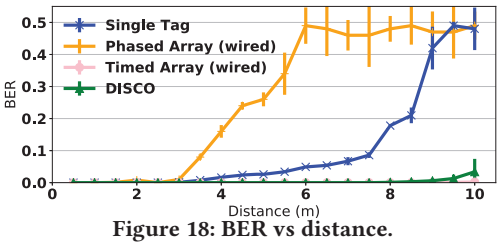

Figure 18: BER vs distance.

Energy Efficiency: For evaluating Disco's energy efficiency, we consider an 8-array with a 24mm aperture size conducting beamforming every 10ms to track smart device mobility reliably at 34Kbps. To ensure synchronization, data sharing, and temporal beamforming within this timeframe, we divide it into three slots: 1 ms for relay-based synchronization, 4ms (+1ms guard interval) for data sharing, and 1ms for temporal beamforming. In this setup, the master tag averages 85.9μW, while each helper tag averages 133.6μW. With an average throughput of 13.6Kbps, translating to 158.3bits/μJ and 101.8bits/μJ energy efficiency for master and helper tags, respectively, Disco surpasses BLE and ZigBee by 3×-5× and 8×-10× with a PCB prototype, making it an enticing low-power choice for IoT applications [46].

6.4 Coexistence

In this section, we evaluate the interference cancellation pipeline outlined in §4.7 to demonstrate Disco's capability to coexist with individuals and smart devices while preserving inaudibility during audible wideband communication. For distances of 1 meter from the array, if the acoustic SNR is below 20 dB, it is unlikely to be audible to humans unless they are very close to the array. We measure the audio SNR at different AoDs, both with and without signal whitening, when the array transmits a modulated wideband signal (0.5 kHz–16.5 kHz) to the exciter device 1 meter away (Figure 21: left: 4-array, right: 8-array). A dedicated measurement device records the received acoustic signal strength at various locations and angles, while the exciter device remains stationary.

Within the main lobe (±20° around the ground truth AoD), the SNR is high (20–30 dB), but it decreases outside due to Disco's beamforming. Before whitening, a few directions outside the main lobe still exhibit high SNR levels (from side lobes), potentially causing audible signals. Signal whitening effectively addresses this, maintaining the SNR below 20 dB throughout the ±20° region for the

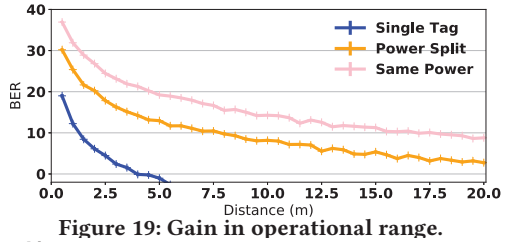

Figure 19: Gain in operational range.

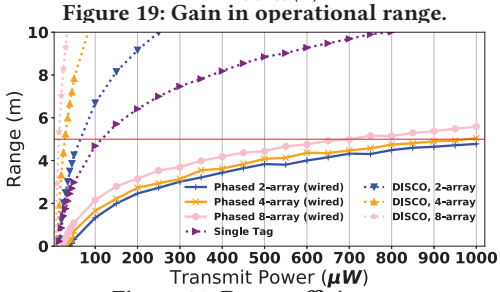

Figure 20: Power efficiency.

8-array. It also suppresses most SNR peaks in the 4-array case, demonstrating the effectiveness of signal whitening in ensuring statistical independence of tag signals outside the desired direction and limiting received power in those locations to that of a single tag (-20 dBm).

Audibility impact within the main lobe: Figure 22 reveals insights into audibility impact within beamformed direction and real-world observations of Disco. The received SNR heatmap shows no audible impact outside beamformed direction. Within it, only 12% of decodable locations within 2-3 m exhibit audibility (3% in entire room), showcasing Disco's commendable coexistence capabilities. In real-world scenarios (Figure 22), when a person obstructs direct path between array and smart device, audibility isn't perceived as exciter signal reception is obstructed, preventing beamforming. Green contours depict regions where individuals perceive beamforming signal (resembling white noise) without disconnection. Notably, only 3.5% of decodable locations within main lobe (0.9% across entire room) have audible impact on individuals. These findings underscore Disco's effectiveness in mitigating audibility concerns in real-world environments.

Robustness Against Ambient Sounds: Figure 23 illustrates Disco's resilience against ambient audio interference. We place a loudspeaker at various angles, playing a mix of sounds (speech, music, etc.) at a fixed distance (3 m) from the smart device communicating with a linear 8-array. We assess the signal to interference ratio (SIR) before and after applying self-interference cancellation to the receiver pipeline. Our tests show a significant improvement in SIR, transforming a non-decodable signal (-7 dB to -2 dB) to a decodable one (12 dB to 17 dB). This enhancement is observed in ≈ 90% of the tested angles. However, there's a small interval (≈ 20°) where the ambient audio source creates challenges due to $\delta t_I - \delta t_S$ being close to zero, degrading the performance of the echo elimination algorithm in [36] (see Section 4.7).

7 IoT Applications

Disco offers an effective solution for various industrial and consumer IoT applications, two examples of which are shown in Fig. 1: (i) monitoring in RF-challenged/prohibited (e.g. industry, certain health facilities) environments and (ii) targeted audio delivery in

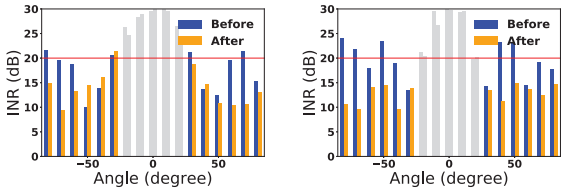

Figure 21: Co-Existence, Exciter AoD = 0°.

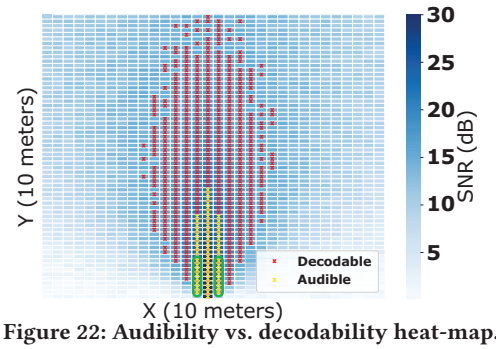

Figure 22: Audibility vs. decodability heat-map.

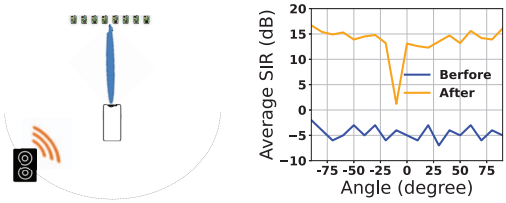

Figure 23: Robustness to interference. Left: Setup. Right: SIR before and after cancellation.

Solution	Latency	Deployment complexity
Disco	< 60 ms	Low
Duty-cycled active RF	> 1s	Low
Backscatter RF	10–100 ms	High

Table 2: Disco advantages for time-critical IoT.

noisy environments such as airports. In addition, Disco offers advantages in terms of low latency and low deployment complexity (Table 2) that make it an attractive choice for a number of time-critical applications. Our tests validate latency under 25 ms in 98% of cases, with the remaining 2% ranging from 25-60 ms due to multi-path.

Dynamic Adaptability with static array deployment: Disco's beamforming scheme requires a static distributed array for optimal performance. Despite this, it boasts swift self-calibration capabilities (§4.4.1), enabling adjustments to array structure and tag positions. This ensures effectiveness in dynamic environments without compromising practicality in industrial IoT. Moreover, Disco necessitates prior knowledge of array structure and parameters for all tags within the distributed array. This enables Disco's near-field temporal beamforming, aligning signals precisely in the time domain toward the receiver device's exact location, differing from traditional phased-array systems.

8 Discussions and Limitations

Enhanced use case scenarios with deployment flexibility: Disco's utility in industrial IoT, especially in smart warehousing, underscores its practicality and adaptability. It efficiently manages varying throughput requirements from a few Kbps to tens of Kbps crucial for continuous sensor data reading. Additionally, in warehouse environments with metal or liquid obstacles, acoustic communication's reliability over RF ensures dependable data transmission. Deploying numerous audio tags in close proximity allows for tailored communication ranges for different items, enhancing design flexibility. Disco's distributed nature enables coordinated communication among multiple tags on the same item, adaptable to diverse industrial settings. Furthermore, its potential extends to personal acoustic spaces, localizing audio to specific listeners while remaining inaudible elsewhere.

Role Assignments and Payload Sharing: Disco's flexibility in master tag assignment allows dynamic changes based on transmission needs. Multiple tags with sensors can alternate between master and helper roles, enhancing system versatility. However, simultaneous announcement of multiple master tags poses collision risks. Disco integrates collision avoidance mechanisms, with potential for further refinement. These contribute to system robustness and scalability in practical deployments, highlighting its potential for various industrial and consumer IoT applications.

Scaling aperture size: Increasing the number of tags (>10) in the array improves Disco's beamforming gains further but also introduces latency, medium access and data sharing overhead between master and helper tags, requiring optimized mechanisms for larger tag arrays.

Improving data rates: Disco currently uses binary modulation on subcarriers in the acoustic band. Investigating higher-order modulations in subsequent designs can enable higher throughput.

Proximity of users: Unlike observers, users holding smart devices within the audible region perceive the beamformed signal. To address this, a dynamic version of Disco can be employed. The master tag adjusts the transmit power of the array based on received exciter signal strength, reducing the audible region by allowing helper tags to transmit at a lower power without impacting decodability at the receiver.

Higher order modulations: The current version of Disco is limited to binary amplitude modulation (OOK). However, our upcoming iterations will introduce support for higher order modulations, which is expected to result in significantly higher data transfer rates, as outlined in §6.3. This improvement will make Disco a more practical choice, but it will come at the cost of a reduced operational range (compared to figure 18).

9 Conclusions

In this work, we proposed Disco, a novel distributed acoustic system that utilizes low-power acoustic tags to create an on-demand array for distributed beamforming. Disco achieves this by creating a virtual distributed 2-speaker system, enabling wireless synchronization and distributed *temporal* beamforming at the tags. The distributed nature of Disco allows for easy construction and deployment of tag arrays, resulting in superior performance in throughput, range, and energy efficiency. We consider Disco to be a groundbreaking acoustic system that paves the way for various low-power applications in RF-challenged environments.

References

[1] Y. Bai, N. Garg, and N. Roy. Spidr: Ultra-low-power acoustic spatial sensing for micro-robot navigation. In *Proceedings of the 20th Annual International Conference*

on Mobile Systems, Applications and Services, MobiSys '22, page 99–113, New York, NY, USA, 2022. Association for Computing Machinery.

[2] Y. Bai, J. Liu, L. Lu, Y. Yang, Y. Chen, and J. Yu. Batcomm: Enabling inaudible acoustic communication with high-throughput for mobile devices. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, SenSys '20, page 205–217, New York, NY, USA, 2020. Association for Computing Machinery.

[3] A. Bannis, H. Y. Noh, and P. Zhang. Bleep: Motor-enabled audio side-channel for constrained uavs. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[4] D. Bharadia, K. R. Joshi, M. Kotaru, and S. Katti. Backfi: High throughput wifi backscatter. *ACM SIGCOMM Computer Communication Review*, 45(4):283–296, 2015.

[5] M. Bhavani, R. Brinda, P. S. Manoharan, and S. Ramalingam. Online voice based smart security and automation system for real time application using artificial intelligence. In *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, pages 756–761, 2022.

[6] T. Chen. Facilitating physical-computer system design through data-driven natural-language interaction. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–6, New York, NY, USA, 2020. Association for Computing Machinery.

[7] T. Chen, J. Chan, and S. Gollakota. Underwater messaging using mobile devices. In *Proceedings of the ACM SIGCOMM 2022 Conference*, SIGCOMM '22, page 545–559, New York, NY, USA, 2022. Association for Computing Machinery.

[8] CUI Devices. *ELECTRET CONDENSER MICROPHONE*.

[9] D. A. Cuji, Z. Li, and M. Stojanovic. Act: an acoustic communications testbed. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–6, 2022.

[10] Y. Fu, S. Wang, L. Zhong, L. Chen, J. Ren, and Y. Zhang. Svoice: Enabling voice communication in silence via acoustic sensing on commodity devices. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, SenSys '22, page 622–636, New York, NY, USA, 2023. Association for Computing Machinery.

[11] R. Ghaffarivardavagh, S. S. Afzal, O. Rodriguez, and F. Adib. Ultra-wideband underwater backscatter via piezoelectric metamaterials. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication*, SIGCOMM '20, page 722–734, New York, NY, USA, 2020. Association for Computing Machinery.

[12] Audio recording android application. https://github.com/Dimowner/AudioRecorder.

[13] Z. Guoming, X. Ji, X. Zhou, D. Qi, and W. Xu. *UltraComm: High-Speed and Inaudible Acoustic Communication*, pages 184–204. Springer, 01 2020.

[14] A. L. Iñiguez Carrillo. Towards a human-computer interaction model for voice user interfaces. In *Proceedings of the IX Latin American Conference on Human Computer Interaction*, CLIHC '19, New York, NY, USA, 2020. Association for Computing Machinery.

[15] M. Itani, T. Chen, T. Yoshioka, and S. Gollakota. Creating speech zones with self-distributing acoustic swarms. *Nature Communications*, 14(1):5684, Sep 2023.

[16] V. Iyer, V. Talla, B. Kellogg, S. Gollakota, and J. Smith. Inter-technology backscatter: Towards internet connectivity for implanted devices. In *Proceedings of the 2016 conference on ACM SIGCOMM 2016 Conference*, pages 356–369. ACM, 2016.

[17] E. Jajuli, M. R. Effendi, L. Kamelia, R. Mardiati, D. Miharja, and E. A. Zaki Hamidi. The implementation of motorcycle security system using voice commands and fingerprint sensors. In *2021 15th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, pages 1–6, 2021.

[18] W. Jiang and W. M. D. Wright. Multi-channel indoor wireless data communication using high-k capacitive ultrasonic transducers in air. In *2013 IEEE International Ultrasonics Symposium (IUS)*, pages 1606–1609, 2013.

[19] B. Kellogg, V. Talla, S. Gollakota, and J. R. Smith. Passive wi-fi: Bringing low power to wi-fi transmissions. In *NSDI*, volume 16, pages 151–164, 2016.

[20] T. Kitagawa, K. Kondo, K. Nakazawa, and Y. Nakajima. Quality evaluation of voice synthesis tools for home appliance voice communications. In *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pages 26–27, 2021.

[21] H. Lee, T. H. Kim, J. W. Choi, and S. Choi. Chirp signal-based aerial acoustic communication for smart devices. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2407–2415, 2015.

[22] Y. Li, Z. Chi, X. Liu, and T. Zhu. Passive-zigbee: Enabling zigbee communication in iot networks with 1000x+ less power consumption. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, SenSys '18, page 159–171, New York, NY, USA, 2018. Association for Computing Machinery.

[23] Q. Liao, Y. Huang, Y. Huang, Y. Zhong, H. Jin, and K. Wu. Magear: Eavesdropping via audio recovery using magnetic side channel. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, MobiSys '22, page 371–383, New York, NY, USA, 2022. Association for Computing Machinery.

[24] Q. Lin, Z. An, and L. Yang. Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices. In *The 25th Annual International Conference on Mobile Computing and Networking*, MobiCom '19, New York, NY, USA, 2019. Association for Computing Machinery.

[25] H. Mun, H. Lee, S. Kim, and Y. Lee. A smart speaker performance measurement tool. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, SAC '20, page 755–762, New York, NY, USA, 2020. Association for Computing Machinery.

[26] The audible spectrum. https://www.ncbi.nlm.nih.gov/books/NBK10924/.

[27] M. Noor Shahida, B. Mardiana, H. Hazura, S. Fauziyah, M. Zahariah, and A. Hanim. Homes appliances controlled using speech recognition in wireless network environment. In *2009 International Conference on Computer Technology and Development*, volume 2, pages 285–288, 2009.

[28] P. Oppermann and C. Renner. Higher-order modulation for acoustic backscatter communication in metals. In *Proceedings of the ACM SIGCOMM 2022 Conference*, SIGCOMM '22, page 576–587, New York, NY, USA, 2022. Association for Computing Machinery.

[29] D. Pal, C. Arpnikanondt, S. Funilkul, and W. Chutimaskul. The adoption analysis of voice-based smart iot products. *IEEE Internet of Things Journal*, 7(11):10852–10867, 2020.

[30] Y. Peng, L. Shangguan, Y. Hu, Y. Qian, X. Lin, X. Chen, D. Fang, and K. Jamieson. Plora: A passive long-range data network from ambient lora transmissions. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '18, pages 147–160, New York, NY, USA, 2018. ACM.

[31] M. Rostami, K. Sundaresan, E. Chai, S. Rangarajan, and D. Ganesan. Redefining passive in backscattering with commodity devices. In *The 26th Annual International Conference on Mobile Computing and Networking*, 2020.

[32] SEIKO EPSO CORPORATION. *CRYSTAL OSCILLATOR (Programmable) OUTPUT: CMOS*. 2020.

[33] S. Sharma and G. Singh. Comparison of voice based virtual assistants fostering indian higher education – a technical perspective. In *2021 International Conference on Technological Advancements and Innovations (ICTAI)*, pages 162–167, 2021.

[34] R. Sivapriyan, N. Sakshi, and T. Vishnu Priya. Comparative analysis of smart voice assistants. In *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, pages 1–6, 2021.

[35] X. Song, B. Yang, G. Yang, R. Chen, E. Forno, W. Chen, and W. Gao. Spirosonic: Monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, MobiCom '20, New York, NY, USA, 2020. Association for Computing Machinery.

[36] F. Su, Y. Zhou, Y. Ma, and H. Liu. A robust acoustic echo canceller based on subband fast kalman algorithm and daul-filter structure. In *2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pages 1–5, 2021.

[37] V. Talla, M. Hessar, B. Kellogg, A. Najafi, J. R. Smith, and S. Gollakota. Lora backscatter: Enabling the vision of ubiquitous connectivity. *arXiv preprint arXiv:1705.05953*, 2017.

[38] Texas Instruments. *10-Bit Micro Power, RRO Digital-to-Analog Converter*.

[39] Texas Instruments. *ADS788x 10-Bit, 8-Bit, 1.25-MSPS, Micro-Power, Miniature SAR Analog-to-Digital Converters*.

[40] Texas Instruments. *MSP430FR596x, MSP430FR594x Mixed-Signal Microcontrollers*.

[41] Texas Instruments. *OPA1671 13-MHz, Low-Noise, Rail-to-Rail, Audio Operational Amplifier*. JANUARY 2019.

[42] Xilinx Inc. *XC2C64A CoolRunner-II CPLD*. Version 2.3.

[43] X. Xu, J. Yu, Y. Chen, Y. Zhu, L. Kong, and M. Li. Breathlistener: Fine-grained breathing monitoring in driving environments utilizing acoustic signals. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '19, page 54–66, New York, NY, USA, 2019. Association for Computing Machinery.

[44] G. Yang and Y. Liang. Backscatter communications over ambient ofdm signals: Transceiver design and performance analysis. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2016.

[45] P. ZHANG, D. Bharadia, K. Joshi, and S. Katti. Enabling backscatter communication among commodity wifi radios. In *Proceedings of the 2016 ACM SIGCOMM Conference*, SIGCOMM '16, pages 611–612, New York, NY, USA, 2016. ACM.

[46] P. Zhang, D. Bharadia, K. Joshi, and S. Katti. Enabling backscatter communication among commodity wifi radios. In *Proceedings of the 2016 conference on ACM SIGCOMM 2016 Conference*, pages 611–612. ACM, 2016.

[47] P. Zhang, D. Bharadia, K. R. Joshi, and S. Katti. Hitchhike: Practical backscatter using commodity wifi. In *SenSys*, pages 259–271, 2016.

[48] P. Zhang, C. Josephson, D. Bharadia, and S. Katti. Freerider: Backscatter communication using commodity radios. In *Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '17, 2017.

[49] Y. R. Zheng, X. Zhu, and M. Tan. Miniature underwater animal tags and smart sensors for civil engineering applications. In *Proceedings of the 14th International Conference on Underwater Networks & Systems*, WUWNet '19, New York, NY, USA, 2020. Association for Computing Machinery.