Satori: In-band Analog Backscatter for Audio Transmission

Yang Zou, Xin Na, Yimiao Sun, Yande Chen, Yuan He*
School of Software and BNRist, Tsinghua University
{zouy23,nx20,sym21,cyd22}@mails.tsinghua.edu.cn,heyuan@tsinghua.edu.cn

Abstract

In IoT applications such as environmental monitoring and industrial security surveillance, audio sensors are increasingly used, among which wireless sensors are preferred. In order to achieve a sustained transmission, low-power wireless technology such as backscatter has been widely considered. However, existing backscatter systems encounter difficulties in audio transmissions due to the high power consumption from the complicated digital processing and fast frequency-shifting clocks. In this paper, we propose Satori, the first-of-its-kind in-band analog backscatter system for audio transmission with ultra-low power consumption. Satori eliminates the need for in-place digital processing by directly embedding analog audio voltages into backscattered WiFi symbols through analog modulation. It also avoids the power consumption of the frequency-shifting clock by transmitting the audio within the excitation WiFi signal's band. We implement the Satori prototype and evaluate it under various settings. The results indicate that Satori can transmit audio at a sampling rate of 41.67 kHz and achieve a SNR exceeding 18 dB.

CCS Concepts

• **Networks** → **Network protocol design**; • **Hardware** → **Wireless integrated network sensors**.

Keywords

Backscatter, Analog, WiFi, Audio, RF Computing

ACM Reference Format:
Yang Zou, Xin Na, Yimiao Sun, Yande Chen, Yuan He. 2025. Satori: In-band Analog Backscatter for Audio Transmission. In *The 23rd Annual International Conference on Mobile Systems, Applications and Services (MobiSys '25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3711875.3729153

1 Introduction

Audio sensors are increasingly used in IoT applications, such as environmental monitoring, smart cities, and factory security surveillance [44, 48, 61]. Since audio signals are susceptible to interference and attenuation, these sensors are typically positioned near sound sources and rely on wired or wireless transmission methods. In industrial settings, where cabling is costly and impractical, wireless sensors are preferred [52]. Generally, these wireless sensors are expected to transmit high-quality audio while minimizing the power consumption.

Backscatter, as a low-power transmission technology[5, 14, 16, 23, 30, 66], has shown potential in achieving audio transmission with minimal energy consumption. Significant efforts have been made toward developing battery-free wireless sensors [25, 34, 39]. However, the existing backscatter systems fall short in realizing the vision of ultra-low-power audio transmission due to the high energy requirements from the following two parts:

• **High-speed digital processing**: Existing backscatter systems typically rely on the analog-to-digital converter (ADC) and processors to digitize and process the audio signal. In high-quality audio transmission tasks, these digital devices operate at a high speed, resulting in excessive power consumption burden on the backscatter tag.

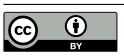
• **Fast frequency-shifting clock**: Since the interference from the in-band excitation will seriously degrade the audio transmission quality, existing approaches often shift the backscattered signal to a different frequency band [55, 66]. However, the frequency shift necessitates a fast clock, leading to additional power consumption.

In this paper, we introduce Satori, the first-of-its-kind in-band analog backscatter system for audio transmission. The Satori tag is excited by the WiFi signal commonly available in industrial environments. As shown in Fig. 1, it directly performs analog modulation and embeds the analog audio signals into the phases of each WiFi symbol, thereby avoiding high-speed digital processing. Satori then transmits the audio directly inside the frequency band of the excitation WiFi signal, eliminating the fast and power-consuming frequency-shifting clock. By leveraging analog modulation and in-band transmission, Satori enables audio transmission with ultra-low power consumption.

However, realizing the idea is non-trivial and presents several technical problems. First, the backscattered signal is often overwhelmed by the strong in-band excitation and is prone to noise. Second, even a minor carrier frequency offset (CFO) induces phase variations in the received signal, which can completely obscure the backscattered signal. Third, the frequency-selective fading in the Tx-Tag-Rx channel can severely weaken the backscattered signal, degrading the audio transmission quality.

To address these technical problems, we propose three key designs for Satori. At the tag end, we design the **in-band analog modulation** by exploring the in-band backscatter channel modulation model. This modulation method reliably embeds the audio signal into each WiFi symbol, even under strong interference from the in-band WiFi excitation. We further optimize the modulation's parameters to enhance the robustness against noise. We design the receiver to capture the baseband signal of each WiFi symbol and perform per-symbol channel estimation to extract the tag's modulation. Then, we propose the **two-step CFO calibration**. It begins by roughly cancelling the CFO through phase rotation of

*Yuan He is the corresponding author.



MobiSys '25, Anaheim, CA, USA

ACM ISBN 979-8-4007-1453-5/2025/06
https://doi.org/10.1145/3711875.3729153

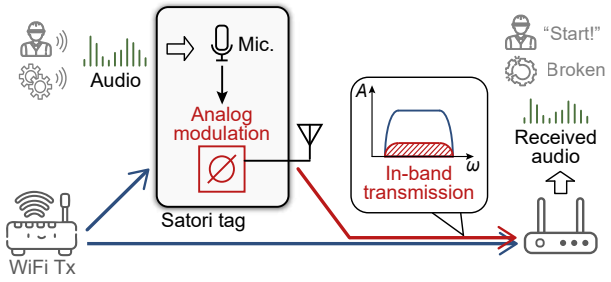


Figure 1: Satori utilizes analog modulation and in-band transmission to transmit audio signals.

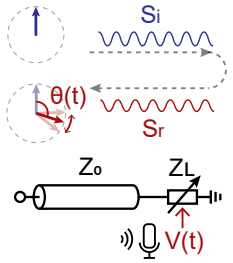
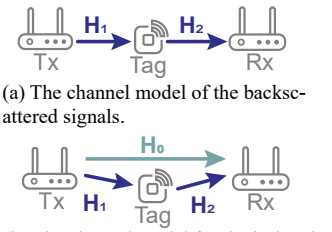


Figure 2: The reflected phase shifts with the analog voltage.



(a) The channel model of the backscattered signals.

(b) The channel model for the in-band backscatter.

Figure 3: The channel model of the backscatter communication.

the received signals. It further uses the cyclic prefix (CP) in every WiFi symbol to precisely track and calibrate the CFO symbol by symbol. Finally, we propose the **fading-adapted audio recovery**. It accurately quantifies the frequency-selective fading experienced by the backscattered signal from the received WiFi signals. Then, a weighted averaging technique is applied to mitigate the fading's impact on audio recovery.

We summarize our contributions as follows:

(1) We present the in-band analog backscatter system for the first time. It achieves audio transmission with ultra-low power consumption, pushing the development of audio sensing in industrial environments.
(2) We propose the design of analog in-band modulation, two-step CFO calibration, and fading-adapted audio recovery, addressing key technical challenges in the analog audio signal embedding and in-band transmission of backscattered signals.
(3) We implement the Satori tag design on PCB and the Satori receiver on a SDR, and conduct evaluations in an indoor environment. The results indicate that Satori can transmit audio with a SNR exceeding 18 dB.

The remainder of this paper is structured as follows: §2 and §3 introduce the design of Satori tag and receiver, followed by the implementation in §4. §5 evaluates the performance of Satori. Then we discuss in §6 and review the related works in §7. Finally, we conclude this paper in §8.

2 Tag Design

This section introduces how the Satori tag performs in-band analog modulation. We first explain how the Satori tag shifts the reflected phase according to the analog audio signal. To fully cancel out the strong interference brought by the in-band WiFi excitation, we propose the in-band analog modulation scheme and optimize its parameters for improving the robustness against noise. Finally, we present the design details of Satori tag's modulation circuit.

2.1 Shifting the Reflected Phase

The Satori tag first uses an ultra-low-power microphone to convert the audio signal to a voltage signal on the tag. This voltage signal is an analog signal that changes in real time with the external audio. We denote this analog voltage signal as $V(t)$, which carries the information of the external sound at this time. Then, we consider using this analog voltage signal to control the tag's reflected phase.

In an RF circuit, the reflected phase refers to the phase difference between the reflected signal S_r and the incident signal S_i, which is primarily determined by the terminal load impedance Z_L as:

$$\theta = \arg(\frac{S_r}{S_i}) = \arg(\frac{Z_L - Z_0}{Z_L + Z_0}), \quad (1)$$

where Z_0 is the characteristic impedance of the transmission line. By employing components with continuously tunable reactance as the terminal load, the reflected phase can be adjusted continuously. We directly feed the analog voltage $V(t)$ into the load, changing the load's reactance, and the reflected phase $\theta(V(t))$ is shifted by the voltage, as shown in Fig. 2. Thus, the tag can shift its reflected phase according to the analog audio signal in real time.

Generally, the reflected phase is expected to be linear with the voltage. This avoids distortions across different voltages and ensures ease of demodulation. Existing works [2, 34] propose a series of designs to make the reflected phase change linearly with the voltage.

2.2 In-band Analog Modulation Scheme

Modeling the Channel. The backscatter tag modulates the audio in the reflected phase. In essence, the tag's modulation alters the wireless channel between the transmitter (Tx) and the receiver (Rx). Considering the scenario shown in Fig. 3(a), a tag shifts the reflected phase linearly with the voltage, tuning the Tx-Tag-Rx channel as follows:

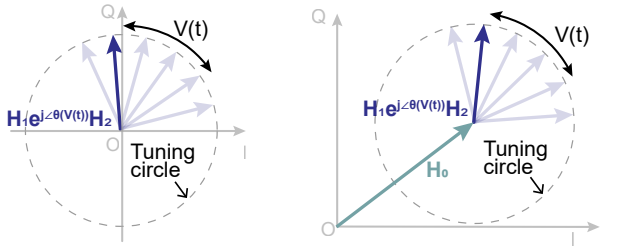
$$H = H_1 e^{j\theta(V(t))} H_2, \quad (2)$$

where H_1 is the channel from WiFi Tx to the tag, H_2 is the channel from the tag to the Rx. $\theta(V(t))$ represents the tag's reflected phase varying linearly with the voltage $V(t)$. We represent the channel H as a vector in the IQ plane, as shown in Fig. 4. When the voltage varies, the tag shifts its reflected phase, and the endpoint of the channel vector traces on a circle that is centered at the origin. This circle represents the backscatter tag's effect on the channel, which we refer to as the "tuning circle".

However, for in-band backscatter, the Rx receives both the excitation from the WiFi Tx and the backscattered signal from the tag, as shown in Fig. 3(b). In this case, the channel can be written as:

$$H = H_0 + H_1 e^{j\theta(V(t))} H_2, \quad (3)$$

where H_0 is the in-band excitation channel. The channel, as shown in Fig. 3(b), still forms a tuning circle on its endpoints. However,



(a) The channel is tuned by the tag and forms a tuning circle.

(b) The circle's center is displaced from the origin in the in-band backscatter.

Figure 4: The vector diagram of the channel. The tag tunes the channel, forming the tuning circle.

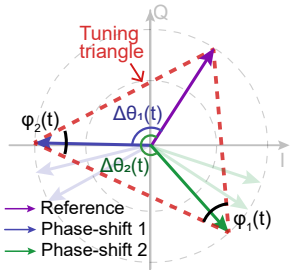


Figure 5: The modulation scheme and the tuning triangle.

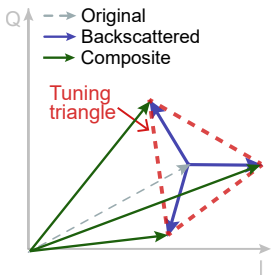


Figure 6: The three composite symbols form the tuning triangle.

due to the impact of H_0, the center is displaced from the origin. Therefore, to accurately extract the tag's tuning effect on the channel H, i.e., to extract the tuning circle, the tag should provide at least **three** points with distinct phases.

Designing the scheme. In order to modulate one analog voltage value in the in-band excitation interfered channel, our tag takes three backscattered WiFi symbols. The first is a **reference symbol**, with a fixed reflected phase, acting as a phase reference. The other two are **phase-shifted symbols**, whose reflected phases correspond to the current voltage value. Due to the wireless channels H_1 and H_2 introducing the same but unknown phase rotations to all three backscattered WiFi symbols, it is infeasible to represent the analog voltage using the phase of a single symbol. Instead, we focus on the phase differences between the phase-shifted symbols and the reference symbol, as shown below:

$$\Delta\theta_1(V(t)) = pV(t) + a_1,\ \Delta\theta_2(V(t)) = pV(t) + a_2, \tag{4}$$

where pV represents the part of the phase difference that changes linearly with the analog voltage V, and p is the slope. a_1 and a_2 are additional fixed phase differences applied by the tag, preventing the three symbols from overlapping at some specific voltages.

Fig. 5 illustrates the phase diagram of the three symbols. Furthermore, we find that these three backscattered WiFi symbols form a triangle. We refer to it as **tunning triangle**. The two internal angles $\varphi_1(t)$ and $\varphi_2(t)$ of the tunning triangle has the following relationship with $\Delta\theta_1(V(t))$ and $\Delta\theta_2(V(t))$:

$$\varphi_1(t) = \frac{1}{2}\Delta\theta_1(V(t)),\ \varphi_2(t) = \frac{1}{2}(2\pi - \Delta\theta_2(V(t))), \tag{5}$$

Thus, the analog voltage value is carried in the tuning triangle formed by the three symbols. The tag periodically modulates the three backscattered WiFi symbols, continuously embedding the varying analog audio voltage.

This modulation scheme in the in-band backscatter scenario is illustrated in Fig. 3(b). The three backscattered WiFi symbols add up with the original WiFi symbols, forming three composite WiFi symbols. As shown in Fig. 6, we find that the endpoints of the three composite WiFi symbols just form the tuning triangle on the IQ plane, indicating that the tuning triangle formed by the modulation can stably carry the voltage in in-band backscatter.

Optimizing the robustness against noise. To improve the modulation scheme's robustness to noise, we optimize the slope p and the fixed phase differences a_1 and a_2 mentioned in Eq. 4. Intuitively, the slope p should be set as large as possible, which means more significant phase shifting, making it more resistant to noise. Next, we design the fixed phase differences a_1 and a_2 by solving an optimization problem: we first generate N analog voltages $V[i]$ as possible inputs in the range $[0, V_0]$, where V_0 is the maximum possible voltage. Then, we simulate the tag's modulation process, add the Gaussian noise to the three received WiFi symbols in the IQ domain and extract the analog voltage $\hat{V}[i]$ from the triangle formed by the three WiFi symbols. We aim to minimize the error between the extracted analog voltage and the original analog voltage, i.e.:

$$\arg\min_{a_1,a_2} \sum_{i=1}^{N} (V[i] - \hat{V}[i])^2, \tag{6}$$

where a_1 and a_2 are ranging from $[0, 360)$.

We first consider the theoretical derivation of the optimal parameters. We treat the extracted $V[i]$ as a random variable, and Eq. 6 implies minimizing the central second-moment of $V[i]$. An ideal approach is to derive the central second-moment of the inner angles of the tuning triangle from the coordinates of WiFi symbols. Then, we use Eq. 5 to obtain the relationship between the central second-moment of $V[i]$ and a_1, a_2. However, this method is challenging because the calculation of the inner angles involves non-linear operations on multiple random variables, and the integral for solving its central second-moment has no explicit analytical solution.

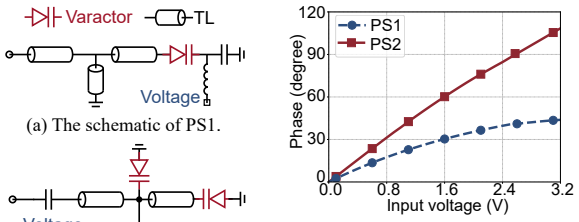
Therefore, we first apply the Monte Carlo method to compute the central second-order moment of $\hat{V}[i]$ (i.e., the mean square error as defined in Eq. 6) and utilize a heuristic simulated annealing algorithm to solve the minimization problem. Since heuristic algorithms may converge to local optimal solutions, we conduct multiple calculations with multiple initial values, initial temperatures, and annealing methods, and select the best result among the calculations. The optimized values of a_1 and a_2 are as follows:

$$a_1 = \frac{\pi}{2} - \frac{V_0}{2p},\ a_2 = \frac{3\pi}{2} - \frac{V_0}{2p}, \tag{7}$$

This means that the phase difference between the two phase-shifted symbols is always π.

2.3 Designing the Modulation Circuit

We first focus on how to modulate the reference symbols and phase-shifted symbols. For the reference symbol, a straightforward method is applied, involving using a fixed load, such as a capacitor.

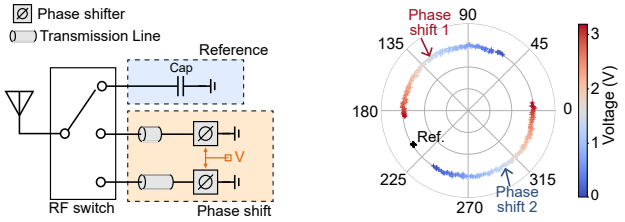


(a) The schematic of PS1.

(b) The schematic of PS2.

(c) The phase shift of PS1 and PS2 versus the input voltage.

Figure 7: Two kinds of reflective-type phase shifters and their performance comparison.



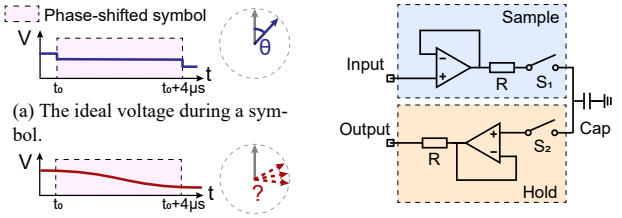
(a) The modulation circuit.

(b) The simulation results of the modulation circuit.

Figure 8: The schematic and the simulation results of our reflective-type phase shifter.

To modulate two phase-shifted symbols whose phases change linearly with the voltage, existing work proposes a reflective phase shifter (PS1) that consists of a varactor and three transmission lines (TL) as shown in Fig. 7(a). When the diode is biased with the analog voltage, its capacitance changes, altering the reflected phase. Although adjusting the TLs in PS1 can increase the phase-shifting range, it also introduces nonlinearity in the phase-voltage relationship. [2] replaces one TL in Fig. 7(a) with another varactor diode, and we find that using multiple varactors provides opportunities for nonlinear compensation. We propose the PS2, which introduces an additional parallel nonlinear varactor after performing impedance matching with a TL. This varactor is biased with the same voltage. By adjusting the characteristic impedance and length of the TL in PS2, the nonlinear phase shifting can be compensated by the second varactor. To achieve this, we use PathWave ADS to optimize the TLs in PS2 for phase-shifting range and linearity, satisfying the assumptions of Equ. 4. Fig. 7(c) shows that PS2 performs better than PS1 in terms of the phase-changing slope and linearity. We limit the maximum phase shift within π by adjusting the TLs in PS2, avoiding the phase wrapping ambiguity in the two phase-modulated symbols. A natural question is whether the shifter has the same phase response across the 20 MHz WiFi bandwidth. In fact, since the WiFi bandwidth is far smaller than the carrier frequency of 2.4 GHz, the shifter introduces almost identical phase changes across the entire band. We measure the reflected phase at different frequencies using a VNA. With a frequency difference of 20 MHz, the maximum phase deviation is only 0.49°, corresponding to an error of less than 5‰.

However, when connecting with analog microphones, the high-frequency audio signal will render the rapidly changing voltage. For a voltage signal with frequency f_v and amplitude A_v, its variation within a 4-μs WiFi symbol can reach $2A_v \sin(2\pi f_v \times 2\mu s)$. Assuming



(a) The ideal voltage during a symbol.

(b) The fast voltage change disrupt the phase-shifted symbol.

(c) The schematic of the sample-and-hold circuit.

Figure 9: The sample-and-hold circuit avoids the fast voltage changes disrupting the phase-shifted symbols.

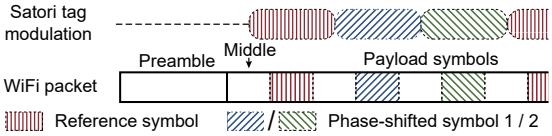


Figure 10: Satori tag's modulation mechanism

that the frequency of the voltage signal is 4 kHz, its variation in 4 μs is 10% of its amplitude. If such a rapidly changing voltage is directly fed into the phase shifter, the reflected phase can vary significantly within a single WiFi symbol, corrupting the symbol, as illustrated in Fig. 9(b). Moreover, rapid voltage changes can cause the two phase-shifted symbols to be embedded with different voltage values, resulting in incorrect voltage embedding based on the tuning triangle. To solve this problem, we propose a sample-and-hold circuit to stabilize the analog voltage during the two phase-shifted symbols, as shown in Fig. 9(c). This circuit consists of two voltage followers, two switches, and a capacitor. During the reference symbol, switch $S1$ closes and $S2$ opens, allowing the first follower to charge the capacitor. Then $S1$ opens, and $S2$ closes before the phase-shifted symbols, allowing the second follower to follow the voltage stored in the capacitor. This circuit effectively stabilizes the voltage signal within the two phase-shifted symbols, ensuring correct voltage embedding.

Quick demonstration. We propose the modulation circuit design as shown in Fig. 8(a). It consists of an RF switch and three loads. The first load is fixed to generate the reference, and the other two are phase shifters. The analog voltage is stabilized by the sample-and-hold circuit before input to phase shifters. By adjusting the transmission line in front of the two phase shifters, we ensure that the phase difference satisfies Eq. 7. We use PathWave Advanced Design System (ADS) to verify our circuit design. Fig. 8(b) shows the reflection coefficients of the three loads under different voltage inputs, indicating that our circuit realizes the optimized parameters.

Finally, the low-power tag can't detect and align the WiFi symbols perfectly and has a clock frequency offset of hundreds of ppms, both of which will cause the symbol synchronization error. Thus, we design the modulation mechanism of Satori tag, as shown in Fig. 10. When an excitation WiFi packet arrives, the tag skips the preamble and switches to the reference state in the middle of the first payload symbol to start modulating the reference symbol. Then, the circuit toggles the switch and modulates the two phase-shifted symbols. The switch period is twice of one WiFi symbol period, ensuring that every even-index payload symbol is fully modulated. Meanwhile, the sample-and-hold circuit works in conjunction with

this modulation mechanism, sampling in two WiFi symbol periods and holding the value for four WiFi symbol periods. This mechanism can tolerate the synchronization error of up to 2 μs (i.e., half of one WiFi symbol duration). The packet detection of backscatter tags can achieve a detection time offset below 1 μs [9, 10, 34]. And for a typical 1 ms WiFi packet, the cumulative time misalignment caused by a frequency offset from a clock with hundreds of ppms remains below 1 μs. Therefore, the proposed mechanism can minimize the synchronization error introduced by both effects.

Although this mechanism takes the time of six WiFi symbols to carry one voltage, it still achieves a sampling rate of 41.67 kHz with the WiFi symbol rate of 250 kHz, sufficient to support audio signal transmission. Regarding the transmission latency, a voltage signal is captured by a sample-and-hold circuit upon the arrival of the first WiFi symbol. Thus, the transmission delay is 6×4 μs = 24 μs. Although the latency is slightly greater than that of backscatters optimized for transmission delay, such as the microphone array backscatter [71] with a latency of 17.2 μs, the 24-μs latency is already sufficient for most practical applications.

3 Receiver Design

As shown in Fig. 11, when the tag shifts the reflected phase of an incident WiFi symbol, the same phase shift is applied to all the WiFi subcarriers of the backscattered symbol. Recall that after the phase shifting, every three WiFi symbols form a tuning triangle, which carries the voltage information. Since the tag modifies the reflection phase of each symbol in the channel to embed audio signals, the receiver needs to capture the baseband subcarriers of all WiFi symbols and perform per-symbol channel estimation to extract the tag's modulation. Specifically, we focus on the pilot subcarriers as they are fixed and known across all WiFi symbols.

Here are the ideal processing steps for the receiver to recover the audio: when the receiver receives a WiFi packet, it first processes it with standard WiFi demodulation. Then, it picks the even-index payload symbols, which are fully modulated by the tag and extracts their pilot subcarriers. Next, the receiver processes these pilot subcarriers in the IQ domain. It calculates the inner angles of the tuning triangle using the coordinates of the pilots from three symbols in the IQ domain. Finally, it computes the embedded voltage values from these inner angles by solving Eq. 5, thereby recovering the analog audio.

3.1 Two-steps CFO Calibration

Although the receiving seems straightforward, the frequency offset, including the *carrier frequency offset* (CFO) and the *sample frequency offset* (SFO), can affect the phase of the received signal. Since Satori doesn't introduce the frequency shifting or change the WiFi symbol rate, the offset is mainly between the transmitter and the receiver. Among the two offsets, the CFO has a greater impact on the phase of the pilots [28]. Since the backscattered signal is only a small part of the received signal in our in-band transmission, the phase rotation caused by the CFO completely overwhelms the triangle formed by the pilots of the received symbols, as shown in Fig. 12.

The standard WiFi receiver uses the short training symbol (STS) and long training symbol (LTS) for CFO calibration; there is still residual CFO (rCFO) in the payload [41]. Although the calibration of the rCFO has been studied in many existing works, these methods only work when the pilot subcarriers are not modulated by the tag [41, 42, 47] and can't be applied in our scenario.

Considering that the rCFO fluctuates slightly around an average value, we propose the two-step CFO calibration method. It first detects and calibrates the average rCFO and then tracks and corrects the fluctuating part.

First step: average CFO correction For the first step, we focus on the reference symbols to detect and correct the average rCFO. The pilot subcarriers of these WiFi symbols can be modelled as:

$$R_k = (H_0 X + H_1 X e^{j\theta_{ref}} H_2) e^{2\pi j \cdot \mathrm{rCFO} \cdot N_p k \tau}, \tag{8}$$

where R_k is the pilot subcarriers of the kth reference symbol. H_0, H_1, and H_2 represent the wireless channel as described in Fig. 3(b). X, θ_{ref}, and τ denote the pilot in the original WiFi symbol, the tag's phase change, and the duration of the WiFi symbol, respectively. And N_p is a constant value of 6, as the interval between every two reference symbols is 6 WiFi symbols.

Since X and θ_{ref} are fixed, and H_0, H_1, and H_2 are stable within a WiFi packet, there is a linear relationship between the phase of R_i and rCFO as follows:

$$\angle R_k = \angle (H_0 X + H_1 X e^{j\theta_{ref}} H_2) + 2\pi \cdot \mathrm{rCFO} \cdot N_p k \tau, \tag{9}$$

where $\angle (H_0 X + H_1 X e^{j\theta_{ref}} H_2)$ is a constant phase.

We use the linear fitting method to estimate the average rCFO and complete the first step of CFO calibration. As shown in Fig. 12(b), after the first step, the received pilot subcarriers have formed some clusters, but they still cannot form a clear tuning triangle for the voltage recovery. This is because the fluctuating part of the rCFO cannot be finely calibrated in the first step.

Second step: fluctuating CFO calibration The second CFO calibration step uses the CP within each symbol to track and calibrate the fluctuating part of the rCFO. For WiFi symbols sampled at 20 MHz, the CP are the first 16 samples, which are the same as the last 16 samples. This allows us to detect the rCFO level as follows:

$$\hat{\mathrm{rCFO}}_n = \frac{N_s}{F_s} \angle \Big(\sum_{t=1}^{16} (x_n^*[t] x_n[t + N_s])\Big), \tag{10}$$

where $\hat{\mathrm{rCFO}}_n$ is and $x_n[t]$ is the the rCFO and the tth sample in the nth WiFi symbol. And N_s represents the number of samples in the interval between the CP and the last 16 samples, which is 64 in WiFi OFDM. F_s is the sampling rate, which is 2×10^7.

We only correct the CFO of the even-index WiFi symbols in the payload section, which are fully modulated by the tag. To further improve the estimation accuracy, we accumulate and update the $rCFO_k$ as follows:

$$\mathrm{rCFO}_n = \begin{cases} \hat{\mathrm{rCFO}}_n & n = 2 \\ (1-\rho) \cdot \mathrm{rCFO}_{n-2} + \rho \cdot \hat{\mathrm{rCFO}}_n & n > 2 \end{cases}, \tag{11}$$

where ρ is the update rate. We apply the rCFO_n to calibrate the nth WiFi symbol, yielding the results shown in Fig. 12(c). The two-step CFO calibration calibrates the rCFO, and the pilot points form a clear tuning triangle, facilitating a better audio recovery.

One may concern that the SFO between the transmitter and the receiver could also impact audio recovery. However, for a certain pilot subcarrier, the phase rotation caused by SFO is also linearly

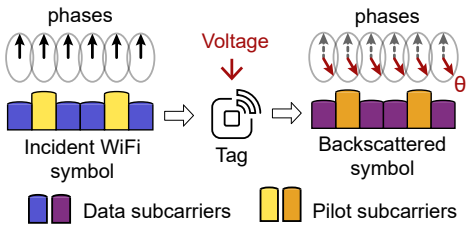
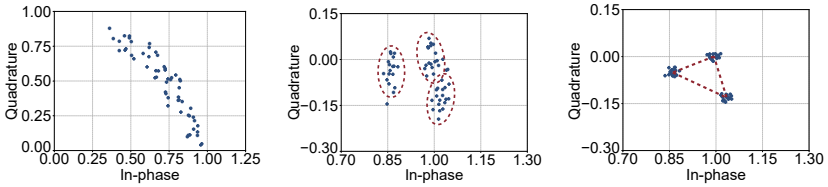


Figure 11: The tag shifts the phases of all subcarriers of the backscattered symbol.



(a) Before the CFO calibration. (b) The 1st step calibration. (c) The 2nd step calibration.

Figure 12: The IQ points of the pilot subcarriers before and after the two-step CFO calibration.

related to the symbol index [41]. Thus, the linear fitting in the first step actually also eliminates the effects of SFO.

3.2 Fading-adaptive Audio Recovery

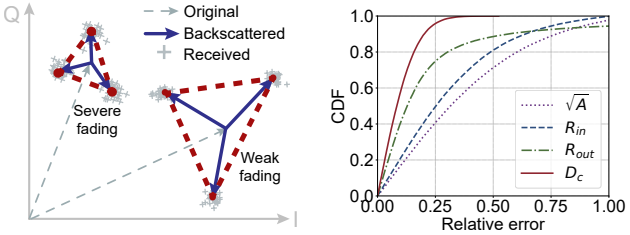
After the calibration, the receiver can extract the tuning triangle from each pilot subcarrier to recover the tag's analog voltage. However, in real-world environments, the frequency-selective channel fading in the Tx-Tag-Rx channel can significantly weaken some pilot subcarriers of the backscattered WiFi signal, making the tuning triangle severely affected by the noise. Since the 20 MHz WiFi signal has four pilot subcarriers, a natural idea is to calculate the voltage from each tuning triangle on the four pilots and then use a weighted average to calculate the final result as follows:

$$\hat{V}(t) = \frac{\sum_{p=1}^{4} w_i \hat{V}_p(t)}{\sum_{p=1}^{4} w_p}, \tag{12}$$

where $\hat{V}_p(t)$ is the analog voltage recovered from the pth pilot by applying Eq. 5 and Eq. 4, and w_p is the weight assigned to the pth pilot subcarrier. For the pilots with more severe fading, a lower weight is assigned to reduce the impact on audio recovery.

However, in in-band transmission, the backscattered signal and the original signal are mixed together, making it challenging to assess the Tx-Tag-Rx channel fading experienced by the backscattered signal from the received signal. To address this issue, we propose the fading-adaptive audio recovery. We observe that when a pilot subcarrier of the backscattered signal undergoes deep fading, its amplitude significantly decreases, resulting in a smaller tuning triangle, as shown in Fig. 13(a). The radius of the circumcircle of the tuning triangle, which is exactly equal to the amplitude of the pilots of the backscattered symbol, can theoretically be used to quantify the fading and act as w_p.

However, the radius of the circumcircle of the triangle is very sensitive to noise and may deviate significantly from the amplitude of the pilots due to the noise. Therefore, we consider other geometric features of the triangle instead of the circumradius R_{out}. Potential choices are the square root of the triangle's area A, the inradius R_{in}, or the sum of the distances from the centroid to the three vertices D_c. They are all proportional to the amplitude of the backscattered symbol and can be used to quantify the fading and act as w_p. We study the sensitivity of these features to noise using a Monte Carlo method. Specifically, we generate an ideal triangle, add the Gaussian random noise, and calculate the relative error of the features of the noisy triangle compared to the ideal triangle.



(a) The fading affects the triangle size on different pilot subcarriers. (b) The CDF of the relative error of the geometric features under noise.

Figure 13: Use the triangle's geometric features to quantify the fading.

In Fig. 13(b), we show the CDF of the relative error for each geometric feature after adding the Gaussian noise to the tuning triangle. We observe that R_{out}, which theoretically equals the pilot amplitude of the backscattered symbol, has a very high 95% relative error. Instead, D_c has both a lower median and 95% relative error than other features. This suggests that D_c more accurately captures the size of the tuning triangle and better quantifies pilot subcarrier fading. Therefore, selecting D_c as the weight improves resilience to deep fades in noisy environments. We adopt D_c as w_p in the receiver implementation and use it to recover the analog audio signal via Eq. 12.

4 Implementation

4.1 Tag

We prototype the Satori tag on a 4-layer printed circuit board (PCB) using commercial off-the-shelf components, as shown in Fig. 14. The design includes three modules: a packet detector, a modulation circuit, and a logic unit. The packet detector and modulation circuit are each connected to 2.4 GHz omnidirectional antennas with 6 dBi gain and interface with the logic unit.

The packet detector is implemented using an envelope detector LT5534 [6] and a comparator TLV3201 [20]. It can detect the arrival of a WiFi packet, allowing the Satori tag to detect the WiFi packet and start modulation.

The modulation circuit consists of two main parts: a sample-and-hold circuit and a reflective circuit. An external low-power microphone provides an analog voltage input to the sample-and-hold circuit. The sample-and-hold circuit includes two ultra-low-power amplifiers TLV9041 [21] as followers, while a CMOS switch

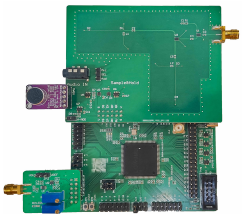


Figure 14: The prototype of the Satori tag.

Figure 15: Experiment setup of the evaluation.

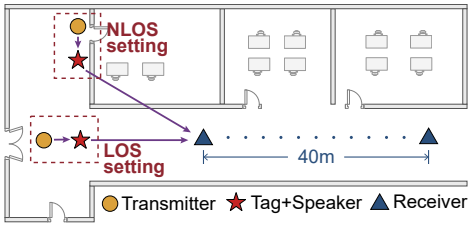


Figure 16: Floorplan of the experiment environment for LOS and NLOS evaluation.

TS5A3159 [22] toggles between sampling and hold modes. The sample-and-hold circuit's output connects to the reflective circuit. The reflective circuit has three terminal loads: two reflective-type phase shifters, each formed by a microstrip line on the PCB, and two varactor diodes MAVR-000120 [33]. The third load is a capacitor of 10 pF in series with a resistor of 9.1 Ω, where we design the register to match the reflection amplitude of phase shifters. We use two RF switches SKY13335 [40] to toggle among loads.

The control unit uses a low-power GW1N-LV4 FPGA [11]. This unit's function only includes checking the output from the detector and controlling the modulation circuit without doing any digital computation or processing. When delivered to production, the logic unit could be implemented using simple gate circuits, consuming a mere 1 to 2 μW of power.

4.2 WiFi Transceiver

We implement the WiFi Tx using the USRP N210 software-defined radio platform. We generate the IEEE 802.11n compliant baseband signal with MATLAB's Wireless Waveform Generator App on a PC and send it to the USRP N210 through GNU Radio. The Tx includes an RF power amplifier that boosts the signal's average power to 20 dBm, complying with local regulations on civil radio. We use the USRP B210 platform for the Rx. We control the USRP B210 with PicoScenes [24] software on a laptop to receive the signal and perform audio extracting via MATLAB. Both the Tx and Rx are equipped with a 2.4 GHz omnidirectional antenna with 6 dBi gain.

5 Evaluation

5.1 Experiment Methodology

We deploy the WiFi Tx, Rx, and the Satori tag in a corridor, with two speakers playing audio signals near the tag. We evaluate Satori in both line-of-sight (LOS) and non-line-of-sight (NLOS) scenarios, as shown in Fig. 16.

In the evaluation, the Tx transmits WiFi packets at a default rate of 2000 packets per second, with a default WiFi traffic duty cycle of 80%. The Tx's transmission power is set to 20 dBm. The speakers play the default audio signal that covers frequencies from 200 Hz to 3600 Hz.

To evaluate the audio transmission quality of Satori, we first record the output of the microphone on the tag at a bitrate of 320 kbps, which serves as the original audio signal. The primary metric in our evaluation is the signal-to-noise ratio (SNR) of the audio signal recovered by the Rx, with respect to the original audio signal.

We take Leggiero [34] and the Microphone Array Backscatter (M.A.B.) [71] as baselines. The former is an analog backscatter system based on ambient WiFi signals, and the latter is the state-of-the-art analog backscatter system for audio sensing. For the M.A.B. tag, the Tx provides a single-tone excitation as the carrier.

We evaluate the overall performance of Satori across different Tx-to-Tag and Tag-to-Rx distances under LOS and NLOS settings in §5.2. We also evaluate the impact of audio frequency, audio strength, the duty cycle of the WiFi traffic, and tag synchronization error on SNR in §5.3. Furthermore, we conduct ablation studies by removing the sample-and-hold circuit, CFO calibration, and fading-adaptive audio recovery from Satori in §5.4. We evaluate the performance of Satori in bursty WiFi traffic and Satori's impact on the goodput of existing WiFi in §5.6. Finally, we estimate the Satori tag power consumption in §5.5 and demonstrate the proof-of-concept applications in §5.7.

5.2 Overall Performance

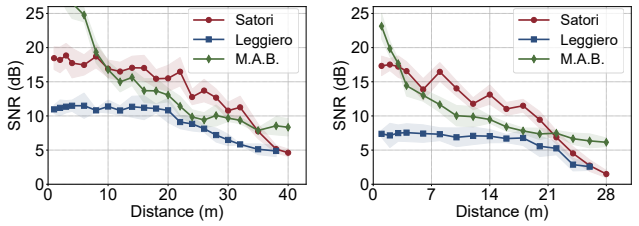
We set the Tx-to-Tag distance to 0.5 m and 2 m and evaluate the SNR of the recovered audio signal across different Tag-to-Rx distances. We evaluate in both LOS and NLOS scenarios.

Fig. 17 shows the SNR of the audio recovered at different Tag-to-Rx distances in LOS scenarios. The shadow represents the variance. We have three observations from the results.

First, Satori achieves an SNR above 18 dB when the Tag-to-Rx is near. As the distance increases, the SNR of the recovered audio decreases due to weak tag signal strength. Satori can recover audio with a 5 dB SNR at distances up to 40 m, sufficient for basic audio sensing applications. When the Tag-to-Tx distance extends to 2 m, the weaker incident WiFi signal further reduces the SNR, yet Satori still supports a communication range of 24 m.

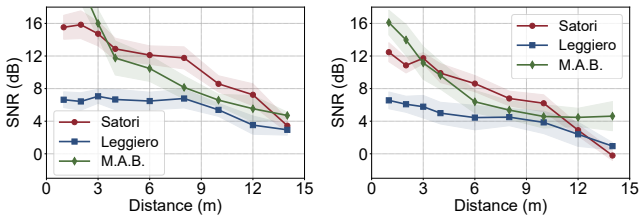
Second, we observe that Satori achieves 5 to 10 dB higher SNR than Leggiero at Tag-to-Rx distances under 30 m, with the Tx-to-Tag distance of 0.5 m. This improvement stems from Satori's symbol-level embedding of analog audio signals, enabling a higher sampling rate for broadband audio transmission. In contrast, Leggiero's packet-level embedding leads to a lower sampling rate. However, as the Tag-to-Rx distance exceeds 35 m, the SNR difference between Satori and Leggiero narrows to 0dB. This is because both Satori and Leggiero suffer from weak tag signal strength, which is the dominant factor reducing the SNR at further distance.

Third, Satori demonstrates SNR performance comparable to or better than M.A.B. at most distances. Because we use smaller Tx power to meet the local radio regulation, the SNR performance of M.A.B. is lower than that reported in [71]. When the Tag-to-Rx



(a) Tx-tag=0.5m (b) Tx-tag=2m

Figure 17: SNR performance versus Tag-to-Rx distance in the LOS setting.



(a) Tx-tag=0.5m (b) Tx-tag=2m

Figure 18: SNR performance versus Tag-to-Rx distance in the NLOS setting.

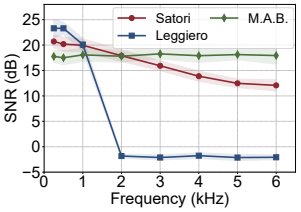


Figure 19: Impact of audio frequency on the SNR.

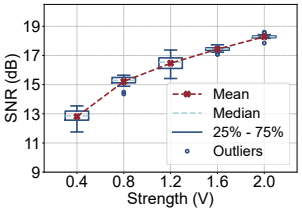


Figure 20: Impact of audio strength on the SNR.

distance is under 8 m, M.A.B. achieves a higher SNR than Satori. This is because the slight residual self-interference hinders Satori from achieving higher SNR, even if the tag and the Rx are very close. Instead, M.A.B. uses a single-tone carrier and frequency shifting, avoiding this limitation. As the Tag-to-Rx distance increases, consistent with [71], M.A.B.'s SNR drops rapidly, whereas Satori maintains a relatively high SNR. When the Tag-to-Rx distance is between 15 and 30 m, Satori's SNR is over 5 dB higher than M.A.B. (with Tx-Tag = 0.5 m).

Fig. 18 shows the recovered audio SNR at various Tag-to-Rx distances in the NLOS scenario. Due to our fading-adaptive audio recovery, Satori achieves up to 16 dB SNR in the more challenging NLOS environment and supports communication over distances exceeding 12 m. In the NLOS scenario, Satori outperforms Leggiero by up to 8 dB. Meanwhile, Satori delivers performance comparable to M.A.B., which uses dedicated devices.

5.3 Impacting Factors

We set the Tag-to-Tx distance to 0.5 m and the Tag-to-Rx distance to 8 m in the LOS scenario. We evaluate the impacting factors, including the audio frequency, the audio strength, the WiFi traffic duty cycle, and the synchronization.

5.3.1 Audio Frequency. We play single-tone audio signals ranging from 250 Hz to 6000 Hz to the Satori tag and the two baselines. We evaluate the SNR of the recovered audio.

Fig. 19 presents the results. For Satori, the SNR decreases at higher frequencies. The SNR of the 6000 Hz audio is 8 dB lower than that of the 250 Hz audio. This is because Satori converts the audio signal into its reflected phase in real time. However, when there is no WiFi signal to reflect, the audio signal is lost. And the loss has a greater impact on the recovery of high-frequency audio. In contrast, M.A.B., using a constant carrier, maintains consistent performance across all frequencies. Leggiero, which embeds audio at the packet level, has a significant performance drop when audio frequencies exceed half the packet rate (i.e., 1000 Hz). Overall, Satori performs well in transmitting the audio of the low-frequency range.

5.3.2 Audio Strength. In practice, the audio signal strength may vary with the distance between the microphone and the sound source. We evaluate the impact of the test audio strength on the recovered audio SNR. In this evaluation, we use the voltage peak-to-peak (Vpp) value output from the microphone as the measure of audio strength. The results are shown in Fig. 20.

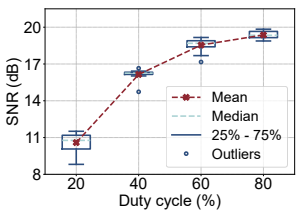
We find that the SNR of the recovered audio increases with the audio strength. This is because larger audio amplitudes lead to more significant phase changes in the backscattered signal, making it more resilient to noise. We also find that when the audio strength decreases from 2.0 Vpp to 0.4 Vpp and the audio signal energy drops by 25× (14 dB), the SNR only decreases by 6 dB. This indicates that Satori has robustness to changes in audio strength.

5.3.3 Duty Cycle of WiFi Traffic. Given that Satori embeds audio into WiFi signals, we set the WiFi traffic duty cycle to 20%, 40%, 60%, and 80% and evaluate the impact of the duty cycles on recovering the broadband test audio. We also evaluate the impact of different duty cycles on audio signals with frequencies of 250 Hz, 500 Hz, 1000 Hz, and 1600 Hz.

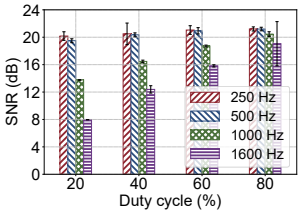
Fig. 21(a) shows the recovered audio SNR across different WiFi duty cycles. We observe that as the duty cycle decreases, Satori's SNR also drops. At an 80% duty cycle, Satori achieves about 19 dB SNR, while at 20%, it reaches around 11 dB. This is because as the duty cycle decreases, more audio signals are lost, reducing the quality of audio recovery. Fig. 21(b) further shows that higher frequencies are more sensitive to duty cycle changes, as the high-frequency audio is more susceptible to the loss.

Although the WiFi traffic duty cycle impacts Satori's performance, some high-throughput devices in industrial environments, such as industrial cameras and automated guided vehicles (AGV), can provide a high enough WiFi traffic duty cycle to ensure Satori's performance.

5.3.4 Synchronization Error. Since Satori needs to start embedding at specific WiFi-pack positions, we evaluate the impact of different synchronization errors on the recovered audio SNR. In this evaluation, we change the synchronization error with a step of 0.8 μs and



(a) Impact of WiFi traffic duty cycle on the recovered test audio.



(b) Impact of WiFi traffic duty cycle on different audio frequencies.

Figure 21: Impact of the WiFi traffic duty cycle on the SNR of the recovered audio.

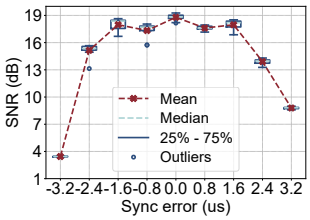


Figure 22: Impact of the synchronization error on the SNR.

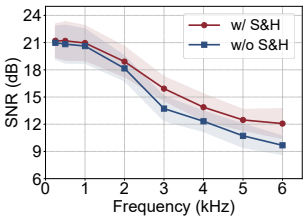


Figure 23: The SNR enhancement from the sample-and-hold circuit.

evaluate the impact of synchronization errors ranging from 0 to ±3.2 μs. The results are shown in Fig. 22.

When the synchronization error is within 1.6 μs, the recovered audio SNR is almost unaffected, remaining above 18 dB. When the synchronization error exceeds 2.4 μs, the SNR drops rapidly. The SNR falls below 10 dB when the synchronization error exceeds 3 μs. The reason is that the tag's toggling cycle is twice that of a WiFi symbol. If the synchronization error is less than half the duration of a WiFi symbol, the even-index payload WiFi symbols can still be fully modulated by the tag and used for audio recovery. Thus, we think Satori is robust to synchronization errors caused by the ultra-low-power packet detection with low performance.

5.4 Ablation Study

We conduct the ablation study by removing the sample-and-hold circuit, CFO calibration, and the weight in fading-adaptive audio recovery, assessing their impact on audio transmission.

5.4.1 Effectiveness of the Sample-and-hold Circuit. We play the audio with different frequencies ranging from 250 Hz to 6000 Hz to Satori, evaluating the SNR of the recovered audio signals of different frequencies with and without the sample-and-hold circuit. The Tx-to-Tag and Tag-to-Rx distances are set to 0.5 m and 8 m, respectively. We show the results in Fig. 23, where the sample-and-hold circuit is denoted as S&H.

When audio frequencies are below 1000 Hz, the SNR remains similar with or without the sample-and-hold circuit. However, as the frequency increases, Satori without the sample-and-hold circuit experiences a faster SNR decline, with a drop of over 2 dB compared to Satori with the sample-and-hold circuit. This is because the sample-and-hold circuit stabilizes the voltage over each backscattered WiFi symbol, improving audio transmission quality, especially at higher frequencies. Given the minor impact of removing the sample-and-hold circuit on low-frequency audio, it can be replaced by a simple low-pass filter to reduce power consumption in applications that don't require high-frequency audio transmission.

5.4.2 SNR Enhancement through the CFO Calibration. We evaluate the impact of CFO calibration on audio transmission by partially or fully removing the two-step calibration. Satori with only the first step is termed "rough calibration," with no calibration as "no calibration," and with full calibration as "full calibration." We set the Tx-to-Tag distances to 0.5 m and 2 m and vary the Tag-to-Rx distance. We compare the SNR of the recovered audio across these CFO calibration modes. Fig. 24 shows the results.

Table 1: The power consumption of the PCB prototype of Satori.

Part	Logic unit	Packet detection	Clock
Power	8.9 mW	23.9 mW	12.9 mW
Part	S&H circuit	Switch	Total
Power	33 μW	31 μW	45.6 mW

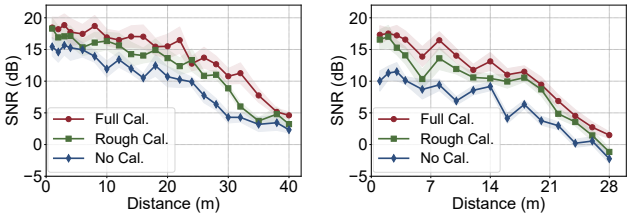
We find that Satori with full calibration achieved the highest SNR, followed by rough calibration, with no calibration yielding the lowest SNR. At closer Tag-to-Rx distances, the SNR difference between full calibration and no calibration is around 4 dB. When the Tag-to-Tx distance increases to 2m, the SNR enhancement of the CFO calibration is more significant. The SNR of the full calibration is over 7 dB higher than that of no calibration. This is because the increased Tx-to-Tag distance weakens the backscattered signals. Since the CFO affects these weaker signals more in in-band transmission, the enhancement of CFO calibration on audio recovery becomes more significant.

5.4.3 Improvement from the Weight in Audio Recovery. We modify the fading-adaptive audio recovery by setting equal weights for all pilot subcarriers (denoted as w/o weight) and compare it to the original weighted method (w/ weight) in the fading-adaptive audio recovery. We vary the Tag-to-Rx distance, evaluating the SNR improvement from the weight of the audio recovery in both LOS and NLOS scenarios.

Fig. 25 shows the results. We find that w/ weight achieves higher SNR in both LOS and NLOS scenarios. In the LOS scenario, the improvement is modest, with a maximum increase of around 5dB at certain distances. This is because the fading difference between subcarriers is slight in LOS scenarios, resulting in a limited SNR improvement from the weight. In the NLOS scenario, w/ weight shows a more significant improvement, with an increase of over 8dB. This is mainly because the NLOS channel has more complex multipath fading, and the frequency-selective fading is more severe than in the LOS scenario. The w/ weight Satori can better adapt to channel fading, improving audio recovery quality.

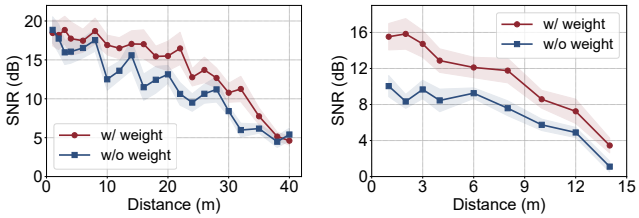
5.5 Power Consumption

We evaluate the power consumption of the PCB prototype of the Satori tag and each of its components. The components we evaluate include the logic unit of the Satori tag (i.e., the FPGA), low-frequency clock generation, packet detection, the sample-and-hold



(a) Tx-tag=0.5m (b) Tx-tag=2m

Figure 24: SNR comparison between the Satori with different CFO calibration settings.



(a) LOS scenario (b) NLOS scenario

Figure 25: SNR comparison between the Satori w/ and w/o the weight in fading-adaptive audio recovery.

circuit, the RF switches, and the phase shifters. The logic unit operates at a core voltage of 1.2 V, while all other components operate at 3.3 V. The power consumption of the FPGA was estimated using the EDA toolbox provided by Gowin [13] For the remaining components, we measure the current on the power supply line using a VICTOR 8155 multimeter [49] and calculate the power consumption by multiplying the measured current by the supply voltage. Since the minimum measurement error of the VICTOR 8155 is 0.1 μA, we improve the current reading accuracy by averaging multiple measurements. For components with a current consumption below this threshold, we estimate the power consumption based on the values provided in the datasheet.

The FPGA core consumes 8.9 mW of power, which includes 6.5 mW of static circuit power and 2.4 mW of dynamic power. The clock generation module consists of a 1 MHz crystal oscillator, requiring a supply current of 3.9 mA and consuming approximately 12.9 mW of power. The power consumption of the packet detection module mainly comes from the active envelope detector LT5534, which consumes 23.8 mW of power in our PCB design. The comparator TLV3201 has a static power consumption of 120 μW, which accounts for only a small fraction. In the sample-and-hold circuit, each voltage follower TLV9041 operates with a current draw of 10 μA. Given the alternating operation of the sample-and-hold working mechanism, their total power consumption amounts to 33 μW. The typical current consumption of the two analog switches in sample-and-hold circuit is less than 20 nA according to the datasheet [22], leading to a power consumption of just 66 nW. Each RF switch requires a current of 5 μA, resulting in a combined power consumption of 31 μW for the two switches. The varactor diodes in the phase shifters are reverse-biased and draw an extremely low bias current of only 1.6 pA [33], as specified in the datasheet. Consequently, the power consumption of the phase shifters is negligible.

In summary, the total power consumption of the PCB prototype of the Satori tag is approximately 46 mW, as detailed in Tab. 1. Since this work focuses on analog backscatter, the most critical modules are the RF switches and phase shifters. Our measurements and estimations indicate that the power consumption of these components remains at the μW level. Compared to existing works, our power optimization for the backscatter tag primarily focuses on eliminating the need for high-speed digital processing and high-frequency clocks required for frequency shifting. The digital backscatter usually uses a low-power MCU (e.g., MSP430[12]) to process audio signals up to 40 kHz, which consumes hundreds of μW. In contrast, our approach directly embeds the analog voltage signal into the

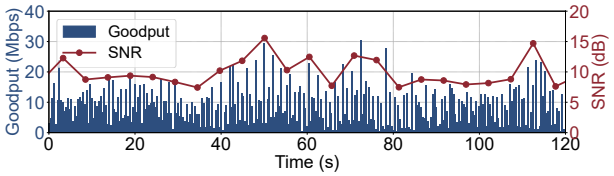


Figure 26: SNR performance in bursty WiFi traffic.

phase of the reflected signal using a varactor-based phase shifter, incurring negligible power overhead. Moreover, our tag only requires a low-speed clock of hundreds of kHz to achieve in-band audio embedding, in contrast to backscatter systems that require clocks of tens of MHz for frequency shifting [34, 66, 67]. Since clock power consumption is generally proportional to clock frequency, this reduction in clock speed decreases power consumption by an order of magnitude. For the logic control unit, an ASIC implementation can further reduce power consumption to the μW level.

5.6 Operating in WiFi Networks

To assess the performance of Satori in WiFi networks with bursty traffic, we use the commercial network interface card (NIC) QCA9300 to transmit burst packets within 2 minutes. This bursty traffic is commonly seen in real-time video streaming. We record the WiFi goodput 3 times per second and measure the SNR of the audio segments recovered by Satori with 5-second intervals. The measurement results are presented in Fig. 26. The SNR of the audio recovered by Satori varies in real-time with the WiFi throughput. This is mainly because a lower throughput often implies a lower duty cycle, which leads to a decrease in the quality of the audio recovered by Satori.

We also evaluate Satori's impact on existing WiFi networks by measuring the WiFi goodput with the modulation settings of MCS=2 (QPSK), MCS=4 (16QAM) and MCS=7 (64QAM). We use another QCA9300 NIC as the WiFi receiver. The distance between WiFi Tx and Rx is set to 4 m, and we gradually decrease the Tag-to-Tx distance from 0.5 m. A blank control experiment without Satori was conducted as a baseline. Since the MCS settings only differ in the modulation order of each subcarrier and the coding scheme, without changing the number of pilots and the symbol rate, Satori operates with the same performance under different MCS settings.

The results are shown in Fig. 27. We find that for MCS=2 and MCS=4, Satori's impact on WiFi goodput is less than 1% at any Tag-to-Tx distance, even though it continuously alters the channel and introduces asynchronous tag modulation to the in-band signal.. This

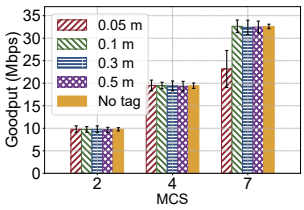


Figure 27: Satori's impact on WiFi networks with different MCS at varying Tx-Tag distances.

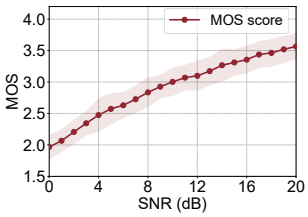


Figure 28: The MOS score corresponding to different SNR.

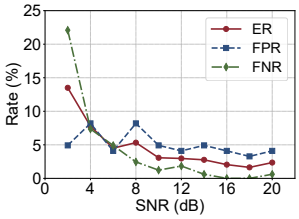


Figure 29: The error rate of industrial fault detection in different SNR.

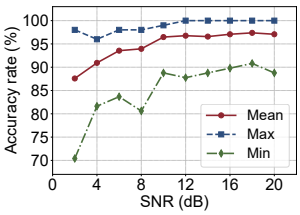


Figure 30: The accuracy of voice command recognition in different SNR.

is because the WiFi receiver can use pilot subcarriers to correct the slight signal changes introduced by the tag. For MCS=7, Satori only impacts WiFi goodput when the Tag-to-Tx distance is less than 5 cm. This is because the tag performs phase modulation and continuously changes the channel within a packet, while the Rx performs channel equalization based on the Channel State Information (CSI) detected from the preamble. As a result, the payload symbols received by the Rx are offset due to the channel changing. When the tag is close to the Tx, the channel change is more obvious and can't be tolerated by high-MCS WiFi, leading to a significant decrease in the goodput of the existing WiFi network.

5.7 Proof-of-Concept Application

To comprehensively demonstrate the performance of Satori in applications such as speech audio transmission, we use VISQOL[19] to measure the Mean Opinion Score (MOS) of the audio transmitted by Satori under different SNRs, as shown in Fig. 28. Generally, a MOS of 3 indicates acceptable speech quality. Satori can achieve this quality when the SNR is 10 dB, which corresponds to to the Tag-Rx distance of 30 m.

We also demonstrate Satori's applications in industrial fault detection and command recognition. For fault detection, we use the MiMii dataset [17] with sounds of machinery under normal and abnormal conditions, processing audio through Satori and classifying with a CAM++ network [54]. We evaluate the error rate (ER), false positive rate (FPR), and false negative rate (FNR) across different SNR levels. For command recognition, we use 10 common control commands in the Google speech commands dataset [57]. We input these audio commands into Satori and classify them via the CAM++ network. We assess the average accuracy across SNR levels and the maximum and minimum accuracy rates among all commands.

Fig. 29 shows the results for industrial fault detection. When SNR is above 6 dB, Satori achieves an average error rate below 5%. For the more critical FNR in fault detection, Satori achieves below 2.5%. Combined with the results in §5.2, this indicates that Satori can reliably transmit the audio signal for industrial equipment faults detection even at a Tag-to-Rx distance of 35 m.

Fig. 30 presents the results for speech command recognition. When the SNR exceeds 10 dB, Satori achieves an average accuracy of over 95%, with the lowest accuracy for individual commands still above 85%. This demonstrates that Satori can support more complex tasks of speech command recognition.

6 Discussion

Low WiFi Duty Cycle: Our design embeds audio signals onto ambient WiFi signals, so the WiFi duty cycle has a non-negligible impact on audio recovery. However, we note that the low WiFi duty cycle primarily affects high-frequency audio components, while most audio information is concentrated in the low frequencies [68]. Leveraging this characteristic, we can apply compressed sensing techniques or utilize trained neural networks for audio recovery to accomplish audio sensing tasks effectively [7].

Compatibility with Existing WiFi: Satori utilizes the pilot subcarriers in WiFi OFDM to transmit audio signals. Consequently, Satori can operate under various WiFi protocols, including the newer IEEE 802.11ax/be, as these protocols all define the pilot subcarriers. However, commercial devices still can't serve as Satori receivers because they don't provide access to pilots on the PHY layer for per-symbol channel estimation. We believe that Satori will be able to run on commercial devices in the future, as an increasing number of PHY interfaces are being provided by commercial devices to enable various applications [60].

Multi-tag/Tx Scenarios: In some scenarios, there may be multiple WiFi Txs and multiple Satori tags deployed. This may lead to two types of collisions: *1)* between tags excited by different Txs and *2)* between tags excited by the same Tx. For the first type, because the CSMA/CA mechanism ensures the Rx receive the signal from only one Tx at a time, the tags excited by different Txs will not collide. For the second type, since the tags can detect the Tx's signal via the envelope detector, a simple downlink can be enabled, and the Tx can coordinate the tags it excites to avoid collisions.

7 Related Works

With the development of IoT sensing systems, the backscatter technology has gained attention as a low-power method for transmitting sensor data[16, 23, 25, 30, 34, 55]. Based on how the sensor information is embedded, the backscatter can be classified into analog and digital types. We introduce related work of both categories and highlight the difference between Satori and these works. Besides, from the perspective of audio sensing, we summarize the relevant works on RF-based audio sensing technologies.

7.1 Analog Backscatter

Typically, the analog backscatter directly embeds analog signals into reflected signals by amplitude, frequency, pulse, or phase modulation, avoiding processing them in the digital domain.

[45, 46] achieves the amplitude modulation by continuously adjusting the reflected signal's amplitude using the MOSFET. Specifically, [45] demonstrates a battery-free audio communication system through amplitude modulation. However, the amplitude modulation inevitably weakens the reflected signal's strength, and they don't handle the self-interference, leading to short communication ranges. [37, 38] achieves the frequency modulation by continuously adjusting the reflected signal's frequency based on the input analog signal. However, these solutions require a voltage-controlled oscillator (VCO) to modulate the frequency of the backscattered signal, resulting in higher power consumption.

[25, 35, 39, 70, 71] employs pulse modulation by varying the width, position, and interval of pulses in the reflected signal. [35] demonstrates high-definition video transmission, showing the potential of analog backscatter for multimedia data transmission. [71] achieves multi-channel audio transmission through pulse position modulation (PPM). However, they require dedicated devices transmitting carrier and frequency shifting, while Satori achieves inband transmission over the ambient WiFi signals, avoiding the deployment cost and high power consumption.

Researchers also explore directly embedding analog signals into the phase of the reflected signal [27, 34, 43]. In [34], the analog signal is embedded in the phase of ESS CSI in WiFi packets, enabling analog voltage signal transmission at the packet level. In contrast, Satori embeds the audio signal into the WiFi signal at the symbol level, enabling high-quality audio transmission. Moreover, [34] requires frequency shifting to move the backscattered signal into another blank band with 20MHz bandwidth. This results in higher power consumption and larger bandwidth overhead. In contrast, Satori avoids this issue through in-band transmission.

7.2 Digital Backscatter

Digital backscatter first converts the sensor data into digital signals and then embeds it into the backscattered signal. Traditional digital backscatter, such as RFID[18, 72], requires dedicated devices to communicate with them. Researchers consider embedding digital signals into ambient signals, such as Wi-Fi[8, 10, 31, 55, 62, 63, 66, 67, 69], LTE[5], BLE[65], TV[30], FM[50], LoRa[15, 16, 23, 36], and even mmWave[3, 32] signals.

Some digital backscatter works transmit data over pilot subcarriers or achieve in-band transmission. [51, 56, 60] use the pilot subcarriers as a single tone and shift them to another band, enabling the digital signal transmission. [1, 9, 26, 28] demonstrate in-band transmission. Specifically, [9] uses quasi-orthogonal OFDM codewords to eliminate self-interference and reach higher throughput. Compared with these works, Satori achieves inband analog audio embedding and leverages the fixed pilot subcarriers to recover the analog modulated signals.

7.3 RF-Based Audio Sensing

From the perspective of audio sensing applications, RF-based wireless audio sensing [4, 29, 64] has gained significant attention. [58, 64] use mmWave to detect the vibrations for audio sensing, which often require LOS to the vibrating target. Some works use WiFi [53] and RFID [4, 59] signals for NLOS audio sensing but are limited by sampling rates and only capture basic audio features. Unlike these works, Satori leverages a sensor-based approach to achieve ultra-low-power audio sensing, enabling the recovery of more complex audio signals.

8 Conclusion

This paper introduces Satori, an in-band analog backscatter system that enables ultra-low-power audio transmission. Satori tag exploits analog modulation to embed the analog audio signal into every WiFi symbol directly, and transmits the audio inside the frequency band of the excitation, significantly reducing the power consumption on the tag. Moreover, the Satori receiver carefully eliminates the effect of CFO and frequency-selective fading, ensuring the quality of audio transmission. From a broader perspective, Satori follows the paradigm of RF computing [14, 15, 34], offering a direct manipulation from analog audio to the phase of RF signals. Our evaluation shows that Satori operates at a sampling rate of 41.67 kHz and achieves an SNR exceeding 18 dB for the transmitted audio in an indoor environment.

Acknowledgments

We sincerely thank our anonymous shepherd and all reviewers for their valuable feedback. This work is supported by the National Natural Science Foundation of China under grant No. 62425207.

References

[1] Ali Abedi, Farzan Dehbashi, Mohammad Hossein Mazaheri, Omid Abari, and Tim Brecht. 2020. Witag: Seamless wifi backscatter communication. In *Proceedings of the ACM SIGCOMM*.

[2] Francois Burdin, Zyad Iskandar, Florence Podevin, and Philippe Ferrari. 2015. Design of Compact Reflection-Type Phase Shifters With High Figure-of-Merit. *IEEE Transactions on Microwave Theory and Techniques* (2015).

[3] Yoon Chae, Zhenzhe Lin, Kang Min Bae, Song Min Kim, and Parth Pathak. 2024. mmComb: High-speed mmWave Commodity WiFi Backscatter. In *Proceedings of the USENIX NSDI*.

[4] Yunzhong Chen, Jiadi Yu, Yingying Chen, Linghe Kong, Yanmin Zhu, and Yi-Chao Chen. 2024. RFSpy: Eavesdropping on Online Conversations with Out-of-Vocabulary Words by Sensing Metal Coil Vibration of Headsets Leveraging RFID. In *Proceedings of the ACM Mobisys*.

[5] Zicheng Chi, Xin Liu, Wei Wang, Yao Yao, and Ting Zhu. 2020. Leveraging ambient lte traffic for ubiquitous passive communication. In *Proceedings of the ACM SIGCOMM*.

[6] Anlog Device. 2024. Analog Devices LT5534 RF Power Detectors. https://www.analog.com/en/products/lt5534.html

[7] David L Donoho. 2006. Compressed sensing. *IEEE Transactions on information theory* (2006).

[8] Caihui Du, Jiahao Liu, Shuai Wang, Rongrong Zhang, Wei Gong, and Jihong Yu. 2023. Timespan-based backscatter using a single cots receiver. In *Proceedings of the ACM MobiSys*.

[9] Caihui Du, Jihong Yu, Rongrong Zhang, Ju Ren, and Jianping An. 2024. Orthcatter: High-throughput In-band OFDM Backscatter with Over-the-Air Code Division. In *Proceedings of the USENIX NSDI*.

[10] Manideep Dunna, Miao Meng, Po-Han Wang, Chi Zhang, Patrick Mercier, and Dinesh Bharadia. 2021. SyncScatter: Enabling WiFi like synchronization and range for WiFi backscatter communication. In *Proceedings of the USENIX NSDI*.

[11] Gowin. 2024. Gowin GW1N-LV4 FPGA. https://www.gowinsemi.com/en/product/detail/46/

[12] Gowin. 2024. TI MSP430FR5969 Low-power MCU. https://www.ti.com/product/MSP430FR5969

[13] Gowin. 2025. Gowin EDA. https://www.gowinsemi.com/en/support/home

[14] Xiuzhen Guo, Yuan He, Zihao Yu, Jiacheng Zhang, Yunhao Liu, and Longfei Shangguan. 2022. RF-transformer: a unified backscatter radio hardware abstraction. In *Proceedings of the ACM MobiCom*.

[15] Xiuzhen Guo, Longfei Shangguan, Yuan He, Nan Jing, Jiacheng Zhang, Haotian Jiang, and Yunhao Liu. 2022. Saiyan: Design and implementation of a low-power demodulator for LoRa backscatter systems. In *Proceedings of the USENIX NSDI*.

[16] Xiuzhen Guo, Longfei Shangguan, Yuan He, Jia Zhang, Haotian Jiang, Awais Ahmad Siddiqi, and Yunhao Liu. 2020. Aloba: Rethinking ON-OFF keying modulation for ambient LoRa backscatter. In *Proceedings of the ACM SenSys*.

[17] Purohit Harsh, Tanabe Ryo, Ichige Kenji, Endo Takashi, Nikaido Yuki, Suefusa Kaori, and Kawaguchi Yohei. 2019. MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection. In *Proceedings of the DCASE*.

[18] Yuan He, Yilun Zheng, Meng Jin, Songzhen Yang, Xiaolong Zheng, and Yunhao Liu. 2019. Red: Rfid-based eccentricity detection for high-speed rotating machinery. *IEEE Transactions on Mobile Computing* (2019).

[19] Andrew Hines, Jan Skoglund, Anil C Kokaram, and Naomi Harte. 2015. ViSQOL: an objective speech quality model. *EURASIP Journal on Audio, Speech, and Music Processing* (2015).

[20] Texas Instruments. 2024. TI TLV3201 Low-power Comparator. https://www.ti.com/product/TLV3201

[21] Texas Instruments. 2024. TI TLV9041 Ultra Low-power Operational Amplifier. https://www.ti.com/product/TLV9041

[22] Texas Instruments. 2024. TI TS5A3159 SPDT Analog Switch. https://www.ti.com/product/TS5A3159

[23] Jinyan Jiang, Zhenqiang Xu, Fan Dang, and Jiliang Wang. 2021. Long-range ambient LoRa backscatter with parallel decoding. In *Proceedings of the ACM MobiCom*.

[24] Zhiping Jiang. 2024. PicoScenes. https://ps.zpj.io/index.html

[25] Wenli Jiao, Yanlin Li, Xiangdong Xi, Ju Wang, Dingyi Fang, and Xiaojiang Chen. 2023. Bioscatter: Low-power sweat sensing with backscatter. In *Proceedings of the ACM Mobisys*.

[26] Bryce Kellogg, Aaron Parks, Shyamnath Gollakota, Joshua R Smith, and David Wetherall. 2014. Wi-Fi backscatter: Internet connectivity for RF-powered devices. In *Proceedings of the ACM SIGCOMM*.

[27] Nabil Khalid, Rashid Mirzavand, Hossein Saghlatoon, Mohammad Mahdi Honari, Ashwin K Iyer, and Pedram Mousavi. 2021. A batteryless RFID sensor architecture with distance ambiguity resolution for smart home IoT applications. *IEEE Internet of Things Journal* (2021).

[28] Taekyung Kim and Wonjun Lee. 2018. Exploiting residual channel for implicit Wi-Fi backscatter networks. In *Proceedings of the IEEE INFOCOM*.

[29] Huining Li, Chenhan Xu, Aditya Singh Rathore, Zhengxiong Li, Hanbin Zhang, Chen Song, Kun Wang, Lu Su, Feng Lin, Kui Ren, et al. 2020. Vocalprint: exploring a resilient and secure voice authentication via mmwave biometric interrogation. In *Proceedings of the ACM Sensys*.

[30] Vincent Liu, Aaron Parks, Vamsi Talla, Shyamnath Gollakota, David Wetherall, and Joshua R Smith. 2013. Ambient backscatter: Wireless communication out of thin air. *ACM SIGCOMM computer communication review* (2013).

[31] Xin Liu, Zicheng Chi, Wei Wang, Yao Yao, Pei Hao, and Ting Zhu. 2021. Verification and redesign of OFDM backscatter. In *Proceedings of the USENIX NSDI*.

[32] Haofan Lu, Mohammad Mazaheri, Reza Rezvani, and Omid Abari. 2023. A Millimeter Wave Backscatter Network for Two-Way Communication and Localization. In *Proceedings of the ACM SIGCOMM*.

[33] MACOM. 2024. MACOM MAVR-000120 Varactor Diode. https://www.macom.com/products/product-detail/MAVR-000120-14110P

[34] Xin Na, Xiuzhen Guo, Zihao Yu, Jia Zhang, Yuan He, and Yunhao Liu. 2023. Leggiero: Analog wifi backscatter with payload transparency. In *Proceedings of the ACM Mobisys*.

[35] Saman Naderiparizi, Mehrdad Hessar, Vamsi Talla, Shyamnath Gollakota, and Joshua R Smith. 2018. Towards Battery-Free HD video streaming. In *Proceedings of the USENIX NSDI*.

[36] Yao Peng, Longfei Shangguan, Yue Hu, Yujie Qian, Xianshang Lin, Xiaojiang Chen, Dingyi Fang, and Kyle Jamieson. 2018. PLoRa: A passive long-range data network from ambient LoRa transmissions. In *Proceedings of the ACM SIGCOMM*.

[37] Dilushi Piumwardane, Madhushanka Padmal, Vaishnavi Ranganathan, Christian Rohner, and Thiemo Voigt. 2022. HarmonicID: An identification system for low-power analog backscatter tags. In *Proceedings of the IEEE RFID*.

[38] Vaishnavi Ranganathan, Sidhant Gupta, Jonathan Lester, Joshua R Smith, and Desney Tan. 2018. Rf bandaid: A fully-analog and passive wireless interface for wearable sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2018).

[39] Ali Saffari, Mehrdad Hessar, Saman Naderiparizi, and Joshua R Smith. 2019. Battery-free wireless video streaming camera system. In *Proceedings of the IEEE RFID*.

[40] Skyworks. 2024. Skyworks SKY13335 SPDT RF Switch. https://www.skyworksinc.com/en/Products/switches/SKY13335-381LF

[41] Essam Sourour, Hussein El-Ghoroury, and Dale McNeill. 2004. Frequency offset estimation and correction in the IEEE 802.11a WLAN. In *Proceedings of IEEE Vehicular Technology*.

[42] Michael Speth, Stefan A Fechtel, Gunnar Fock, and Heinrich Meyr. 1999. Optimum receiver design for wireless broad-band systems using OFDM. I. *IEEE Transactions on communications* (1999).

[43] Xue Sun, Jie Xiong, Chao Feng, Xiaohui Li, Jiayi Zhang, Binghao Li, Dingyi Fang, and Xiaojiang Chen. 2024. Gastag: A Gas Sensing Paradigm using Graphene-based Tags. In *Proceedings of the ACM MobiCom*.

[44] Yimiao Sun, Yuan He, Jiacheng Zhang, Xin Na, Yande Chen, Weiguo Wang, and Xiuzhen Guo. 2023. BIFROST: Reinventing WiFi signals based on dispersion effect for accurate indoor localization. In *Proceedings of the ACM Sensys*.

[45] Vamsi Talla, Bryce Kellogg, Shyamnath Gollakota, and Joshua R Smith. 2017. Battery-free cellphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2017).

[46] Vamsi Talla and Joshua R Smith. 2013. Hybrid analog-digital backscatter: A new approach for battery-free sensing. In *Proceedings of the IEEE RFID*.

[47] Jit Ken Tan. 2006. *An adaptive orthogonal frequency division multiplexing baseband modem for wideband wireless channels*. Ph. D. Dissertation.

[48] Lili Tang, Hui Tian, Hui Huang, Shuangjin Shi, and Qingzhi Ji. 2023. A survey of mechanical fault diagnosis based on audio signal analysis. *Elsevier Measurement* (2023).

[49] Victor. 2025. Victor 8155. http://www.china-victor.com/content/show/721/496

[50] Anran Wang, Vikram Iyer, Vamsi Talla, Joshua R Smith, and Shyamnath Gollakota. 2017. FM backscatter: Enabling connected cities and smart fabrics. In *Proceedings of the USENIX NSDI*.

[51] Bingbing Wang, Fengyuan Zhu, Wenhui Li, Zeming Yang, Meng Jin, and Xiaohua Tian. 2024. Frequency-agile OFDM Backscatter. In *Proceedings of the ACM Mobisys*.

[52] Feng Wang and Jiangchuan Liu. 2010. Networked wireless sensor data collection: issues, challenges, and approaches. *IEEE Communications Surveys & Tutorials* (2010).

[53] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M Ni. 2014. We can hear you with Wi-Fi!. In *Proceedings of the ACM MobiCom*.

[54] Hui Wang, Siqi Zheng, Yafeng Chen, Luyao Cheng, and Qian Chen. 2023. CAM++: A Fast and Efficient Network for Speaker Verification Using Context-Aware Masking. *arXiv preprint arXiv:2303.00332* (2023).

[55] Qiwei Wang, Si Chen, Jia Zhao, and Wei Gong. 2021. Rapidrider: Efficient wifi backscatter with uncontrolled ambient signals. In *Proceedings of the IEEE INFOCOM*.

[56] Qiwei Wang, Jia Zhao, and Wei Gong. 2024. PilotScatter: High-Throughput OFDM Backscatter via Pilot Tones. *IEEE Transactions on Wireless Communications* (2024).

[57] Pete Warden. 2018. Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv preprint arXiv:1804.03209* (2018).

[58] Chenhan Xu, Zhengxiong Li, Hanbin Zhang, Aditya Singh Rathore, Huining Li, Chen Song, Kun Wang, and Wenyao Xu. 2019. Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 14–26.

[59] Panlong Yang, Yuanhao Feng, Jie Xiong, Ziyang Chen, and Xiang-Yang Li. 2020. Rf-ear: Contactless multi-device vibration sensing and identification using cots rfid. In *Proceedings of the IEEE INFOCOM*.

[60] Yifan Yang, Longzhi Yuan, Jia Zhao, and Wei Gong. 2022. Content-agnostic backscatter from thin air. In *Proceedings of the ACM Mobisys*.

[61] Koji Yatani and Khai N Truong. 2012. Bodyscope: a wearable acoustic sensor for activity recognition. In *Proceedings of the ACM UbiComp*.

[62] Jihong Yu, Caihui Du, Jiahao Liu, Rongrong Zhang, and Shuai Wang. 2023. SubScatter: Subcarrier-Level OFDM Backscatter. In *Proceedings of the IEEE INFOCOM*.

[63] Longzhi Yuan and Wei Gong. 2023. Enabling native wifi connectivity for ambient backscatter. In *Proceedings of the ACM MobiSys*.

[64] Jia Zhang, Yinian Zhou, Rui Xi, Shuai Li, Junchen Guo, and Yuan He. 2022. Ambiear: mmwave based voice recognition in nlos scenarios. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* (2022).

[65] Maolin Zhang, Si Chen, Jia Zhao, and Wei Gong. 2021. Commodity-level BLE backscatter. In *Proceedings of the ACM Mobisys*.

[66] Pengyu Zhang, Dinesh Bharadia, Kiran Joshi, and Sachin Katti. 2016. Hitchhike: Practical backscatter using commodity wifi. In *Proceedings of the ACM SenSys*.

[67] Pengyu Zhang, Colleen Josephson, Dinesh Bharadia, and Sachin Katti. 2017. Freerider: Backscatter communication using commodity radios. In *Proceedings of the ACM CoNEXT*.

[68] Tong Zhang and C-C Jay Kuo. 2001. Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on speech and audio processing* (2001).

[69] Jia Zhao, Wei Gong, and Jiangchuan Liu. 2018. Spatial stream backscatter using commodity wifi. In *Proceedings of the ACM MobiSys*.

[70] Jia Zhao, Wei Gong, and Jiangchuan Liu. 2020. Towards scalable backscatter sensor mesh with decodable relay and distributed excitation. In *Proceedings of the ACM Mobisys*.

[71] Jia Zhao, Wei Gong, and Jiangchuan Liu. 2021. Microphone array backscatter: An application-driven design for lightweight spatial sound recording over the air. In *Proceedings of the ACM MobiCom*.

[72] Yang Zou, Xin Na, Xiuzhen Guo, Yimiao Sun, and Yuan He. 2024. Trident: Interference Avoidance in Multi-reader Backscatter Network via Frequency-space Division. In *Proceedings of the IEEE INFOCOM*.