WINC: A Wireless IoT Network for Multi-Noise Source Cancellation

Ishani Janveja
janveja2@illinois.edu
UIUC
Urbana, USA

Jiaming Wang
jw27@illinois.edu
UIUC
Urbana, USA

Junfeng Guan
junfeng.guan@epfl.ch
UIUC
Urbana, USA

Suraj Jog
surajjog@microsoft.com
Microsoft Research
Redmond, USA

Haitham Hassanieh
haitham.alhassanieh@epfl.ch
EPFL
Lausanne, Switzerland

ABSTRACT

This paper introduces Wireless IoT-based Noise Cancellation (WINC) which defines a framework for leveraging a wireless network of IoT microphones to enhance active noise cancellation in noise-canceling headphones. The IoT microphones forward ambient noise to the headphone over the wireless link which travels a million times faster than sound and gives the headphone a *future lookahead* into the incoming noise. While leveraging wireless *lookahead* has been explored in past work, prior systems are limited to a single noise source. WINC, however, can simultaneously cancel multiple noise sources by using a network of IoT nodes. Scaling wireless *lookahead* aware noise cancellation is non-trivial because the computational and protocol delays can defeat the purpose of leveraging wireless *lookahead*. WINC introduces a novel algorithm that operates in the frequency domain to efficiently cancel multiple noise sources. We implement and evaluate WINC to show that it can cancel three noise sources and outperforms past work and state-of-the-art headphones without requiring completely blocking the users' ears.

CCS CONCEPTS

• **Networks** → **Sensor networks**; • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Hardware** → *Digital signal processing*; *Sound-based input / output*.

KEYWORDS

Multi-source Noise Cancellation, Active Noise Cancellation, Acoustics, Internet of Things, Edge Computing, Frequency Domain Adaptive Filter, Earable

ACM Reference Format:
Ishani Janveja, Jiaming Wang, Junfeng Guan, Suraj Jog, and Haitham Hassanieh. 2023. WINC: A Wireless IoT Network for Multi-Noise Source Cancellation. In *The 22nd International Conference on Information Processing in Sensor Networks (IPSN '23), May 09–12, 2023, San Antonio, TX, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3583120.3586964

IPSN '23, May 09–12, 2023, San Antonio, TX, USA

ACM ISBN 979-8-4007-0118-4/23/05…$15.00
https://doi.org/10.1145/3583120.3586964

1 INTRODUCTION

Noise cancellation is becoming ever more necessary as our work and living environments are becoming saturated with noise pollution. Long-term exposure to high-level ambient noise can lead to numerous health problems like high blood pressure, heart diseases, sleep disorders, stress, as well as cognitive impairment in children [24]. In fact, according to the World Health Organization, noise is the second-largest environmental cause of health problems, just after air pollution [1]. Besides, emerging applications like VR need noise cancellation to provide an acoustically immersive experience.

Today's solution for noise cancellation is to use over-the-ear headphones or tight blocking earphones like the Bose NC700 [6], the Sony WH1000XM4 [26], or the Apple AirPod Max [4]. These headphones use a technique known as ANC (Active Noise Cancellation) where a reference microphone (also known as the feedforward microphone) is placed on the outer shell of the headphone to record the ambient noise, which is then passed through a filter to create an anti-noise signal. This anti-noise sound is played by a speaker inside the headphone and interferes destructively with the ambient noise to cancel it out at the ear drum. This design, however, suffers from two major limitations. First, since the reference microphone is so close to the ear, the digital processor has very limited time before the noise reaches the ear drum ($\approx 30\mu s$ [20]) to compute and generate the proper anti-noise signal, which results in imperfect cancellation at low frequencies and almost no cancellation at high frequencies because they vary quickly.[1]

Besides, a single reference microphone is not able to resolve and cancel noises coming from multiple sources located in different directions, and that is why some latest noise-cancelling headphones like Bose NC700 [6] and Apple AirPod Max [4] are equipped with multiple reference microphones. However, as we prove in Sec. 4.2, simply adding more reference microphones to the headphone is suboptimal. This is because the reference microphones confined in a small area on the headphone capture very similar mixtures of noises and provide little extra information as opposed to a single reference microphone. Therefore, the ANC system is not able to isolate the

[1]Noise-cancelling headphones like Bose, Sony, and Apple deal with this by sealing the ear completely with sound absorbing material.

noise sources and generate anti-noise signals corresponding to each noise source. In order to take full advantage of the spatial diversity of the reference microphones, we move the reference microphones away from the headphone and disperse them in the environment, so the reference mics are far apart from each other. The remote reference mics use a wireless IoT network to forward captured noise signals to the headphone through wireless links.

In addition to improving the spatial diversity, moving reference microphones away from the headphone also allows the ANC system to *look ahead* into the future for the incoming noise samples to predict more accurate anti-noise signals. This is possible because the wireless reference microphones in the environment are closer to the noise sources, and they forward the captured noise samples through wirelss links at the speed of light to the headphone, which is almost a million times faster than the speed of sound. Therefore, the ANC system can access the reference signals much earlier than when the acoustic noise reaches the headphone. We refer to the resulting time difference as the *future lookahead*. It provides the ANC system with much more time to process and generate the anti-noise signal, so as to achieve better cancellation. Past work, MUTE [20], has tried to leverage the lookahead and demonstrated better noise cancellation compared to COTS ANC headphones especially for high-frequency noises. However, MUTE [20] is limited to a single noise source and a single reference microphone. It would not work in practical settings with multiple noise sources.

In this paper, we present WINC (Wireless IoT-based Noise Cancellation) where we scale LANC (Lookahead Active Noise Cancellation) to multiple noise sources by using a network of wireless IoT microphones scattered in the environment to capture ambient noises from various locations and forward them to the noise cancelling headphone. By using M reference microphones WINC is able to cancel M noise signals while fully benefiting from the look-ahead time provided by the wireless relays. This scaling, however, is not as trivial as replicating MUTE [20]. First, each reference microphone does not necessarily capture only one noise source but can capture noise from multiple sources at a time making it harder to disentangle these noise signals. Moreover, depending on the locations of the noise sources and the references mics, different sources will exhibit different look-ahead times. Hence, there is not a single consistent look-ahead time for all noise sources that allows us to know how much extra computation time we have to compute and generate the anti-noise signal. Furthermore, canceling M noise sources requires $M\times$ more computation which increases the processing delay and reduces the advantage of having a look-ahead time. Besides, LANC has drawbacks of increased computation complexity and slower convergence rate of the adaptive filter. Finally, on the networking front, we must ensure that the wireless network does not add any processing or medium access delays that counteract the benefit of having a look-ahead time by using wireless.

To address the above issues and scale to multiple noise sources, WINC tackles the problem of efficiently generating the anti-noise signal in the frequency domain. Processing the anti-noise signal in the frequency domain naturally decomposes the long time-domain anti-noise filter with many parameters into multiple independent sub-problems at different frequencies. This decomposition can significantly speed up the convergence of the iterative anti-noise filter estimation. Besides, common everyday noises do not have constant

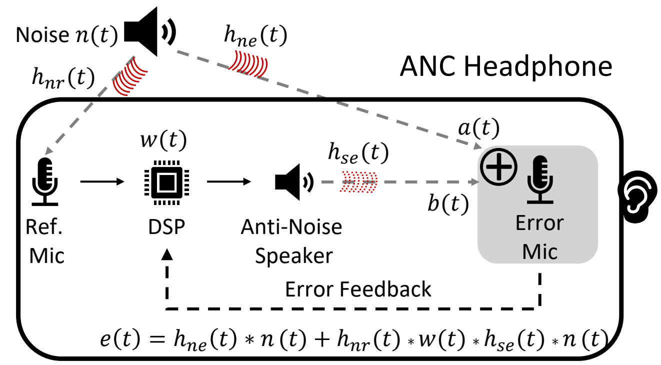


Figure 1: Conventional ANC system overview with one noise source.

energy across all frequencies, and the multiple noise sources can exhibit very different power spectra. Therefore, operating in the frequency domain allows us to adjust the amount of computation on each frequency based on the amount of energy that needs to be canceled. In this way, we can achieve better cancellation with less computation. Last but not least, operating in the frequency domain also enables frequency-selective noise cancellation, that can be used for tuning and customizing the noise cancellation function based on the user's preference and hearing ability (e.g., some users are more sensitive to certain frequencies than others). Although the high-level idea of WINC may seem as simple as a time-to-frequency conversion, translating it into a practical solution requires careful algorithmic and system design which we introduce in detail in section 5.

For the wireless IoT microphone network, WINC avoids digital wireless modulations and protocols like Bluetooth, WiFi, Zigbee, etc, because they introduce latencies for processing and medium access, that can significant reduce the lookahead time budget and defeat the purpose of using wireless IoT relays. Therefore, WINC uses analog frequency modulation (FM) as the wireless modulation scheme for the IoTs, which is the standard for COTS wireless microphones. Besides, different IoT radios are frequency-division multiplexed to allow multiple IoT nodes to transmit and simultaneously forward the reference noise signal.

We built a prototype of WINC and tested it in different environments with up to 3 noise sources. Our results show:

- WINC outperforms the active noise cancellation on Bose headphones by 15dB to 28dB and the overall cancellation by blocking the ear by around 9dB. Note that WINC does not require blocking the ear.
- When cancelling multiple noise sources, WINC outperforms the state-of-the-art prior work MUTE [20] by 16dB to 27dB. Even after extending MUTE [20] to have multiple reference microphones, WINC still achieves 7dB to 14dB better median cancellation on all frequencies and by around 15dB on frequencies above 1000Hz.
- Due to spatial diversity, WINC achieves 7 dB better median cancellation on all frequencies with three noise sources by dispersing three reference microphones with 3-meter range between each, compared to that with 0.5-meter range.

2 PRIMER

2.1 Active Noise Cancellation

A typical active noise cancellation (ANC) systems, as illustrated in Fig. 1, consists four components:

(1) A reference microphone capturing noise signals that are used as feed-forward reference signals in the ANC algorithm.
(2) Digital signal processing (DSP) unit running the ANC algorithm to compute the anti-noise signal.
(3) Anti-noise speaker playing anti-noise signal to cancel the noise signal at the user's ear-drum.
(4) Error microphone monitoring residual noise close to the ear-drum and provide feedback to the ANC algorithm.

2.2 Adaptive Filtering Algorithm for ANC

State-of-the-art ANC headphones today use an adaptive filtering based algorithm for active noise cancellation. We describe the algorithm here with a simple example considering a single noise source and one reference microphone. Suppose noise signal $n(t)$ arrives at the error mic through channel $h_{ne}(t)$, so the received noise signal $a(t)$ at the error mic is:

$$a(t) = n(t) * h_{ne}(t) \tag{1}$$

In order to cancel $a(t)$ at the error mic, the ANC utilizes the signal $r(t)$ recorded by the reference mic M_r, which is the same noise source $n(t)$ but having passed through a different channel $h_{nr}(t)$. Hence, the signal $r(t)$ at the reference mic is

$$r(t) = n(t) * h_{nr}(t) \tag{2}$$

At the DSP, the reference signal $r(t)$ is passed through a filter with filter response $w(t)$, which is then sent to the anti-noise speaker. Next, the anti-noise speaker plays the processed sound $r(t) * w(t)$ and the resultant anti-noise that reaches the human ear will be:

$$\begin{aligned} b(t) &= r(t) * w(t) * h_{se}(t) \\ &= n(t) * h_{nr}(t) * w(t) * h_{se}(t) \end{aligned} \tag{3}$$

Thus, the overall signal received by the error mic placed at the human ear is $e(t) = a(t) + b(t)$. Substituting Eq. 3 and Eq. 1, we get:

$$e(t) = h_{ne}(t) * n(t) + h_{nr}(t) * w(t) * h_{se}(t) * n(t) \tag{4}$$

To get perfect noise cancellation, we need $e(t) = 0$. Hence, the goal is to find a filter response $w(t)$ such that $e(t)$ is as close to 0 as possible. Hence, from Eq. 4, we can see that $w(t)$ should be set to:

$$w_{ideal}(t) = -h_{nr}^{-1}(t) * h_{ne}(t) * h_{se}^{-1}(t) \tag{5}$$

Therefore, for perfect noise cancellation, ANC needs to estimate all 3 channels to obtain to the optimal $w(t)$. However, this is not easy to obtain because: (1) computing the inverse channel requires future samples for precise estimation, and (2) the channels are continuously varying over time so ANC needs to keep re-estimating the channel constantly.

So instead of trying to estimate the precise channels (closed form solution), ANC resorts to an adaptive filtering solution that tries to directly optimize $w(t)$ from the error signal $e(t)$. That is, adaptive filtering tries to find the solution to the optimization problem, defined as:

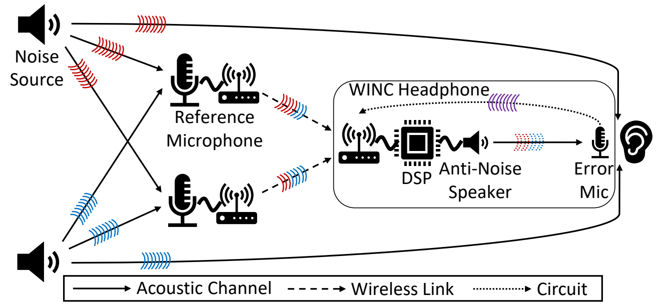


Figure 2: System Overview of WINC's Multi-Noise Source Multi-Reference ANC Over Wireless IoT Networks.

$$\min_{w} \quad e(t)^2 \tag{6}$$

Adaptive filtering techniques use gradient descent optimization tools to find the solution to the above problem, i.e. adjusting the values of the vector $w(t)$ in the direction in which the residual error $e(t)$ reduces the fastest.

$$\begin{aligned} w(\tau) &\leftarrow w(\tau) - \frac{\alpha}{2}\frac{\partial e^2(t)}{\partial w(\tau)} \\ \frac{\partial e^2(t)}{\partial w(\tau)} &= 2e(t)\left(h_{se}(t) * r(t-\tau)\right) \end{aligned} \tag{7}$$

ANC continues the above optimization until convergence and the resulting $w(t)$ obtained is then used for computing the anti-noise signal. This is the basic principle behind the ANC technique used in most commercial devices today. With this background, we will now go deeper into the design of WINC, and explain how we extend it to multiple noise sources, as well as improve its cancellation performance by leveraging adaptive filtering in the frequency domain.

3 SYSTEM OVERVIEW

WINC aims to scale ANC to multiple noise sources in the environment, leveraging a wireless IoT network, as illustrated in Fig. 2. It moves the reference microphones from the ANC headphone (Fig. 1) to multiple wireless IoT-based microphone relays. The IoT-based reference microphones capture noise signals from multiple sources (the blue and red curves) at various locations and forward them to the WINC headphone over the wireless link. The WINC headphone computes and plays the anti-noise signal (the dotted curves) using the anti-noise speaker to cancel the multi-noise source. The residual noise (the purple curves) is recorded by the error microphone and used to update the anti-noise generating parameters. It is important to emphasize that the reference microphones are not necessarily deployed close to any certain noise source, so in general, each reference signal is *a combination of multiple noises*. In this way, we not only increase the number of reference microphones, but also introduce spatial diversity and look-ahead to the reference signals, which are the keys to scaling ANC to multiple noise sources. We describe the spatial diversity and look-ahead in more detail in Sec. 4.2 and Sec. 4.3 respectively.

Apart from these, we also jointly innovate the adaptive filtering-based ANC algorithm and the formulation of the feedback error signal to improve the cancellation performance over multiple different noise sources. Instead of maintaining and updating all time-domain filters using a single time-domain error term, we decompose the error term to multiple orthogonal frequency bins in the frequency domain and update each frequency of the filters separately. We describe our frequency-domain adaptive filtering algorithm and how it leverages the frequency-domain error signal to achieve faster convergence in Sec. 5.

4 NETWORKED MULTI-SOURCE LANC

In this section, we first highlight how multi-noise source multi-reference ANC fundamentally differs from single-noise source single-reference ANC (in Sec. 2.2). We model the optimization problem and provide an ideal as well as an iterative solution to the problem. Following this is the discussion on how WINC leverages the network of microphones to provide channel diversity (Sec. 4.2) and increase lookahead (Sec. 4.3). These are the key contributors to WINC's performance gains over the existing state-of-the-art noise cancelling systems.

4.1 Problem Modeling

Leveraging multiple reference microphones to cancel multiple noise sources is not as trivial as replicating the single-reference ANC algorithm multiple times for each pair of noise source and reference mic. It entails several challenges.

First, each reference mic does not receive signals from only a single noise source. In fact, depending on the deployment of the reference mics and the locations of the noise sources, each reference mic captures a unique combination of noise signals from multiple sources travelling through different acoustic channels. This is known as the "cross-talk" effect.

Moreover, the multiple noise sources are at random unknown locations, so the resulting geometric relations between the noise sources, reference mics, and the user are unpredictable. One cannot assume that every reference signal contains lookahead corresponding to all noise signals. Some noise signals may reach the error mic even before they arrive at some of the reference mics. Every reference signal will also contain different amounts of lookahead with respect to different noise sources, and some of the lookahead can even be negative, or delayed. Therefore, without knowing the acoustic channel between all noise sources and all reference mics, one cannot decide which reference signal to use for canceling which noise signal.

In addition to the reference signals where multiple noise sources are inseparable, the residual noises from all sources also sum up at the error microphone and become entangled. Therefore, one cannot perfectly isolate the error signal components corresponding to each noise source or reference signal either. Hence, the multi-source ANC algorithm has to utilize a common error signal as feedback to update the filters for every reference signal.

Because of the multi-source entanglement in both reference and error signals, we have to jointly cancel all noise signals using all reference signals and a cumulative error signal. Figure 3 illustrates WINC's system model using M reference mics to generate anti-noise

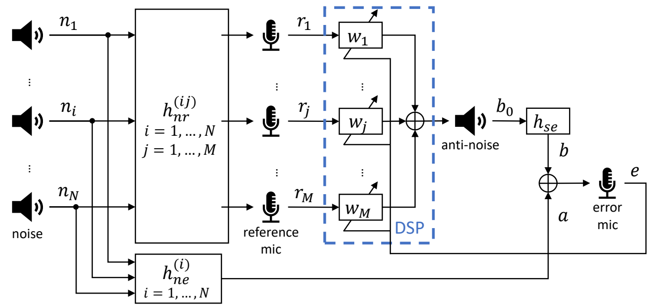


Figure 3: System model with N noise sources and M reference mics. Each reference signal r_j will be processed by a corresponding filter w_j in the DSP, then the resultants are summed up and played by the anti-noise speaker to cancel the noise n_i that directly hits the error mics. The residual error e is used to update the filters w_j.

and cancel N noise sources at the error mic. With each reference signal processed by an ANC filter $w_j(t), j = 1, \dots, M$, the error signal can be modeled as:

$$\begin{aligned} e(t) &= \sum_{i=1}^{N} n_i(t) * h_{ne}^{(i)}(t) + \left(\sum_{j=1}^{M} r_j(t) * w_j(t) \right) * h_{se}(t) \\ &= \sum_{i=1}^{N} n_i(t) * \left(h_{ne}^{(i)}(t) + h_{se}(t) * \sum_{j=1}^{M} h_{nr}^{(ij)}(t) * w_j(t) \right) \end{aligned} \tag{8}$$

The ideal solution to achieve perfect cancellation ($e(t) = 0$) is the solution of the following equation set, which is independent of the unknown noise signals $n_i(t)$.

$$h_{ne}^{(i)}(t) + h_{se}(t) * \sum_{j=1}^{M} h_{nr}^{(ij)}(t) * w_j(t) = 0; \quad i = 1, \dots, N \tag{9}$$

From Eq. 9, we can infer that the optimal filter for reference mic j alone, i.e., $w_j(t)$, will depend on all the channel values between every noise source and every mic in the system. This means that all the ANC filters will jointly cancel all the noise sources, as opposed to every filter being responsible for cancelling a particular noise source independently. However, solving Eq. 9 is impractical since the acoustic channels between the noise source and any microphone (both $h_{ne}^{(i)}(t)$ and $h_{nr}^{(ij)}(t)$) can vary and hence cannot be pre-estimated. Instead, the problem is solved through an optimization formulation as follows:

$$\min_{w_j(t), j=1,\dots,M} \quad e^2(t) \tag{10}$$

Thus, the optimal filters $w_j(t)$ for this multi-source, multi-reference ANC can be found using adaptive filtering algorithm as follows:

$$\begin{aligned} w_j(\tau) &\leftarrow w_j(\tau) - \frac{\alpha}{2} \frac{\partial e^2(t)}{\partial w_j(\tau)} \\ \frac{\partial e^2(t)}{\partial w_j(\tau)} &= 2e(t) \left(h_{se}(t) * r_j(t - \tau) \right) \end{aligned} \tag{11}$$

The iterative solution Eq. 11 is dependent on two variables: (1) $e(t)$ and $r_j(t)$ are captured in real time; and (2) $h_{se}(t)$ can be estimated during the initialization of the system and barely varies.

This forms the basis of WINC's multi-source noise cancellation algorithm.

4.2 Channel Diversity

Although WINC and some COTS ANC headphones with multiple reference mics follow the same multi-noise source multi-reference system model, the locations of reference mics make a huge difference. WINC's reference microphones are dispersed in the environment, and the spatial diversity of the reference mics allow WINC to leverage an opportunity known as the channel diversity. In an environment with multipath reflections, the difference in the channel response is usually closely related to the location of the device. Statistically, the closer the receivers are to each other, the more similar their channel responses will be. Such an effect has been well studied in the wireless MIMO communication area, where spatial multiplexing of multiple streams highly relies on the channel difference between pair transmitter and receiver antennas. This is the reason why in MIMO theory, the farther apart we spread the transmitter or receiver antennas, the better performance we achieve due to the spatial diversity gain. This phenomenon also carries over to the acoustic modality.

Eq. 9 defines the condition for the multi-source ANC system to achieve perfect cancellation. As a linear equation set, Eq. 9 has a solution (i.e., consistent) when the number of independent equations is fewer than or equal to the number of independent variables. Otherwise, Eq. 9 is inconsistent and has no solution, which means that we may not find a set of filters that can perfectly cancel any noise signals. In this model, the number of independent equations/variables are decided by the following factors:

- The number of independent equations is determined by the locations of the noise sources. Noise sources at the same location can be treated as one as they have identical channels $h_{ne}^{(i)}$ and $h_{nr}^{(ij)}, \forall j$. In this case, the number of independent equations reduces and the problem becomes easier to solve.
- The number of independent variables is determined by the locations of reference mics. Reference mics at the same location have the same channel $h_{nr}^{(ij)}, \forall i$, and hence, can also be combined. In this case, the system reduces to one with fewer reference mics, and the problem becomes harder to solve because of fewer independent variables. Similarly, partially dependent variables (close-by reference mics with similar channel) will degrade the convergence of the adaptive filtering as well, particularly under the influence of hardware noise. This is what we refer to as channel diversity.
- The consistency of Eq. 9 also depends on the channel between the anti-noise speaker and the error mic h_{se}. However, in most applications, the anti-noise speaker and error mic are at fixed locations in the headphone, resulting in a fixed h_{se}. Therefore, we lessen its influence here.

Conventional ANC headphones cannot fully leverage the potential spatial diversity gains from multiple reference mics, because the location of the on-headphone reference mics are confined by the size of the headphone, which is only around 20 cm (half wavelength of 850 Hz). The confined size of these headphones largely limits the channel diversity gain we can potentially achieve by adding more mics. On the contrary, WINC has the freedom to disperse reference mics all around the environment, which can provide much better channel diversity for the ANC system and improve performance of noise cancellation with multiple noise sources.

4.3 Lookahead

In addition to the channel diversity gain, the wireless reference microphones provide a peek into the noise signal that enables the ANC system *look-ahead* to the samples much earlier than when they reach the user. Lookahead-aware ANC (LANC) systems have two-fold benefit: (i) extra time budget for processing the anti-noise signal, (ii) better estimation of the non-causal inverse-channel in the adaptive filter (shown in Eq. 5). Therefore, LANC can achieve much better cancellation than conventional ANC headphones [20].

By deploying the reference mics on wireless IoT relays instead of the outer shell of the ANC headphone, WINC is able to move some of the reference mics closer to the noise sources, which reduces the propagation time between these two. Then the IoT relays equipped with wireless transmitters forward the reference signals to the ANC headphone over a wireless link with minimum delay. Because wireless (RF) signals propagate much faster than acoustic signals in air, the relayed reference signals arrive much earlier in time to the DSP unit on the ANC headphone compared to the conventional design on the ANC headphones. This provides the ANC algorithm with significantly more time to process and compute the anti-noise signal, before the corresponding noise hits the eardrum.

Lookahead can also improve the noise cancellation performance by allowing the ANC algorithm to utilize future noise signal samples to estimate the non-causal inverse channel. The reason can be explained from the perspective of either the ideal solution or the iterative solution. On one hand, as can be seen from Eq. 5, the ideal anti-noise generating filter $w(t)$ should be an inverse of the channel $h_{se}(t)$ and $h_{nr}(t)$, which means $w(t)$ is non-causal ($w(\tau) \neq 0, \exists \tau < 0$) with large probability. So, the ideal anti-signal should be dependent not only on past noise signals, but also on future ones. On the other, from the iterative solution in Sec. 2, we can see that Eq. 7 contains the term $r(t-\tau)$ in the gradient computation, implying that the update step in the gradient descent optimization requires future samples of the reference signal $r(t-\tau)$, $(\tau < 0)$. Hence, as we see more future reference samples, the more accurate non-causal filter $w(t)$ we can estimate. In conclusion, the cancellation performance is largely decided by how many "future" samples the reference mics can capture, or in other words, how much lookahead can we obtain from the wireless IoT relays.

5 FREQUENCY-DOMAIN ADAPTIVE FILTERING

Although noise cancellation using wireless IoT provides many-fold benefits, it is a double-edged sword. The improved noise cancellation performance is at the cost of more computation complexity and slower convergence rate of the adaptive filtering algorithm. In this section, we explain these unique challenges caused by the lookahead in the reference signal and present our solution to this challenge through the frequency-domain adaptive filtering optimization technique.

5.1 Convergence Challenge of LANC

According to Sec. 4.1, to jointly cancel N noise sources using M reference signals, WINC's adaptive filtering needs to update $M \times L$ parameters, where L is the number of taps in each anti-noise filter. As shown in Eq. 5, the ideal anti-noise filter ANC tries to estimate is $h_{nr}^{-1} * h_{ne} * h_{se}^{-1}$. Given that the channel from the anti-noise speaker to the error microphone h_{se} is fixed, h_{se}^{-1} is constant. Therefore, the number of anti-noise filter taps L is determined by $h_{nr}^{-1} * h_{ne}$, in other words, the difference between h_{nr} and h_{ne}. With more lookahead in the reference signal, the distance between h_{nr} and h_{ne} becomes larger, which leads to a larger number of taps L in the anti-noise filter.

To explain this more clearly, let us consider a simple example where every acoustic channel is just a simple delay channel (no attenuation or multipath), i.e., $h(t) = \delta(t - \tau)$ where $\delta(t)$ is a unit impulse and τ represents the propagation delay between the noise source and the microphone. For more simplicity, we can assume that channel between the anti-noise speaker and the error mic is $h_{se}(t) = \delta(t)$. Thus, for single-noise-single-reference scenario, the ideal anti-noise filter can be derived using Eq. 5:

$$\begin{aligned} w_{ideal}(t) &= -h_{nr}^{-1}(t) * h_{ne}(t) * h_{se}^{-1}(t) \\ &= -\delta^{-1}(t - \tau_{nr}) * \delta(t - \tau_{ne}) * \delta^{-1}(t) \\ &= -\delta(t - (\tau_{ne} - \tau_{nr})) \end{aligned} \tag{12}$$

We define $\tau_\Delta \triangleq \tau_{ne} - \tau_{nr}$ and assume $\tau_\Delta \geq 0$. [2] Notice that the length L of the anti-noise filter $w(t)$ has to be at least longer than τ_Δ so that the propagation delay induced by the channel can be captured accurately and hence the anti-noise be generated properly to cancel the environmental noise. Because the locations of the noise sources are unknown in practice, the ANC system has to use an anti-noise filter that is long enough to cover any possible value of τ_Δ. Therefore, the minimum L is equal to the *acoustic propagation delay between the reference mic and the error mic*, which corresponds to the condition when the noise source, the reference mic, and the error mic lie on the same line.

Compared to conventional ANC headphones, where the reference microphone and error microphone are very close to each other, the distance between the reference mic and the error mic in a LANC systems is much longer. Therefore, the anti-noise filters for LANC systems also need a lot more taps than those of ANC headphones to achieve reasonable cancellation performance.

Past work like MUTE [20] have to cope with this increased computational burden to leverage the benefits of look-ahead. However, MUTE is limited to a single noise source and reference mic, so it only needs to estimate one anti-noise filter with L taps. Issues arise when trying to scale such a system to cancel multiple noise sources at a time, since we now need to add multiple reference mics $(1, \ldots, M)$, and in turn our optimization problem now needs to optimize for $M \times L$ parameters.

Moreover, the large amount of variables are strongly dependent on each other. Although we scale the number of reference mics to empower channel diversity, the system is still limited to a single error mic on the headphone and hence, single feedback signal. As shown in Eq. 11, at every time step, a single error mic sample is used to update all M anti-noise filters with L taps. It is direct to see that the gradient of $w_j(t)$ is a function of the error signal $e(t)$, which is a resultant of all antinoise filters (Eq. 8). Trying to solve the entire large optimization problem with all $M \times L$ parameters would be extremely unstable from an optimization point of view, and getting the gradient descent to converge to the optimal solution is extremely hard when the problem has a large amount of interdependent variables.

Hence, although time-domain adaptive filtering can work reasonably well for ANC headphones and single-source-single-reference LANC (e.g., MUTE), the time-domain approach cannot scale for multi-noise source multi-reference LANC due to the long convergence time.

[2]Here, the assumption is required when all channels are simple delay channel. Otherwise, the reference mic barely contributes to the cancellation of this noise source, since it receives the noise later than the error mic. But in practice, such reference mics are also useful considering multipath effect, and adaptive filter can automatically assign suitable filter values.

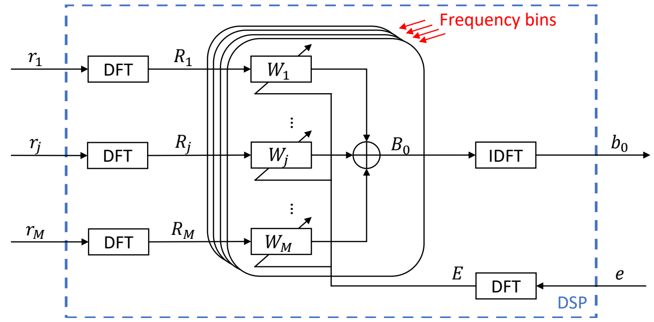


Figure 4: Detailed DSP overview with frequency domain algorithm. The algorithm first performs FFT of all input signals $r_j(t)$ and $e(t)$, which is processed by the frequency domain filter $W_j(f)$ to construct the anti-noise signal. After this, the anti-noise $B_0(f)$ is converted back to time domain $b_0(t)$ and sent out by the anti-noise speaker.

5.2 Frequency Domain Optimization: Independent Sub-Problems

To overcome the convergence challenge of multi-source multi-reference LANC, we try to decompose the large optimization problem into multiple sub-problems with fewer parameters each. Besides, we incorporate signal processing knowledge into this optimization decomposition to come up with a frequency-domain adaptive filtering solution. Decomposition is a popular strategy in solving optimization problems, and a good division of optimization problems should meet the requirements that: (1) the sub problems are independent and can be solved individually; (2) the solution to the original problem can be constructed from the solutions to each sub problem.

The natural orthogonality property of DFT helps decompose the optimization problem. WINC uses DFT to convert time-domain convolutions into frequency-domain multiplications and thus achieves the independence of the parameters in each frequency bin. Operating in the frequency domain allows us to parallelize the adaptive filtering algorithm since we can optimize the error corresponding to each frequency bin separately.

In the frequency domain, the error at each frequency bin f can be represented as:

$$E(f) = \sum_{i=1}^{N} N_j(f) H_{ne}^{(i)}(f) + \sum_{j=1}^{M} R_j(f) H_{se}(f) W_j(f) \quad \text{for each frequency } f \tag{13}$$

The optimization of the sub-problem for each frequency bin can then be formulated as:

$$\min_{W_j(f), j=1,\ldots,M} |E(f)|^2 \tag{14}$$

The adaptive filter taps can now be calculated in parallel at each frequency to minimize the error at that frequency.

$$W_j(f) \leftarrow W_j(f) - \alpha_j E^H(f) R_j(f) H_{se}(f) \tag{15}$$

WINC divides the time-domain optimization problem Eq. 10 (with ML variables) into L independent frequency-domain problems Eq. 14 (with M variables each). This greatly reduces the number of interacting taps from ML to M, thus making the problem parallelizable and improving the convergence of adaptive filtering.

It is proved that Eq. 14 keeps a convex loss function (see proof in Appx. A), so global optimum is guaranteed and is able to reconstruct the optimum of Eq. 10. Moreover, since the proof does not make any assumption of the channels between any noise source and any mic, the convexity holds when any source or any mic moves. Hence, the frequency domain adaptive filtering is feasible in dynamic environments.

5.3 Power Adaptive Step Size

Converting the time domain optimization to frequency domain untangles the dependency between the anti-noise filter taps and divides the optimization into L independent sub-problems with M variables each. The reduced dependence can be further leveraged to improve the convergence rate of the adaptive filter by assigning each filter tap (corresponding to a frequency bin) a different step size. To also keep it robust to sudden variations in signal power at a particular frequency, these independent step sizes should be adaptive. Thus, in WINC we use adaptive step size for each frequency:

$$W_j(f) \leftarrow W_j(f) - \mu_j(f) E^H(f) R_j(f) H_{se}(f) \tag{16}$$

Here $\mu_j(f)$ is the adaptive step size for j^{th} filter, and is given by:

$$\mu_j(f) = \frac{\alpha_j(f)}{P_j(f)} \tag{17}$$

In Eq. 17, $\alpha_j(f)$ is a constant and $P_j(f)$ is the measure of mean power in this specific frequency bin of the reference signal. The core intuition behind scaling the step size inversely relative to the reference signal power is that when there is a sudden variation in the input reference signal, the fixed step size may cause a large variation in the anti-noise filter and thus a sudden increase in error, which may lead to an unstable state in the sub-optimization. On the other-hand, for frequencies with lower power, this adaptive step size leads to faster convergence. Figure 8b in Sec. 7.3 demonstrates the gains in convergence rate that we achieve in WINC with this power adaptive step size.

Algorithm 1 WINC: Wireless IoT-based Noise Cancellation

Initialization Phase:
Play n_s at anti-noise speaker.
$N_s \leftarrow FFT\{n_s\}$ ▷ All FFT are on the latest L samples
Record e_s at error mic.
$E_s \leftarrow FFT\{e_s\}$
for $f = 1, f \le L, f{+}{+}$ **do**
  $H_{se}(f) \leftarrow \frac{E_s(f)}{N_s(f)}$
  for $j = 1, j \le M, j{+}{+}$ **do**
    $W_j(f) \leftarrow 0$
  end for
end for
Noise Cancellation Phase:
while True **do**
  Record the error $e(t)$ at error mic.
  Record the reference $r_j(t)$ at M reference mics.
  $E \leftarrow FFT\{e\}$
  for $f = 1, f \le L, f{+}{+}$ **do**
    for $j = 1, j \le M, j{+}{+}$ **do**
      $R_j \leftarrow FFT\{r_j\}$
      $P_j(f) \leftarrow R_j^H(f) R_j(f)$
      $\mu_j(f) \leftarrow \frac{\alpha_j(f)}{P_j(f)}$
      $W_j(f) \leftarrow W_j(f) - \mu_j(f) E_j^H(f) R_j(f) H_{se}(f)$
    end for
    $B_0(f) \leftarrow \sum_{j=1}^{M} W_j(f) R_j(f)$
  end for
  $b_0 \leftarrow IFFT\{B_0\}$
  Play $b_0(t)$ at anti-noise speaker.
  $t \leftarrow t + 1$
end while

5.4 Bridging Time Domain Measurements to Frequency Domain Adaptive Filtering

The workflow of WINC is divided into two phases: initialization and noise cancellation. In the initialization phase, the anti-noise speaker plays a training sequence $s[n]$, while the error mic records the sound to compute the channel h_{se}. The following error needs to be minimized in order to obtain a good estimate of h_{se}:

$$\begin{aligned} error[n] &= e[n] + h'_{se}[n] * s[n] \\ &= h_{se}[n] * s[n] + h'_{se}[n] * s[n] \end{aligned} \tag{18}$$

$e[n]$ is the error mic sample at the n^{th} instant. The training sequence $s[n]$ is a known random noise, and h_{se} is an adaptive filter that uses LMS to adjust its values and thereby minimize the error. Therefore, as the estimate h'_{se} of the antinoise speaker to error mic channel approaches close to its ideal value h_{se}, $error[n]$ becomes negligible. This helps WINC estimate the antinoise speaker to error mic channel at a wide band of frequencies used during our experiments (1-3kHz). The known random sequence is played for 10 seconds. to let the error converge and filter values remain stable for sufficient time. Such computation is only an one-time operation, since the channel remains almost the same as a resultant as a person's ear shape. Due to the convexity of the problem, all anti-noise filter taps are initialized as zeros for simplicity.

In the noise cancellation phase (Fig. 4), WINC first computes DFT over the new mic sample and the latest $L-1$ ($L >= 1024$) samples of each signal, which will be used to update the anti-noise filter values with gradient Eq. 15. Then, the reference signals are passed through the anti-noise filters and summed up to get each frequency component of the anti-noise. Finally, IDFT is performed to get the time-domain anti-noise signal, which is played by the anti-noise speaker. The loop repeats for the next set of signals.

Although WINC requires computing frequency components for every new sample, the computation remains the same as the time domain algorithm. Typically, DFT can be implemented in the sliding window manner (e.g. STFT), which uses the result in the previous window to generate the new one. Besides, IDFT only needs to compute the last sample. Thus, the complexity remains as $O(L)$ per anti-noise sample, which is no more than the time-domain adaptive filtering. A detailed sample-by-sample pseudocode is shown in Alg. 1, and WINC can also be implemented in a block-by-block manner keeping the same computational complexity per sample.

6 SYSTEM IMPLEMENTATION

Wireless Relay: Similar to most wireless microphone systems, WINC's wireless relays use analog FM modulation, even though compared to digital modulations analog FM suffers more from distortions. This is because digital modulations require sampling the audio signal at the IoT relays using Analog-to-Digital converters (ADCs) and then converting the digital audio data back to analog acoustic signals at the receiver of the WINC processor. This additional round of analog-to-digital and digital-to-analog conversions introduces significant latency, reducing the amount of lookahead we can obtain from the wireless relaying. Multiple wireless IoT relays are frequency-division multiplexed (FDM) on different frequency channels, so they can simultaneously forward the reference signals to the headphone. Note that the number of FDM channels only depends on the number of IoT relays, because all users can subscribe to all the FDM channels and receive from all reference microphones. In our experimental setup, we implement the wireless IoT relays and the wireless receiver on the WINC headphone using two sets of Gem Sound wireless microphone systems [27].

WINC Headphone: Figure 5 shows our experimental hardware setup for the WINC headphone, which consists of a wireless receiver module, a DSP unit along with the anti-noise speaker and error microphones. The wireless receiver module demodulates the reference signals forwarded by the IoT relays over the wireless channel. We use TI K2G DSP + ARM processor [12] evaluation module with an audio daughter card. TI K2G supports simultaneous multi-channel audio input and output, with up to 8 analog input channels and 16 analog output channels, which is lacking in most off-the-shelf development DSP boards. In our experiments, the K2G processor takes three analog reference audio signals and the error signals as inputs and outputs the anti-noise audio signals.

Limitations of our hardware: A key challenge in enabling multi-source cancellation with multiple reference microphones is the processing capability of the DSP unit on the headphone. It is required to sample all the reference signals from the reference microphones, estimate the adaptive filters associated with each reference mic accurately, and generate the antinoise signal within the time budget provided by the lookahead. To sample reference signals, the ADCs/DACs should be able to support sampling frequencies of up to 44kHz. To generate accurate antinoise signal, we use single-precision floating point format for the adaptive filter weights. Therefore, the DSP should be equipped with memory that can support 32-bit data. Most importantly, the DSP has to be able to continuously update an adaptive filter with at least 1000 taps for every reference microphone, because a typical acoustic channel in a room has many delayed echoes and requires at least 1000 taps for suitable accuracy. However, the general-purpose ARM-based DSP board we use in our experimental setup is not optimized for audio applications and runs a real-time OS to support multiple other tasks and routines onboard. As a consequence, we were limited to a total of 600 filter taps at a sampling rate of 6.3kHz, which is further divided among multiple adaptive filters. This huge gap in the number of filter taps leads to degradation in cancellation performance.

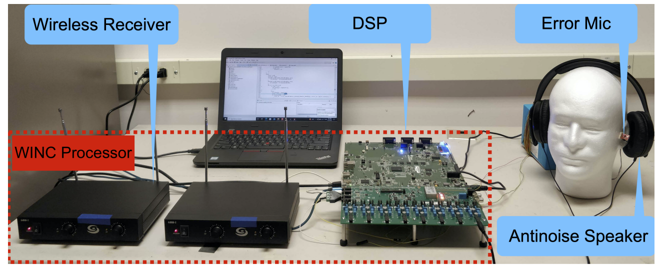


Figure 5: Hardware setup of WINC headphone

Emulating Hardware: Due to computational limitation of our experimental hardware, we evaluated our system using emulations where we are not limited by the hardware computation capability or delays introduced by slow hardware and are able to thoroughly evaluate the performance of our system. We carefully design our emulations to keep them as close to real-time experiments as possible. Here, we synchronously record the reference and error mic signals. However, instead of performing cancellation in real time, we save these signals and run our cancellation algorithm offline. Note that the reference and error mic recordings also encode real-world channels. The only difference is the absence of antinoise speaker to error mic channel (h_{se}). We also account for this channel in our emulations by estimating h_{se} online and convolving the anti-noise samples with this channel in the emulations. Also observe that the error microphone samples cannot be used as is to update the adaptive filters in emulations. In real-time noise cancellation, the error microphone hears residual as the noise and antinoise signals interfere destructively. Therefore, in the emulations, the resultant error e' is calculated by summing up the recorded error mic samples and the estimated antinoise samples. To sum up, the offline algorithm first performs FFT of all recorded reference mic signals $r_j(t)$ which is processed by the frequency domain filters $W_j(f)$ to construct the anti-noise signal. After this, the anti-noise $B_0(f)$ is converted back to time domain $b_0(t)$ and convolved with channel h_{se}. The resultant anti-noise signal b_0' is summed up with the recorded error mic signal e(t) to give the effective error e' which is used to derive the new estimate of adaptive filters. Similar to online processing on a DSP (Fig. 4), this emulation runs for each time step t.

The results presented in the next section are obtained through these emulated experiments.

7 EXPERIMENTAL EVALUATION

In this section, we first explain our evaluation metrics and baselines, and then we present the main results that compare the overall noise cancellation performance of WINC against some state-of-the-art baselines. After that, we demonstrate some micro-benchmarks results that verify and highlight each contributor of WINC's improved noise cancellation performance. Finally, we show an extended function of WINC due to the use of frequency-domain adaptive filtering.

7.1 Evaluation Design

We conduct controlled experiments in indoor environments using portable speakers placed at various locations to play pre-recorded noise signals. This allows us to control the number and locations of the noise sources, as well as the noise signal parameters such as frequency spectrum, volume, etc. Specifically, the performance of WINC has been tested in two different kinds of environments - a conference room and a general laboratory space. We varied the distance of the reference microphones with respect to the error microphones by placing them 50cm, 1.5m and 3m away across different experiments to test performance with different levels of spatial diversity. Lastly, we also vary the location of the noise sources in our experiments to vary the acoustic channels. Each noise signal always has at least one reference microphone that provides a positive lookahead. We test WINC in the presence of up to 3 noise sources, which is typical in indoor spaces. The noise sources and all microphones remain stationary throughout the experiment.

Noise Signals: We use three different types of noise signals to evaluate the system.

- *Gaussian*: We primarily use white Gaussian noise (WGN) signals covering the entire spectrum from 0 Hz to 3 kHz. These frequencies are well within the range of everyday sounds such as low rumbling from motorized devices to high pitched bird chirps from outside. Moreover, we wish to evaluate our system for wideband noise since it is the hardest to design ANC systems that do so.
- *Human Voices*: In addition to WGN noise, we also use recordings of most commonly found real-world noises as the noise signals. The first type of such real-world noise signal is human voices, for which we have separated audio clip for male and female talking.
- *Construction Site Noise: The other type of commonly found everyday noise signal that we use in our experiments are construction site noises from heavy machinery.*

Baselines: We compare WINC against four baselines:

- *COTS ANC Headphone Overall*: We compare WINC against one of the most popular commercial off-the-shelf (COTS) noise canceling headphone - Bose QC 35. It follows the standard ANC system architecture with a single reference microphone on the outer shell of the headphone.
- *COTS ANC Headphone without Passive Blockage*: Although COTS headphones can provide reasonably good noise cancellation, it heavily relies on the passive isolation of the earcups. To focus on the ANC algorithm performance, we try to decouple the contribution of the passive isolation. To do so, we measure the noise cancellation of the COTS headphone twice with the ANC function turned on and off (only passive isolation). Then we compute the cancellation difference, which provides us with the cancellation performance of ANC algorithm alone in COTS headphones.
- *MUTE [20]*: We also compare our system against state-of-the-art LANC prior work, i.e., MUTE [20], which uses the standard time-domain adaptive filtering solution. Specifically, MUTE has only one reference microphone.
- *WINC (time domain)*: We also evaluate the performance difference due to the time-domain and frequency-domain adaptive filtering algorithm. Therefore, we skip DFT/IDFT in WINC, and use time-domain adaptive filtering algorithm for anti-noise filter update. We refer to this baseline as WINC (time domain).

Metrics: We use the following evaluation metrics throughout our experiments unless explicitly stated otherwise.

- *Noise Cancellation Ratio (in dB)*: We use the noise cancellation ratio as the main evaluation metric, because it's consistent across all kinds of noises and scenarios. We extract the cancellation ratio by differentiating the noise signal power at the error microphone emulating user's eardrum with and without the cancellation method under test, using the following equation:

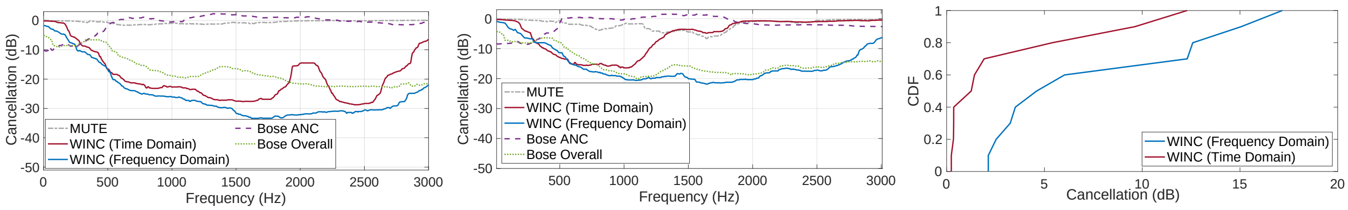
$$R_{\text{Cancellation}} = 20 * log_{10}\left(\frac{|\text{error signal}_{\text{with cancellation}}|}{|\text{error signal}_{\text{without cancellation}}|}\right)$$

- *Cancellation Across Different Frequencies*: In addition to the overall cancellation ratio, we are also interested in the cancellation ratio at different frequencies. Therefore, we perform spectral analysis of the original and suppressed noise noises by applying short-time Fourier transform, and then compute the cancellation ratio for each frequency bin. We also compute the median over different time windows.

7.2 Main Results

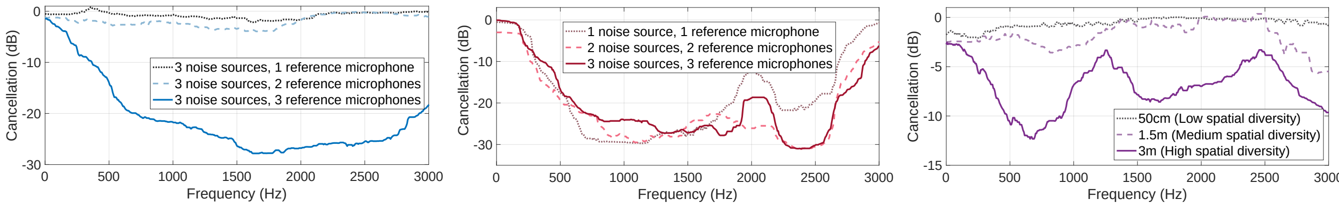
To evaluate the multi-source noise cancellation performance of our system, we create an environment with three noise sources at random locations, and we also deploy three IoT relays to cancel the multi-source noise. We use WINC and the three baseline methods to cancel the same pre-recorded noises and compare the results.

In the first set of experiments, we use wideband white Gaussian noise as the noise signals, which helps us understand the performance of our system across a wide range of frequencies. The results are shown in Fig. 6a. First, one can see that median cancellation offered by WINC outperforms the COTS ANC baseline by ~9 dB across all frequencies, due to the spatial diversity and limited lookahead. The performance gap is more significant when we decouple the passive isolation, as the ANC function of COTS headphone alone is ~28 dB worse than WINC without the passive



(a) (b) (c)

Figure 6: Cancellation of three noise sources. (a) Wideband Gaussian Noise; (b) Real-world Noise (c) Overall performance.



(a) (b) (c)

Figure 7: (a) Cancellation of three noise sources with increasing # of reference mics. (b) Scaling # of reference microphones with # of sources. (c) Spatial diversity in IoT relays improves cancellation performance.

isolation. This results confirm that COTS ANC headphones heavily rely on the passive isolation to cancel noises, especially those in the higher frequencies to make up for the limited lookahead. Moreover, frequency-domain WINC performs ~7 dB better on than time-domain WINC and the gain is more obvious in the higher frequencies, where WINC can achieve as high as ~15 dB more cancellation than time-domain WINC. This is because the advantages of WINC's frequency-domain adaptive filtering are especially pronounced at higher frequencies, since inaccurate filter weights will lead harm the higher frequencies which are very sensitive to even small phase offsets.

In the second set of experiments, we compare the performance of our system against baselines in canceling real-world noise signals, including female and male talking, as well as construction site noises. Figure 6b shows the WINC can achieve ~2 dB, ~15 dB, and ~14 dB more average cancellation than the COTS overall, ANC-alone COTS, and time-domain WINC baselines respectively. These results echo the same observation that we see in case of wideband Gaussian noises.

It is worth noting that our system is implemented without any passive isolation, and we expect the performance of WINC to be even much better when combining passive isolation and WINC's frequency-domain adaptive filtering.

Finally, we also test our system in different locations, microphone setups and different noise sources to measure our performance in more diverse scenarios. The cumulative results of cancellation we get with time-domain WINC versus frequency-domain WINC are presented in Fig. 6c, where we show the CDFs of the median cancellation ratio. It can be seen that frequency-domain WINC achieves a 3.5 dB higher median cancellation than time-domain WINC.

7.3 Microbenchmarks

We also design the following micro-benchmarks to demonstrate some key aspects of WINC.

A. Fewer References than Sources: To verify if multi-source ANC benefits from the number of microphones, we conduct experiments in the presence of 3 noise sources and vary the number of microphones from 1 to 3. Figure 7a shows that increasing the number of reference microphones can exponentially improve the cancellation performance. Such improvement is not obvious when the number of reference mics is fewer than the number of noise sources (1 and 2 reference mics), but it shows a great leap when the reference mics can provide enough channel diversity (3 reference mics) to separate individual noise signals from their mixture.

B. Equal References and Sources: Next, we conduct experiment where the number of wireless IoT relays varies along with the number of noise sources. Figure 7b compares the performance of WINC cancelling 1/2/3 of noise sources with the same number of reference mics. In all cases, WINC is able to separate independent noise signals with adequate reference mics, which theoretically could achieve perfect cancellation considering no hardware noise. Thus, the three plots indicate comparable performance.

C. Channel Diversity of Wireless IoT Relays: To quantify the benefits of spreading out the reference mics (more channel diversity), we evaluate WINC's performance in a three-source-three-reference scenario with varying distance among reference mics. Figure 7c demonstrates the cancellation across frequencies when the three reference mics are 0.5/1.5/3 meters away from each other. Cancellation in 1.5-meter case shows minor improvement over that in 0.5-meter case, which indicates that the mic separation is still not enough to fully leverage the advantage of 3 reference mics. But when the distance further grows up to 3 meters, WINC is able to achieve a median of 7 dB across all frequencies cancelling wide-band Gaussian noise.

D. Cancellation with different number of taps:
As we have discussed in Sec. 5, the drawback of LANC is the need to increase the number of adaptive filter taps, which makes time-domain adaptive filtering hard to converge. To verify the necessity

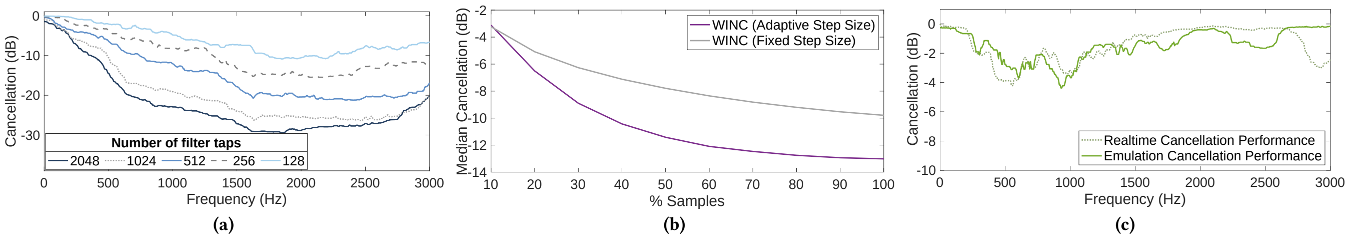


Figure 8: (a) Multi-source noise cancellation with an increasing # of adaptive filter taps. (b) Effect of Frequency dependent adaptation rate. (c) DSP vs Emulated Performance with same number of taps.

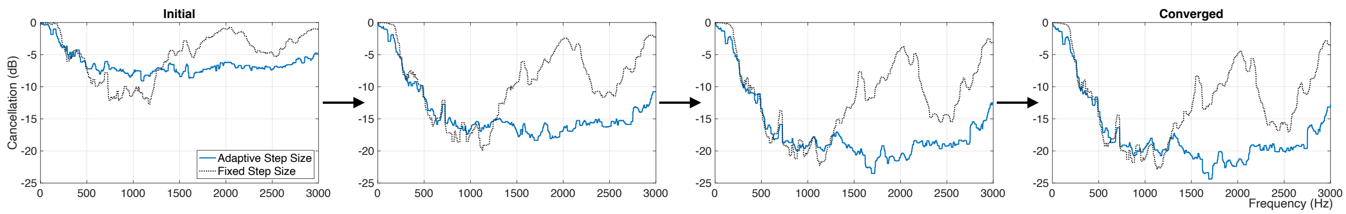


Figure 9: Cancellation of different frequencies over time using adaptive and fixed step size schemes in frequency domain cancellation.

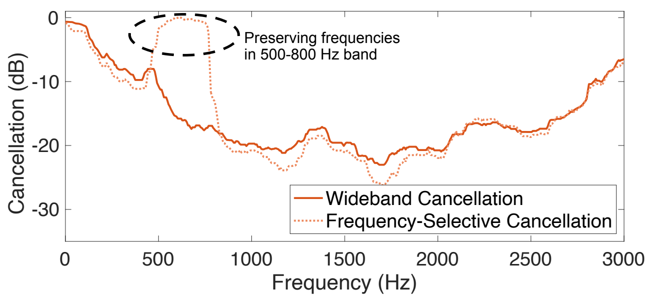


Figure 10: Frequency-Selective Noise Cancellation.

of having sufficient adaptive filter taps, we compare the best cancellation achievable with a different number of adaptive filter taps in Fig. 8a. The cancellation of wide-band Gaussian noise obviously increases as the number of anti-noise filter taps increases from 128 to 2048. Each doubling of the filter taps leads to a 3-5 dB median improvement across all frequencies. It seems that such benefit has not yet reached its limit with longer anti-noise filter.

E. Convergence with adaptive step size: Section 5.3 describes a power adaptive step-size for each independent frequency bin of a filter. To demonstrate the benefit of this method, we compare its performance with a fixed-step-size version. Figure 8b shows the overall cancellation of the two step sizes as a function of time. The final cancellation achieved using the power-adaptive step size is ~4 dB better compared to that achieved using a fixed step size. The advantage also can be seen in the convergence rate, where the solution converges faster to around the optimum with varying step size. More details are provided in the time trace of frequency-related performance. Figure 9 shows that the fixed step size may not be suitable for the noise above 1500 Hz, while adaptive step size can achieve a much more uniform noise reduction across a wide band.

F. Offline Emulation and Real-Time DSP Comparison: Finally, we also verify that our offline emulation can faithfully represent the cancellation performance on the real-time DSP board, so every result can be replicated with a low-overhead DSP board. Figure 8c shows that when using 300 anti-noise filter taps, real-time DSP and offline emulation achieve similar cancellation performance across the spectrum.

7.4 Frequency Selective Noise Cancellation

In addition to all the benefits we've demonstrated throughout the paper, frequency-domain adaptive filtering also allows for frequency-selective noise cancellation. Users can selectively preserve certain frequencies while suppressing all the others. This selectivity can be convenient in everyday lives, for instance, users would like to be aware if a distant fire alarm goes off in the building. Moreover, because not everyone are equally sensitive to all frequencies, privately customized WINC system is able to provide unique user experience under various scenarios. Figure 10 shows that by setting the step size for frequencies between 500-800Hz to 0, we can prevent canceling them out. Thus, WINC can pause updating the anti-noise filter for a certain frequency when the noise volume is below a threshold (e.g. human hearing sensitivity curve), which reserves computation resources for other purposes.

8 RELATED WORK

The literature in active noise cancellation is extremely rich with various system implementations and application [3, 5, 8, 13–15, 20, 25, 28, 30]. In this section, we will cover the ones that are most related to this work in the following aspects: ANC for multiple noise sources and convergence of adaptive filtering in ANC.

ANC for multiple noise sources: Several works have explored different ideas on systems employing multiple reference microphones to cope with multiple noise sources. Some of the older systems such as [16] require pre-estimating the channels between noise sources and reference microphones and are not suitable to dynamic environment. A common lacking in the more recent systems is that unlike WINC, their system architecture does not leverage lookahead and spatial diversity. For instance, in [8, 10, 17, 28] the reference microphones are placed close to the user which is similar to standard commercial noise canceling headphones like Bose QC. They also lack the benefits of using the frequency domain adaptive filtering algorithm.

A parallel body of work on multi-noise source cancellation using multiple reference microphones focuses on selecting the appropriate reference signals out of the many obtained through different mics. [10] deploys multiple microphones around the user and use the time of arrival to select the best reference signal for cancellation. In [2], on the other hand, the system removes irrelevant reference signals based on power spectrum density. More recently, a coherence based method has been applied to select appropriate reference signals [21, 23]. We believe these reference signal selection algorithms are complementary to our system and can be applied prior to the frequency domain adaptive filtering. By selecting only a subset of reference signals to be used as input for the noise cancellation algorithm, the number of antinoise filters to be optimized can be reduced, thereby diminishing computational burden on the system. Thus, this technique can help WINC to selectively cancel the dominant noise sources in the environment.

Convergence of adaptive filtering in ANC: Various work have been done to improve convergence of adaptive filtering in ANC from multiple aspects, such as modifying the optimization problem [18, 19], selecting appropriate adaptive step [9, 11], etc. More relevant work is to improve the convergence of ANC by error signal separation. A theoretical analysis in [7] revealed the slow convergence using one error sample to update all ANC filter taps. Thus, it uses band-pass filter to split the narrowband noise signals of pre-known frequencies. [29] introduces cascaded adaptive filtering to remove unwanted disturbance from the error signal, and [22] modified the cascaded structure to separate the error signal into the ones corresponding to each reference signal. However, these methods add an extra adaptive filters for noise separation, and the required number of adaptive filters is 2x the number of reference microphones. WINC differs from prior work in that it uses adaptive step sizes for the frequency domain filter taps to achieve uniform cancellation over a wide band of frequencies.

9 DISCUSSION & LIMITATIONS

In this section, we discuss various aspects of WINC and also address limitations in our current system design.

Scaling WINC: When a new dominant noise source is added, WINC requires an additional reference microphone to get the best cancellation performance, as we demonstrate in our experiments (Fig. 7a). For each additional reference microphone, WINC requires another wireless relay operating on a designated FM channel to avoid interference among the relays and from surrounding RF sources. A new antinoise filter is also required that takes the new reference signal as input and generates the corresponding antinoise signal to minimize the error. Therefore, the wireless spectrum usage, DSP computation resources, and memory usage scale linearly with the number of reference microphones.

Cancellation at the eardrum: In this work, all filters are optimized for the residual signal captured by the error mic. However, since the error mic is placed around the auricle of headphones, the error signal is different from what a person can hear at eardrums. [5] places another microphone in the inner-ear to measure the difference beforehand, which is removed later during cancellation. Such initialization is required for every individual and even every usage. Instead, WINC provides the opportunity to encompass the general hearing sensitivity contour and target at different residual power in each frequency. This can generally and effectively provide a more balanced experience to users.

Trade-off among diversity, look-ahead and computation: Channel diversity and look-ahead are the crucial and interacting components in WINC. Channel diversity is crucial for multiple-source noise cancellation, but the diversity provided by each extra reference microphone is limited by the distance to the error mic (i.e. look-ahead). Look-ahead is the key technique to compute the non-causal part of the ANC filter, but longer look-ahead leads to extra computation load, which has a negative effect on the cancellation. It is worth to study and find a suitable deployment of reference microphones to achieve a balance between these factors.

Usability and mobility: A major concern with a system like WINC that uses a network of microphones is constraints on usability of the system. We envision that the system will be deployed in indoor spaces such as offices and study rooms. With the proliferation of voice assistants and voice automated smart-homes and spaces, we believe that such an infrastructure will be readily available and make our system more accessible to users.

The key requirement for a noise cancelling system to work in mobile scenarios is that the adaptive antinoise filter should reconverge within the coherence time to accommodate the change in the acoustic channels. Since human motion, and hence the induced change in acoustic channels, is much slower than the speed at which the DSP board is expected to generate the antinoise samples, noise cancellation is expected to work in non-stationary scenarios. However, a short-coming of WINC as well as existing work on multi-noise source cancellation is that they are evaluated for stationary scenarios due to bulky implementation of the system. We therefore leave mobility analysis for future work.

10 CONCLUSION

This paper takes a further step on employing wireless network in active noise cancellation. It keeps the non-causal gain of wireless look-ahead and, in the meanwhile, highlights the more pronounced channel diversity gain of using multiple wireless reference microphones than traditional on-headphone microphones, which greatly improves the cancellation for multiple noise sources. The challenge in adaptive filter convergence is addressed by frequency-domain signal processing, which also opens up opportunities for user experience improvements in frequency-sensitive applications.

REFERENCES

[1] European Environment Agency. 2022. Noise pollution is a major problem, both for human health and the environment. https://www.eea.europa.eu/articles/noise-pollution-is-a-major
[2] Masaichi Akiho, Miki Haseyama, and Hideo Kitajima. 1999. A practical method to reduce a number of reference signals for the ANC system. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, Vol. 4. IEEE, 2387–2390.

[3] M. Al-Kindi and J. Dunlop. 1987. A low distortion adaptive noise cancellation structure for real time applications. In *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 12. 2153–2156.
[4] Apple. 2022. AirPods Pro and AirPods Max Active Noise Cancellation and Transparency mode - Apple Support. https://support.apple.com/en-us/HT210643
[5] Piero Rivera Benois, Reinhild Roden, Matthias Blau, and Simon Doclo. 2022. Optimization of a fixed virtual sensing feedback ANC controller for in-ear headphones with multiple loudspeakers. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8717–8721.
[6] Bose. 2022. Smart Noise Cancelling Headphones 700 | Bose. https://www.bose.com/en_us/products/headphones/noise_cancelling_headphones/noise-cancelling-headphones-700.html
[7] Cheng-Yuan Chang and Sen M Kuo. 2013. Complete parallel narrowband active noise control systems. *IEEE transactions on audio, speech, and language processing* 21, 9 (2013), 1979–1986.
[8] Jordan Cheer, Vinal Patel, and Simone Fontana. 2019. The application of a multi-reference control strategy to noise cancelling headphones. *The Journal of the Acoustical Society of America* 145, 5 (2019), 3095–3103.
[9] K Gomathi, V Saravanan, et al. 2016. Variable step size for improving convergence of FxLMS algorithm. *Procedia Technology* 25 (2016), 420–426.
[10] Satoru Hase, Yoshinobu Kajikawa, Lichuan Liu, and Sen M Kuo. 2015. Multi-channel ANC system using optimized reference microphones based on time difference of arrival. In *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 305–309.
[11] Boyan Huang, Yegui Xiao, Jinwei Sun, and Guo Wei. 2012. A variable step-size FXLMS algorithm for narrowband active noise control. *IEEE transactions on audio, speech, and language processing* 21, 2 (2012), 301–312.
[12] Texas Instruments Inc. 2019. EVMK2G Development kit TIcom. https://www.ti.com/tool/EVMK2G
[13] Bhan Lam, Dongyuan Shi, Woon-Seng Gan, Stephen J Elliott, and Masaharu Nishimura. 2020. Active control of broadband sound through the open aperture of a full-sized domestic window. *Scientific reports* 10, 1 (2020), 1–7.
[14] Paul Lueg. 1936. Process of silencing sound oscillations. *US patent 2043416* (1936).
[15] Yu Luo and Lina Pu. 2022. EC-ANC: Edge Case-Enhanced Active Noise Cancellation for True Wireless Stereo Earbuds. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022), 1436–1447.
[16] WB Mikhael and PD Hill. 1988. Acoustic noise cancellation in a multiple noise source environment. In *1988., IEEE International Symposium on Circuits and Systems*. IEEE, 2399–2402.
[17] Vinal Patel, Jordan Cheer, and Simone Fontana. 2020. Design and Implementation of an Active Noise Control Headphone With Directional Hear-Through Capability. *IEEE Transactions on Consumer Electronics* 66, 1 (2020), 32–40.
[18] Markus Rupp and Ali H Sayed. 1995. Modified FxLMS algorithms with improved convergence performance. In *Conference Record of The Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, Vol. 2. IEEE, 1255–1259.
[19] Markus Rupp and Ali H Sayed. 1998. Robust fxlms algorithms with improved convergence performance. *IEEE Transactions on Speech and Audio Processing* 6, 1 (1998), 78–85.
[20] Sheng Shen, Nirupam Roy, Junfeng Guan, Haitham Hassanieh, and Romit Roy Choudhury. 2018. MUTE: Bringing IoT to noise cancellation. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. 282–296.
[21] Xiaoyi Shen, Dongyuan Shi, and Woon-Seng Gan. 2021. A wireless reference active noise control headphone using coherence based selection technique. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 7983–7987.
[22] Xiaoyi Shen, Dongyuan Shi, and Woon-Seng Gan. 2022. A Hybrid Approach to Combine Wireless and Earcup Microphones for ANC Headphones with Error Separation Module. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8702–8706.
[23] Xiaoyi Shen, Dongyuan Shi, Santi Peksi, and Woon-Seng Gan. 2022. A multi-channel wireless active noise control headphone with coherence-based weight determination algorithm. *Journal of Signal Processing Systems* 94, 8 (2022), 811–819.
[24] National Geographic Society. 2022. Noise Pollution. https://education.nationalgeographic.org/resource/noise-pollution
[25] M.M. Sondhi, D.R. Morgan, and J.L. Hall. 1995. Stereophonic acoustic echo cancellation-an overview of the fundamental problem. *IEEE Signal Processing Letters* 2, 8 (1995), 148–151.
[26] Sony. 2022. Sony WH-1000XM4 Wireless Noise-Cancelling Headphones | WH-1000XM4 |Sony US. https://electronics.sony.com/audio/headphones/headband/p/wh1000xm4-b
[27] Gem Sound. 2022. GMW 1,2,4. https://www.gemsound.com/gmw-124
[28] László Sujbert and Attila Szarvas. 2018. Noise-canceling office chair with multiple reference microphones. *Applied Sciences* 8, 9 (2018), 1702.
[29] Xu Sun and Sen M Kuo. 2007. Active narrowband noise control systems using cascading adaptive filters. *IEEE transactions on audio, speech, and language processing* 15, 2 (2007), 586–592.
[30] Dunmin Zheng and Michael Anthony Zuniga. 2006. Telephonic handset employing feed-forward noise cancellation. US Patent 7,031,460.

APPENDIX

A CONVEXITY OF EQ. 14

Here we provide a proof for the convexity of the optimization problem Eq. 14. We first simplify Eq. 13 as:

$$E(f) = a_0(f) + \sum_{j=1}^{M} a_j(f) W_j(f) = \mathbf{a}^H \mathbf{W} \tag{19}$$

where the new variables are defined as:

$$\begin{aligned} a_0(f) &\triangleq \sum_{i=1}^{N} N_i(f) H_{ne}^{(i)}(f) \\ a_j(f) &\triangleq R_j(f) H_{se}(f) \quad j = 1, \dots, M \\ \mathbf{a} &\triangleq \begin{bmatrix} a_0(f) & a_1(f) & \cdots & a_M(f) \end{bmatrix}^H \\ \mathbf{W} &\triangleq \begin{bmatrix} 1 & W_1(f) & \cdots & W_M(f) \end{bmatrix}^T \end{aligned} \tag{20}$$

By the definition of convexity, for any $\mathbf{W} \in \mathbb{C}$, $\Delta\mathbf{W} \in \mathbb{C}$ and $\lambda \in [0, 1]$, the loss $\mathcal{L}(\mathbf{W}) = |\mathbf{a}^H \mathbf{W}|^2$ has to satisfy the following inequality:

$$\mathcal{L}(\mathbf{W} + \lambda\Delta\mathbf{W}) \leq (1 - \lambda)\mathcal{L}(\mathbf{W}) + \lambda\mathcal{L}(\mathbf{W} + \Delta\mathbf{W}) \tag{21}$$

The left and right side of the inequality is computed as:

$$\begin{aligned} \text{Left} &= \mathcal{L}(\mathbf{W} + \lambda\Delta\mathbf{W}) \\ &= |\mathbf{W} + \lambda\Delta\mathbf{W}|^2 \\ &= [\mathbf{a}^H(\mathbf{W} + \lambda\Delta\mathbf{W})]^H [\mathbf{a}^H(\mathbf{W} + \lambda\Delta\mathbf{W})] \\ &= (\mathbf{W} + \lambda\Delta\mathbf{W})^H \mathbf{a}\mathbf{a}^H (\mathbf{W} + \lambda\Delta\mathbf{W}) \\ &= \mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} + \lambda\mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W} \\ &\quad + \lambda\Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} + \lambda^2\Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W} \end{aligned} \tag{22}$$

$$\begin{aligned} \text{Right} &= (1 - \lambda)\mathcal{L}(\mathbf{W}) + \lambda\mathcal{L}(\mathbf{W} + \Delta\mathbf{W}) \\ &= (1 - \lambda)|\mathbf{a}^H\mathbf{W}|^2 + \lambda|\mathbf{a}^H(\mathbf{W} + \Delta\mathbf{W})|^2 \\ &= (1 - \lambda)[\mathbf{a}^H\mathbf{W}]^H[\mathbf{a}^H\mathbf{W}] \\ &\quad + \lambda[\mathbf{a}^H(\mathbf{W} + \Delta\mathbf{W})]^H[\mathbf{a}^H(\mathbf{W} + \Delta\mathbf{W})] \\ &= (1 - \lambda)\mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} + \lambda(\mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} \\ &\quad + \mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W} + \Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} + \Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W}) \\ &= \mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} + \lambda\mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W} \\ &\quad + \lambda\Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\mathbf{W} + \lambda\Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W} \end{aligned} \tag{23}$$

Subtracting Eq. 23 from Eq. 22,

$$\begin{aligned} \text{Left} - \text{Right} &= (\lambda^2 - \lambda)\Delta\mathbf{W}^H\mathbf{a}\mathbf{a}^H\Delta\mathbf{W} \\ &= \lambda(\lambda - 1)|\mathbf{a}^H\Delta\mathbf{W}|^2 \\ &\leq 0 \end{aligned} \tag{24}$$

Since Left ≤ Right always holds, Eq. 14 has a convex loss.