On the View-and-Channel Aggregation Gain in Integrated Sensing and Edge AI

Xu Chen, Khaled B. Letaief, and Kaibin Huang

Abstract

Sensing and edge *artificial intelligence* (AI) are two key features of the *sixth-generation* (6G) mobile networks. Their natural integration, termed *Integrated sensing and edge AI* (ISEA), is envisioned to automate wide-ranging *Internet-of-Tings* (IoT) applications. To achieve a high sensing accuracy, features of multiple sensor views are uploaded to an edge server for aggregation and inference using a large-scale AI model. The view aggregation is realized efficiently using over-the-air computing (AirComp), which also aggregates channels to suppress channel noise. As ISEA is at its nascent stage, there still lacks an analytical framework for quantifying the fundamental performance gains from view-and-channel aggregation, which motivates this work. Our framework is based on a well-established distribution model of multi-view sensing data where the classic Gaussian-mixture model is modified by adding sub-spaces matrices to represent individual sensor observation perspectives. Based on the model and linear classification, we study the End-to-End sensing (inference) uncertainty, a popular measure of inference accuracy, of the said ISEA system by a novel, tractable approach involving designing a scaling-tight uncertainty surrogate function, global discriminant gain, distribution of receive Signal-to-Noise Ratio (SNR), and channel induced discriminant loss. As a result, we prove that the E2E sensing uncertainty diminishes at an *exponential* rate as the number of views/sensors grows, where the rate is proportional to global discriminant gain. Given AirComp and channel distortion, we further show that the exponential scaling remains but the rate is reduced by a linear factor representing the channel induced discriminant loss. Furthermore, in the case of many spatial degrees of freedom, we benchmark AirComp against equally fast, traditional analog orthogonal access. The comparative performance analysis reveals a sensing-accuracy crossing point between the schemes corresponding to equal receive array size and sensor number. This leads to the proposal of a scheme for adaptive access-mode switching to enhance ISEA performance. Last, the insights from our framework are validated by experiments using a convolutional neural network model and real-world dataset.

X. Chen and K. Huang are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong. Khaled B. Letaief is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. Corresponding author: K. Huang (Email: huangkb@eee.hku.hk).

Index Terms

Integrated sensing and edge AI, Over-the-air computation, Multi-antenna communication

I. Introduction

In June 2023, the International Telecommunication Union (ITU-R) finalized six major usage scenarios for 6G. While others represent a scaled-up version of 5G, two are new – *Integrated AI and Communications* (IAAC) and *Integrated Sensing and Communication* (ISAC). IAAC reflects the 6G vision of edge intelligence, referring to ubiquitous distributed AI model training and inference at the network edge to support *Internet-of-Tings* (IoT) applications [1], [2]. On the other hand, ISAC will leverage edge devices as distributed sensors and network-scale cooperation to enable 6G networks to have multi-view observations of the physical world in real-time [3], [4]. The natural fusion of the two distinctive 6G functions, termed *Integrated Sensing and Edge AI* (ISEA), shall provide a powerful platform for automating a broad range of IoT applications including auto-pilot, robotic control, digital twins, augmented reality, and localization and tracking [1], [4]. Unleashing the full potential of ISEA calls for a new goal-oriented design approach that integrates sensing, AI, and communication to optimize the end-to-end (E2E) performance [5], [6]. In this work, we contribute to the theoretic characterization of the E2E performance of ISEA, thereby laying a foundation for goal-oriented designs.

A common backbone architecture for ISEA, called *multi-view convolutional neural network* (MVCNN), wirelessly connects distributed sensors to an edge server [7]. Each sensor uses a lightweight neural network model for feature extraction from local sensing data and then uploads the local features for aggregation and inference at the server using a pre-trained deep neural network model supporting multi-modal computer vision [6], [8], [9]. Local and server models are jointly trained as a single global model to maximize the E2E sensing (or inference) accuracy. This pertains to the common approach in edge AI, called split inference [10]. By treating local and server models as components splitting the global model, relevant techniques can enable the adaptation of the splitting point to balance the device computation load and performance requirements in terms of, e.g., E2E latency and communication resources [11]–[13]. The mentioned feature aggregation, commonly referred to as multi-view pooling, is a key operation of MVCNN that exploits multiple sensor observations to improve sensing accuracy. The server operation fuses received local features into an aggregated feature map that is input into the

global model (e.g., classifier) to generate a label identifying a target object/event. Element-wise averaging and maximization over local feature vectors are two popular aggregation functions termed average-pooling and max-pooling, respectively (see, e.g., [6]). Via view aggregation, multi-view sensing can attain an accuracy significantly higher than that of the single-view case especially when there are many sensors [7], [14]. However, the implementation of ISEA is confronted by a communication bottleneck resulting from the transmission of high-dimensional features by potentially a large cluster of sensors.

Massive access techniques for 5G are insufficient for tackling the communication bottleneck of edge AI, which includes ISEA as a special case. Such techniques, for example, grant-free massive access, assume low-rate sporadic transmission by many low-complexity sensors monitoring environmental variables such as humidity and temperature [15], [16]. In contrast, 6G sensors are usually multi-modal devices (e.g., cameras and LIDAR) deployed in data-intensive computer vision applications such as surveillance, autonomous driving, and drone swarms [17]. The challenges are escalated by the tactile applications targeted by 6G, such as augmented reality and remote robotics, which demand air latency below 1 milli-second [18]. The search for solutions motivates researchers to depart from the traditional communication-computing separation approach and advocate a paradigm shift towards the mentioned goal-oriented designs that target a specific task, such as distributed learning or sensing, and aim at maximizing the corresponding E2E system performance [1], [2]. One natural design approach for new paradigm is to customize existing techniques from multi-view sensing and edge AI, for example, sensor scheduling [19], feature compression [10], and hierarchical pooling [20], using an E2E metric (e.g., E2E sensing accuracy or latency) and targeting a specific air-interface technology (e.g., MIMO, OFDMA, and adaptive power control). An alternative, more revolutionary approach is to design new physical-layer technologies fully integrating computing and communication. In this vein, a representative class of techniques as considered in this work, called *over-the-air computation* (AirComp), integrates multi-access and nomographic functional computation (e.g., averaging and geometric mean) to solve the scalability problem in traditional multi-access that divides radio resources [21]. AirComp's basic principle is to exploit the wave superposition property to achieve over-the-air aggregation of uncoded analog signals simultaneously transmitted by multiple devices. The scalability as a result of simultaneous access makes AirComp a popular air-interface technology for supporting fast and efficient distributed computing in 6G operations such as distributed learning [22], inference [23], and sensing [6], [24]. In addition, the use of

uncoded analog transmission in AirComp is another factor contributing to the technology's ultra-low-latency while the resultant unreliability can be coped with by the robustness of data-analytics techniques or an AI algorithm [6], [25], [26]. In particular, AirComp has been extensively studied for implementing over-the-air aggregation of local model updates in federated learning (FL) systems, leading to the emergence of an area called over-the-air FL [27]. Diversified design issues have been investigated including gradient sparsification [22], beamforming [28], precoding [29], power control [30], broadband transmission [31].

Most recently, researchers also explored the applications of AirComp to realize over-the-air view aggregation in ISEA systems [6], [24], [32], [33]. In [6], max-pooling, which is not directly AirComputable, is realized using AirComp via p-norm approximation of maximization. The parameter of the approximation function is optimized to balance the noise effect and approximation error. The optimization still uses the generic metric of AirComp error (i.e., the error in computed function values with respect to the noiseless case) instead of the E2E sensing accuracy though the two metrics are related by a derived inequality. Similarly, the AirComp error is adopted in [32] as the performance metric to optimize a receive beamformer in a system supporting integrated MIMO radar sensing and AirComp. On the other hand, a different metric, discrimination gain, has been proposed to approximately measure the sensing accuracy with tractability. In [33], the effects of sensing, computation, and communication on the discrimination gain are quantified and controlled by designing a task-oriented resource management approach so as to optimize the E2E performance. The average discrimination gain for an individual feature dimension is further considered in [24] to facilitate importance-aware beamforming that adapts effective channel gains of different sensors according to their importance levels accounting for both average discriminant gains and channel states. Wireless for ISEA is still a nascent area where prior work largely focuses on algorithmic designs. There still lacks a systematic framework for analyzing E2E performance. Specifically, in the aspect of multi-view sensing, the sensing accuracy sees continuous improvements with the growth of the number of sensors providing view diversity. There exist few results on quantifying the scaling law. On the other hand, in the aspect of air interface, the AirComp error diminishes with the increase of the number of links due to aggregation and exploitation of (channel) spatial diversity [34]. The consideration of E2E performance for ISEA naturally couples the two aspects and gives rise to the following open research questions we attempt to answer sequentially in this work.

1) **(View Aggregation)** Consider ISEA without channel distortion. *How does the accuracy*

of multi-view sensing improve as the number of sensors grows?

2) **(View-and-Channel Aggregation)** Consider ISEA with wireless channels and using AirComp. *How does the E2E sensing accuracy improve as the number of sensors grows?*
3) **(Optimality of AirComp)** AirComp supports simultaneous access when spatial *degrees of freedom* (DoFs) are insufficient for orthogonal access. When many spatial DoFs are available, is AirComp still optimal for fast ISEA?

By making an attempt to answer these questions, we derive a theoretic framework for quantifying the E2E performance of an ISEA system implemented on the MVCNN architecture with an AirComp-based air interface. Key models and assumptions are summarized as follows. First, a well-established mathematical model for multi-view sensing is adopted [35], [36]. In this real-data validated model, features extracted from sensor observations (e.g., images) are described as low-rank projections of a high-dimensional ground-truth feature map, where a projection matrix, called *observation matrix*, reflects the spatial relationship between the associated sensor and the target object. Second, the feature map is assumed to distribute following the classic Gaussian mixture model (GMM) widely used in statistical learning [37] and deep learning (see, e.g., [38]). The model comprises multiple Gaussian clusters, each of which is tagged with an object-class label. Third, channel coefficients of the multiuser *single-input-multi-output* (SIMO) uplink channel are assumed to be independent and identically distributed (i.i.d.) Rayleigh fading, representing spatial diversity from rich scattering and spatially separated sensors. Last, the E2E sensing accuracy is measured by the popular metric of sensing uncertainty that is computed as the entropy of posteriors of object classes conditioned on observations [39], [40].

Then the key contributions and findings of this work are summarized as follows.

- **View Aggregation Gain**: To answer Research Question 1 for the noise-free case, the E2E sensing uncertainty is derived as a function of multiple factors including the number of sensors, number of (object) classes, and average differentiability of class pairs measured using the well known Mahalanobis distance. The average class differentiability is *with respect to* (w.r.t.) the feature sub-space defined using the *global observation matrix* that cascades the observation matrices of all sensors. The derivation exploits the tractability of GMM to derive asymptotically tight bounds on sensing uncertainty. The derived function reveals that view-aggregation gains in two aspects. On one hand, the monotonic reduction of sensing uncertainty w.r.t. the sensor population reflects its resultant suppression of *sensing*

noise. On the other hand, the function is also a monotone decreasing w.r.t. the average class differentiability that is in turn enhanced as a growing number of suitably scheduled sensors boosts the rank of the global observation matrix. When the number of views is large, the uncertainty function is shown to exhibit a simplified form *linearly proportional* to the number of classes but diminish at an *exponential rate* linearly proportional to the number of views/sensors.

- **View-and-Channel Aggregation Gain**: To address Research Question 2, we consider ISEA with channel distortion induced by AirComp for implementing multi-view aggregation. Building on the preceding analysis and applying random-matrix theory, the sensing uncertainty is shown to scale similarly as its noiseless counterpart except for an additional linear scaling factor for the exponential decay rate. The factor represents the negative channel effect on average class differentiability and is proved to be a monotone decreasing function of the effective receive SNR after AirComp. The analysis reveals that as the number of sensors increases, aggregation suppresses noise sufficiently fast such that the exponential decay of sensing uncertainty in the noiseless case is retained albeit at a slower exponential rate.
- **AirComp versus Orthogonal Access**: To answer Research Question 3, AirComp is benchmarked against analog orthogonal access, both of which support low-latency uncoded analog transmission [25], [26]. AirComp's main advantage lies in supporting spatial simultaneous access even when spatial DoFs are insufficient for orthogonal access. However, as the receive array size increases, we show the existence of a crossing point (with array size and number of sensors approximately equal) above which analog orthogonal access outperforms AirComp in terms of sensing uncertainty. This motivates an adaptive scheme that switches between AirComp and analog orthogonal access depending on the available spatial DoFs.
- **Experiments**: The preceding analytical results are validated in ISEA experiments using both synthetic (i.e., GMM) and real datasets (i.e., ModelNet [7]).

The remainder of this paper is organized as follows. The multi-view sensing and communication models are elaborated in Section II. The analysis results for noiseless and AirComp-based view aggregation cases are presented in Section III and Section IV, respectively. In Section V, we benchmark AirComp against analog orthogonal access, followed by performing experiments in Section VI.

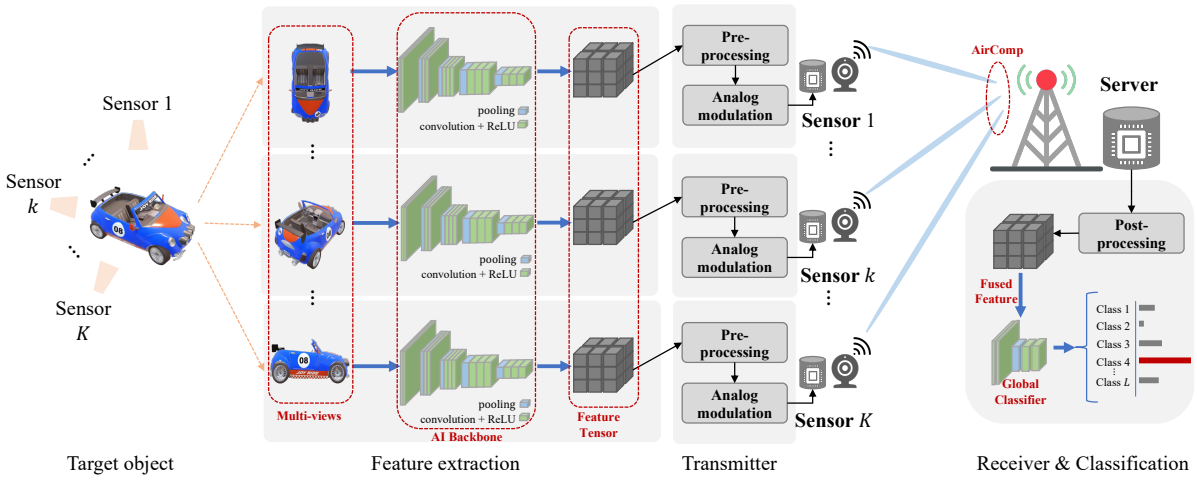


Fig. 1. A system integrating multi-view sensing and edge AI.

II. System, Models, and Metrics

Consider an ISEA system where a server realizes remote object detection by leveraging AI-based multi-view sensing over K distributed sensors, as illustrated in Fig. 1. Relevant operations, models, and metrics are described in the following sub-sections.

A. Multi-View Sensing Model

1) Local Data Distribution: Each sensor, say sensor k, feeds its captured raw data (e.g. images) into a pre-trained model to generate a feature map, denoted by $\mathbf{f}_k \in \mathbb{R}^M$, that comprises M real features. The distribution of $\mathbf{f}_k$ is given as follows. First, let $\mathbf{g} \in \mathbb{R}^M$ be the ground-truth feature map corresponding to the current object and be assumed to have a uniform prior distribution over L classes [24], [39]:

$$\Pr\left(\mathbf{g} = \boldsymbol{\mu}_\ell\right) = \frac{1}{L}, \ \forall \ell, \tag{1}$$

where $\boldsymbol{\mu}_\ell$ denotes the centroid of the ℓ-th class in the feature space. Then, due to the limited physical view of the sensors, $\mathbf{f}_k$ represents a low-dimensional projection of $\mathbf{g}$ [35], [36]. Adopting a well-established multi-view sensing model in the literature (see, e.g. [35]), we can relate the feature map $\mathbf{f}_k$ to $\mathbf{g}$ as

$$\mathbf{f}_k = \mathbf{P}_k \mathbf{g} + \mathbf{w}_k, \tag{2}$$

where $\mathbf{P}_k$ is the low-rank observation matrix of sensor k and $\mathbf{w}_k$ represents the inherent sensing noise following an *independent and identically distributed* (i.i.d.) Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{C})$. The observation matrices $\{\mathbf{P}_k\}$ and the covariance matrix $\mathbf{C}$ can be learned by using subspace-representation networks [35], [36] and are considered to be available to both devices and the server. It follows from (2) that local feature maps follow the distribution of a *Gaussian mixture model* (GMM) [37]:

$$\mathbf{f}_k \sim \frac{1}{L} \sum_{\ell=1}^{L} \mathcal{N}(\mathbf{P}_k \boldsymbol{\mu}_\ell, \mathbf{C}), \tag{3}$$

where $\mathcal{N}(\mathbf{P}_k \boldsymbol{\mu}_\ell, \mathbf{C})$ represents a Gaussian distribution with mean $\mathbf{P}_k \boldsymbol{\mu}_\ell$ and covariance $\mathbf{C}$.

2) Global Classification: Next, $\{\mathbf{f}_k\}$ are uploaded to the server for classification as follows. They are first fused into a single feature map, denoted by $\bar{\mathbf{f}}$, which is known as view aggregation (or pooling). The popular average aggregation is adopted as $\bar{\mathbf{f}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{f}_k$ [7]. Then, $\bar{\mathbf{f}}$ is fed into a classifier for inference. We consider two types of classifiers.

- **Linear Classification:** A linear classifier is considered in analysis for tractability [37]. Consider a case of two classes ($L = 2$), the linear classifier distinguishes the pair of classes by using a classification boundary between their clusters, which is defined as a hyperplane in the feature space

$$\mathcal{H}(\boldsymbol{\alpha}, \beta) = \{\mathbf{f} : \boldsymbol{\alpha}^\top \mathbf{f} + \beta = 0\}. \tag{4}$$

  The optimal label of an input feature map $\bar{\mathbf{f}}$ is assigned as one class if $\bar{\mathbf{f}}$ is determined to be above the hyperplane (i.e., $\boldsymbol{\alpha}^\top \bar{\mathbf{f}} + \beta \geq 0$); otherwise, $\bar{\mathbf{f}}$ is labeled as the other class. In a general case with $L > 2$, there are $L(L-1)/2$ classification boundaries, and the optimal result can be obtained via sequential conduction of the one-versus-one classification. Given equal priors of classes, $\{\mathbf{f}_k\}$ follow the distribution in (3). The optimal L-class linear classifier is the *maximum likelihood* (ML) design [37]:

$$\ell^\star = \arg\max_{\ell} \ \log \Pr\left(\bar{\mathbf{f}} | \boldsymbol{\mu}_\ell\right). \tag{5}$$

- **CNN Classification:** MVCNN model is adopted in experiments. The model consists of two parts, $\mathcal{F}_1$ and $\mathcal{F}_2$, that are employed at sensors and the server, respectively. The sub-model $\mathcal{F}_1$ is identical for sensors and used to extract local features from sensing data. After view aggregation, the obtained aggregated feature vector $\bar{\mathbf{f}}$ is fed into $\mathcal{F}_2$ that outputs scores for individual classes. The class with the highest score is selected as the prediction result.

B. Multi-Access Models

For the ISEA system in Fig. 1, we mainly consider analog multi-access techniques for enabling efficient simultaneous access (i.e., view-and-channel aggregation). We also explore ISEA with noiseless feature aggregation in Section III, aligning with scenarios involving reliable digital transmission. In the class of analog transmission, we primarily adopt AirComp for feature aggregation. To investigate its optimality, we further consider analog orthogonal access, namely orthogonal access with fast analog transmission [26], as a benchmark scheme. It achieves the same multi-access latency as AirComp but requires receive spatial DoF to be equal to or exceed the number of sensors. The assumptions and operations of the schemes are described as follows. The server and sensors are equipped with N-element array and a single antenna, respectively. Assuming a frequency non-selective channel, time is slotted and each slot is used for transmitting one symbol. Block fading is considered such that the channel remains unchanged over a coherence duration comprising T time slots. Symbol-level synchronization is assumed over all sensors.

1) Analog Transmission: In an arbitrary time slot, say slot t, sensors simultaneously transmit their linear analog modulated data symbol, $\{x_{k,t}\}$, leading to the server receiving a symbol vector:

$$\mathbf{y}_t = \sum_k \rho_k \mathbf{h}_k x_{k,t} + \mathbf{z}_t, \tag{6}$$

where ρ_k represents transmit power, $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$ denotes the channel vector of sensor k, and $\mathbf{z}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$ models additive channel noise. Assuming Rayleigh fading, $\mathbf{h}_k$ is composed of i.i.d. $\mathcal{CN}(0,1)$ entries and is independent between sensors. Let $\nu^2 \triangleq \mathsf{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_{k,t}^2\right]$ be the variance of transmitted symbols over a channel coherence block. Each sensor is constrained by a power budget of P, i.e., $\rho_k^2\nu^2 \leq P$. Then, the transmit SNR is defined as $\gamma = \frac{P}{\sigma^2}$.

Let $\mathbf{x}_k = [x_{k,1}, \cdots, x_{k,M}]$ denote the symbol vector transmitted from sensor k over M time slots with $M \leq T$. The data vector in (6) received over M time slots can be aggregated into a matrix symbol, $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_M]$:

$$\mathbf{Y} = \sum_k \rho_k \mathbf{h}_k \mathbf{x}_k + \mathbf{Z}, \tag{7}$$

where $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_M]$.

2) Receiver for AirComp: The symbol vector $\mathbf{x}_k$ in (7) is computed as $\mathbf{x}_k = (\mathbf{f}_k - \mathbf{f}_k^{\mathsf{avg}})^\top$ with $\mathbf{f}_k^{\mathsf{avg}} = \mathsf{E}[\mathbf{f}_k]$ to have zero mean. The parameter $\mathbf{f}_k^{\mathsf{avg}}$ is also available for both the server to receive the features. Following the AirComp literature, *zero-forcing* (ZF) transmit power control is adopted to overcome channel distortion, $\rho_k = (\mathbf{b}^H \mathbf{h}_k)^{-1}$ with $\mathbf{b} \in \mathbb{C}^N$ being a receive combiner (see, e.g., [21]). The output is given as $\mathbf{s} = \mathbf{Y}^H \mathbf{b} = \sum_k \mathbf{f}_k - \mathbf{f}_k^{\mathsf{avg}} + \mathbf{Z}^H \mathbf{b}$. From $\mathbf{s}$, a noisy version of $\bar{\mathbf{f}}$, denoted by $\tilde{\mathbf{f}}$, can be obtained by the following post-processing:

$$\tilde{\mathbf{f}} = \frac{1}{K}\mathbf{s} + \frac{1}{K}\sum_k \mathbf{f}_k^{\mathsf{avg}} = \bar{\mathbf{f}} + \frac{1}{K}\mathbf{Z}^H \mathbf{b}. \tag{8}$$

3) Receiver for Orthogonal Multi-Access: The simultaneous data streams in (7) are orthogonalized via receive beamforming. It is optimal to maximize transmit power at each sensor as $\rho_k = \frac{\sqrt{P}}{\nu}$. Then the received symbol matrix can be rewritten as $\mathbf{Y} = \sum_k \frac{\sqrt{P}}{\nu}\mathbf{h}_k \mathbf{x}_k + \mathbf{Z}$. Let $\mathbf{e}_k = [0, \cdots, 1, \cdots, 0]^\top$ denote the standard basis vector with the k-th element being 1. The data stream of sensor k, denoted by $\mathbf{s}_k$, can be extracted from $\mathbf{Y}$ using a ZF beamformer $\mathbf{b}_k = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_k$ with $\mathbf{H} = [\mathbf{h}_1, \cdots, \mathbf{h}_K]$ [41]:

$$\mathbf{s}_k = \mathbf{Y}^H \mathbf{b}_k = \frac{\sqrt{P}}{\nu}\left(\bar{\mathbf{f}}_k - \mathbf{f}_k^{\mathsf{avg}}\right) + \mathbf{Z}^H \mathbf{b}_k. \tag{9}$$

By slight abuse of notation, let $\tilde{\mathbf{f}}$ also denote the noisy version of the desired aggregated feature vector in the current case. Then it can be obtained by the following post-processing:

$$\tilde{\mathbf{f}} = \frac{1}{K}\sum_k \left(\frac{\nu}{\sqrt{P}}\mathbf{s}_k + \mathbf{f}_k^{\mathsf{avg}}\right) = \bar{\mathbf{f}} + \frac{\nu}{K\sqrt{P}}\mathbf{Z}^H \sum_k \mathbf{b}_k. \tag{10}$$

C. E2E Performance Metric

Typically, the sensing (inference) accuracy is defined as the probability of correct classification. For the purpose of tractable analysis, we also consider the following two relevant metrics for evaluating the E2E sensing performance.

1) Sensing Uncertainty: As a popular measure related to sensing (inference) accuracy, the metric is defined as the entropy of posteriors of classification classes given the aggregated feature [40]. Mathematically, given the aggregated feature map $\tilde{\mathbf{f}}$, its sensing uncertainty, denoted by H, is given as

$$H = \mathsf{E}_{\tilde{\mathbf{f}}}\left[-\sum_{\ell=1}^{L} \Pr\left(\boldsymbol{\mu}_\ell | \tilde{\mathbf{f}}\right) \log \Pr\left(\boldsymbol{\mu}_\ell | \tilde{\mathbf{f}}\right)\right]. \tag{11}$$

2) Discrimination Gain: The sensing accuracy is largely determined by the discernibility between a pair of classes that can be measured by *discrimination gains* computed as their *symmetric Kullback-Leibler (KL) divergence* [39]. Considering sensor k, the local discrimination gain between classes ℓ and ℓ', denoted as $G_k(\ell, \ell')$, can be computed as

$$\begin{aligned} G_k(\ell,\ell') =& \mathsf{KL}\left(\mathcal{N}\left(\mathbf{P}_k\boldsymbol{\mu}_\ell, \mathbf{C}\right) || \mathcal{N}\left(\mathbf{P}_k\boldsymbol{\mu}_{\ell'}, \mathbf{C}\right)\right) \\ &+ \mathsf{KL}\left(\mathcal{N}\left(\mathbf{P}_k\boldsymbol{\mu}_{\ell'}, \mathbf{C}\right) || \mathcal{N}\left(\mathbf{P}_k\boldsymbol{\mu}_\ell, \mathbf{C}\right)\right) \\ =& \left(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}\right)^\top \mathbf{P}_k \mathbf{C}^{-1} \mathbf{P}_k \left(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}\right). \end{aligned} \tag{12}$$

The global discriminant gain is derived in the next section.

III. Multi-view Aggregation Gain without Channel Distortion

In this section, we consider the scenario with the absence of channel noise and focus on analyzing the E2E performance of an ISEA system in terms of sensing uncertainty. The tractable analysis consists of three steps presented in separate sub-sections, namely characterizing the distribution of aggregated features, designing a suitable surrogate function for sensing uncertainty, and deriving the scaling laws of sensing uncertainty.

A. Aggregated Feature Distribution

The computation of sensing uncertainty in (11) relies on an explicit distribution of the aggregated feature map, $\bar{\mathbf{f}}$, at the input to the classifier. Based on the GMM model of local features, the desired result is derived as shown below.

Lemma 1 (Distribution of Aggregated Feature Map)**.** Based on the distribution of local feature maps in (3) and in the absence of channel noise, the aggregated feature $\bar{\mathbf{f}}$ follows a Gaussian-mixture distribution given as

$$\bar{\mathbf{f}} \sim \frac{1}{L}\sum_{\ell=1}^{L} \mathcal{N}\left(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell, \frac{1}{K}\mathbf{C}\right),$$

where $\bar{\mathbf{P}} = \frac{1}{K}\sum_k \mathbf{P}_k$ denotes the average of local observation matrices, termed *global observation matrix*.

Proof. (See Appendix A). □

Comparing Lemma 1 and (3), it is observed that the multi-view aggregation retains the distribution of features except for 1) replacing the cluster centroid of individual classes with their projections onto the global observation matrix and 2) narrowing the cluster size by reducing covariance by the factor $1/K$. These two factors contribute to the multi-view aggregation gain in sensing performance, which is quantified in the sequel.

Using Lemma 1 and (5), the optimal linear classifier with $\bar{\mathbf{f}}$ as input can be written as

$$\ell^\star = \arg\min_{\ell} \ \left(\bar{\mathbf{f}} - \bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)^\top \mathbf{C}^{-1} \left(\bar{\mathbf{f}} - \bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right). \tag{13}$$

The uniqueness of the inferred label $\ell^\star$ is guaranteed by the independence among views and $\bar{\mathbf{f}}$ being a continuous random variable.

B. Surrogate Function for Sensing Uncertainty

For the purpose of tractability, we propose a simpler but scaling-tight surrogate function for sensing uncertainty. To this end, let $G_{\ell,\ell'}$ denote the global discrimination gain of $\bar{\mathbf{f}}$. Using the local discriminant gain in (12), the expression of $G_{\ell,\ell'}$ can be obtained as shown below.

Lemma 2 (Global Discrimination Gain)**.** Using $\bar{\mathbf{f}}$ for classification, the global discrimination gain between class ℓ and ℓ' is given as $G_{\ell,\ell'} = KD_{\ell,\ell'}$ where

$$D_{\ell,\ell'} = \left(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}\right)^\top \bar{\mathbf{P}}\mathbf{C}^{-1}\bar{\mathbf{P}} \left(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}\right).$$

In Lemma 2, $D_{\ell,\ell'}$ is equal to the Mahalanobis distance between classes ℓ and ℓ', which reflects their differentiability [42]. Then, leveraging the distribution of $\bar{\mathbf{f}}$ in Lemma 1 and optimization theory, the resulting sensing uncertainty is obtained as follows.

Proposition 1 (Sensing Uncertainty)**.** In the case of linear classification, the E2E sensing uncertainty, H, in (11) can be bounded as

$$\frac{1}{L}\sum_{\ell=1}^{L} \log\left[1 + \sum_{\ell'\neq\ell} \exp\left(-\frac{D_{\ell,\ell'}}{2}K\right)\right] \leq H \leq$$
$$\frac{1}{L}\sum_{\ell=1}^{L} \log\left[1 + \sum_{\ell'\neq\ell} \exp\left(-\frac{D_{\ell,\ell'}}{cM+2}K\right)\right] + C_a,$$

where $c > 0$ is arbitrary, the constant $C_a = \log \frac{ce^{\frac{1}{c}}}{1+c}$, and $D_{\ell,\ell'}$ is given in Lemma 2.

Proof. (See Appendix B). □

The relatively complex expression of H in (11) does not allow tractable analysis of its scaling laws. Nevertheless, its bounds in Proposition 1 suggests that H can be approximated by the following surrogate function:

$$H_{\mathsf{s}} = \frac{1}{L}\sum_{\ell=1}^{L}\log\left[1+\sum_{\ell'\neq\ell}\exp\left(-\kappa D_{\ell,\ell'}K\right)\right], \tag{14}$$

where κ can be $\frac{1}{2}$ and $\frac{1}{cM+2}$ corresponding to the lower and upper bounds in Proposition 1, respectively. The function is found to follow a similar scaling law as H, which is essential for subsequent asymptotic analysis. The scaling-tight property of H_{s} is validated using the following numerical example.

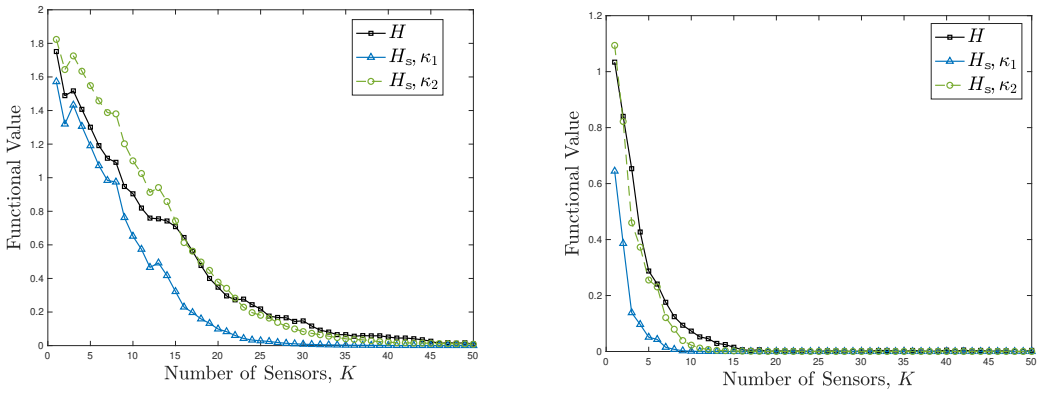
Example 1 (Numerical Validation)**.** Let the feature dimension and the number of classes be equal: $M = L = 5$. The covariance of the local feature distribution is set as $\mathbf{C} = 0.1\mathbf{I}_M$. The curves of exact sensing uncertainty and surrogate function in (14) are plotted in Fig. 2 for a variable number of views/sensors. By comparing the curves, one can observe that the sensing uncertainty is not necessarily a monotonically decreasing function w.r.t. K (e.g., from $K = 2$ to $K = 3$) when K and the local observation DoFs, referring to $\mathsf{rank}(\mathbf{P}_k)$, are small. The monotonicity of multi-view aggregation w.r.t. K emerges when local sensors acquire sufficiently many observation DoFs (i.e., 2), as shown in Fig. 2(b). Based on observations from Fig. 2, we can conclude that the surrogate function accurately captures the scaling law of sensing uncertainty including reflecting the said glitches of monotonicity.

C. Aggregation Gain

1) Simplifying Uncertainty Surrogate: To facilitate the analysis of multi-view aggregation gain, we simplify the expression of uncertainty surrogate function in (14) by Taylor expansion. First, define the average class separation distance as

$$\bar{D} = \frac{1}{L(L-1)}\sum_{\ell=1}^{L}\sum_{\ell'\neq\ell} D_{\ell,\ell'}, \tag{15}$$

where $D_{\ell,\ell'}$ is given in Lemma 2.



(a) Observation DoF = 1 (b) Observation DoF = 2

Fig. 2. Numerical validation on the surrogate sensing uncertainty, $\kappa_1 = \frac{1}{0.5M+1}$, $\kappa_2 = \frac{1}{M+1}$.

Proposition 2 (Surrogate Function Expansion)**.** In the absence of channel noise, the surrogate function of sensing uncertainty in (14) can be written as

$$H_{\mathsf{s}} = \log\left[1 + (L-1)\exp\left(-\kappa \bar{D} K\right)\right] + C_b,$$

where $C_b = \mathcal{O}\left(\frac{1}{L(L-1)}\sum_{\ell}\sum_{\ell \neq \ell'}(D_{\ell,\ell'} - \bar{D})^2\right)$.

Proof. (See Appendix C). □

In Proposition 2, the residual term C_b is negligible when between-class differentiability is similar, i.e.,

$$D_{\ell,\ell'} \approx D_{l,l'}, \ \forall \ell, \ell', l, l'. \tag{16}$$

To simplify notation, we assume that this is the case and hence $C_b \approx 0$ in the subsequent analysis. As a result, the sensing uncertainty surrogate reduces to

$$H_{\mathsf{s}} \approx \log\left[1 + (L-1)\exp\left(-\kappa \bar{D} K\right)\right]. \tag{17}$$

It is worth mentioning that our following analysis also holds for the case of $C_b \neq 0$ which, however, complicates notation and makes analysis tedious.

Based on (17), H_{s} is observed to be a monotonically decreasing function of the product of the views' number and the average differentiability, say $K \cdot \bar{D}$. The product reflects two aspects of multi-view aggregation gain. On one hand, as K grows, the aggregation over more views

suppresses the variances of sensing data clusters and thereby decreases the sensing uncertainty. On the other hand, a growing number of suitably scheduled sensors also enhances the average class differentiability represented by $\bar{D}$ as elaborated shortly.

2) Characterizing Average Class Separation Distance: By substituting Lemma 2 into (15), $\bar{D}$ can be written as

$$\bar{D} = \mathsf{Tr}(\bar{\mathbf{P}}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D}), \tag{18}$$

where $\mathbf{D} = \frac{1}{L(L-1)}\sum_{\ell=1}^{L}\sum_{\ell'\neq\ell}(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top$ reflects the average of pairwise class separation matrices, $\{(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top\}$. Therefore, $\bar{D}$ measures the components of $\mathbf{D}$ projected onto the subspace spanned by $\bar{\mathbf{P}}$. The global (multi-view) observation matrix $\bar{\mathbf{P}}$ has a larger rank than each of its local components $\{\mathbf{P}_k\}$. However, including more views/sensors does not necessarily increase the value $\bar{D}$. It may even decrease by the addition of a sensor contributing little useful information in its observation sub-space, as observed in Example 1. This suggests the need of designing a sensor scheduler using the criterion of maximizing $\bar{D}$ when the views are limited.

3) Main Result: With sufficient independent views in aggregation, the average of random observation matrices $\bar{\mathbf{P}}$ converges to its expectation, and thus the trace value $\mathsf{Tr}(\bar{\mathbf{P}}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D})$ in (18) reduces into a constant given as:

$$\xi \triangleq \mathsf{Tr}(\mathsf{E}[\mathbf{P}_k]\mathbf{C}^{-1}\mathsf{E}[\mathbf{P}_k]\mathbf{D}). \tag{19}$$

Then the sensing uncertainty H_{s} in (17) converges to

$$H_{\mathsf{s}} = \log\left[1 + (L-1)\exp\left(-\kappa\xi K\right)\right]. \tag{20}$$

Then the main result of this section is obtained as follows.

Main Result 1 (View Aggregation Gain)**.** For a large number of views, the sensing uncertainty with noiseless multi-view aggregation is exponentially decreasing w.r.t. K:

$$\boxed{H_{\mathsf{s}} \approx (L-1)\exp(-\kappa\xi K), \qquad K \gg 1,} \tag{21}$$

where κ can be $\frac{1}{2}$ and $\frac{1}{cM+2}$ according to the definition in (14) and ξ follows (19).

IV. Multi-view Aggregation Gain with Channel Distortion

In this section, we build on the results in the preceding section to quantify the view-and-channel aggregation gain for the E2E sensing performance of an ISEA system with AirComp over fading channels. In particular, the results on the distribution of aggregated features and surrogate functions for sensing uncertainty are extended to account for channel distortion. We further obtain useful results on receive SNR distribution and channel-induced sensing-accuracy loss.

A. Aggregated Feature Distribution

First, as the transmitted feature maps are real, the real part of the aggregated feature map $\tilde{\mathbf{f}}$ is extracted as $\tilde{\mathbf{f}}^{\mathsf{re}} = \Re\{\tilde{\mathbf{f}}\}$. The distribution of $\tilde{\mathbf{f}}^{\mathsf{re}}$ is derived as shown below.

Lemma 3. Given AirComp in (8), the aggregated feature map follows the Gaussian-mixture distribution:

$$\tilde{\mathbf{f}}^{\mathsf{re}} \sim \frac{1}{L}\sum_{\ell=1}^{L}\mathcal{N}\left(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell, \frac{1}{K}\mathbf{C} + \frac{1}{\gamma_{\mathsf{air}}}\mathbf{I}_M\right),$$

where $\gamma_{\mathsf{air}} \triangleq \frac{2K^2}{\sigma^2\|\mathbf{b}\|^2}$ denotes the *effective receive SNR*.

Proof. (See Appendix D). □

It follows from Lemma 3 and (5) that the optimal ML classifier in the current case is given as

$$\ell^\star = \arg\min_{\ell}\ (\tilde{\mathbf{f}}^{\mathsf{re}} - \bar{\mathbf{P}}\boldsymbol{\mu}_\ell)^\top \left(\frac{1}{K}\mathbf{C} + \frac{1}{\gamma_{\mathsf{air}}}\mathbf{I}_M\right)^{-1}(\tilde{\mathbf{f}}^{\mathsf{re}} - \bar{\mathbf{P}}\boldsymbol{\mu}_\ell). \tag{22}$$

B. Surrogate Function for Sensing Uncertainty

Given the similar forms of aggregated feature distributions in Lemmas 1 and 3, the needed surrogate function can be derived similarly as its noiseless counterpart in Section III. First, based on Lemma 2 and 3, the global discrimination gain of $\tilde{\mathbf{f}}^{\mathsf{re}}$ is given as $\tilde{G}_{\ell,\ell'} = K\tilde{D}_{\ell,\ell'}(\gamma_{\mathsf{air}})$ where the average class separation distance with channel distortion is given as

$$\tilde{D}_{\ell,\ell'}(\gamma_{\mathsf{air}}) = (\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})^\top \bar{\mathbf{P}}(\mathbf{C} + \frac{K}{\gamma_{\mathsf{air}}}\mathbf{I}_M)^{-1}\bar{\mathbf{P}}\,(\boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'})\,. \tag{23}$$

One can observe that the effect of channel distortion on the discrimination gain is regulated by a single parameter – the effective receive SNR γ_{air}. Using the above result and following

the procedure as deriving Proposition 1, the sensing uncertainty with channel distortion can be bounded as follows.

Corollary 1 (Sensing Uncertainty with Channel Distortion)**.** Consider the ISEA system with linear classification and AirComp. The resulting sensing uncertainty can be bounded as

$$\frac{1}{L}\sum_{\ell=1}^{L}\log\left[1+\sum_{\ell'\neq\ell}\exp\left(-\frac{\tilde{D}_{\ell,\ell'}(\gamma_{\mathsf{air}})}{2}K\right)\right]\leq H(\gamma_{\mathsf{air}})\leq$$
$$\frac{1}{L}\sum_{\ell=1}^{L}\log\left[1+\sum_{\ell'\neq\ell}\exp\left(-\frac{\tilde{D}_{\ell,\ell'}(\gamma_{\mathsf{air}})}{cM+2}K\right)\right]+C_a,$$

where $c>0$, C_a is the constant same as in Proposition 1, and $\tilde{D}_{\ell,\ell'}(\gamma_{\mathsf{air}})$ is given in (23).

It follows that the noisy counterpart of the uncertainty surrogate function in (17) can be obtained as

$$H_{\mathsf{s}}(\gamma_{\mathsf{air}})=\log\left[1+(L-1)\exp\left(-\kappa\tilde{D}(\gamma_{\mathsf{air}})K\right)\right], \tag{24}$$

where $\tilde{D}(\gamma_{\mathsf{air}})\triangleq\frac{1}{L(L-1)}\sum_{\ell=1}^{L}\sum_{\ell'\neq\ell}\tilde{D}_{\ell,\ell'}(\gamma_{\mathsf{air}})=\mathsf{Tr}(\bar{\mathbf{P}}(\mathbf{C}+\frac{K}{\gamma_{\mathsf{air}}}\mathbf{I}_M)^{-1}\bar{\mathbf{P}}\mathbf{D})$.

C. Distribution of Effective Receive SNR

The dependence of sensing uncertainty on the effective receive SNR, γ_{air}, as reflected in (24) suggests the need of analyzing its distribution, which is carried out as follows.

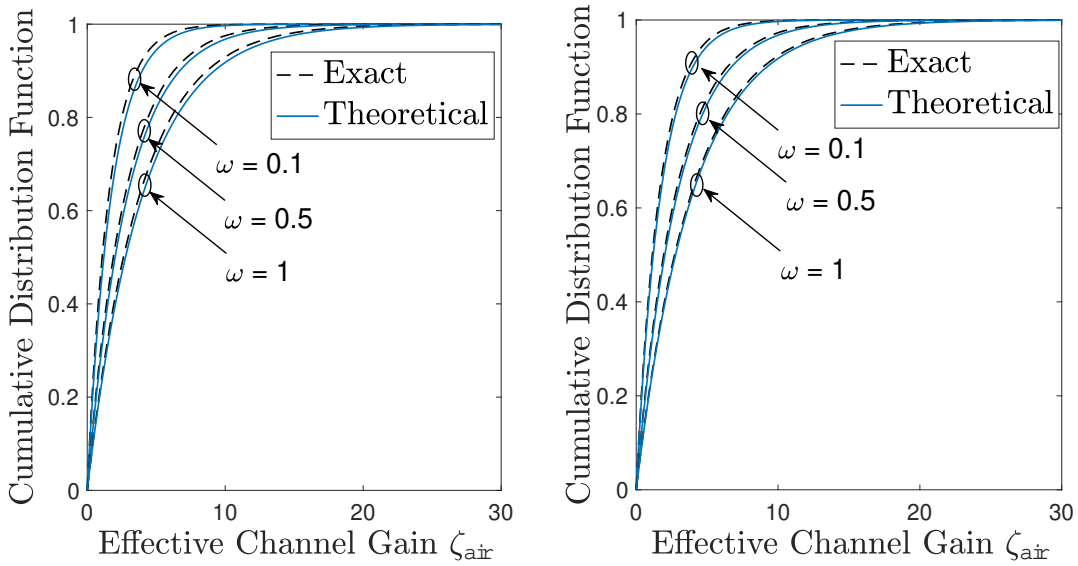
Lemma 4. The sensing uncertainty function $H_{\mathsf{s}}(\gamma_{\mathsf{air}})$ in (24) is monotonically decreasing w.r.t. γ_{air}.

Proof. (See Appendix E). □

Under the power constraint $\frac{\nu^2}{\mathbf{b}^H\mathbf{h}_k\mathbf{h}_k^H\mathbf{b}}\leq P$, the effective receive SNR maximized by optimal beamforming is given as [21]

$$\gamma_{\mathsf{air}}=\frac{2K^2\gamma}{\nu^2}\cdot\min_k\ \mathbf{v}^H\mathbf{h}_k\mathbf{h}_k^H\mathbf{v}, \tag{25}$$

where $\mathbf{v}$ denotes the optimal receive beamformer computed as the first eigenvector of the channel matrix $\mathbf{H}=[\mathbf{h}_1,\mathbf{h}_2,\cdots,\mathbf{h}_K]$. As revealed in (25), γ_{air} is limited by the weakest link due to signal magnitude alignment of AirComp. The alignment term, $\min_k\ \mathbf{v}^H\mathbf{h}_k\mathbf{h}_k^H\mathbf{v}$, will reduce to zero as $K\to\infty$. However, its value scaled up proportional to K, say $\zeta_{\mathsf{air}}\triangleq K\cdot\min_k\ \mathbf{v}^H\mathbf{h}_k\mathbf{h}_k^H\mathbf{v}$, can



(a) $K = 100$ (b) $K = 200$

Fig. 3. Numerical validation of Lemma 5.

converge to an exponential random variable with a fixed parameter. This gives the distribution of γ_{air} as follows.

Lemma 5 (Asymptotic Distribution of Effective Receive SNR)**.** For a large number of sensors ($K \to \infty$) and a proportional array size $N = \omega K$ with $\omega \in (0, \infty)$, the effective receive SNR resulting from AirComp $\gamma_{\mathsf{air}} = 2K\frac{\gamma}{\nu^2} \cdot \zeta_{\mathsf{air}}$ with ζ_{air} being an exponential random variable:

$$\zeta_{\mathsf{air}} \sim \mathrm{Exp}\left(\frac{1}{(1+\sqrt{\omega})^2}\right).$$

Proof. (See Appendix F). □

The asymptotic distribution of ζ_{air} is numerically validated in Fig. 3. The distribution can be further extended to the case with a fixed array size, in which the continuously increasing number of sensors will lead to $\omega = 0$ and $\zeta_{\mathsf{air}} \sim \mathrm{Exp}\,(1)$. Using Lemma 5, the scaling property of exponential distributions yields the asymptotic distribution of the receive SNR:

$$\frac{\gamma_{\mathsf{air}}}{K} \sim \mathrm{Exp}\left(\frac{\nu^2}{2\gamma(1+\sqrt{\omega})^2}\right), \qquad K \to \infty, N = \omega K. \tag{26}$$

Remark 1 (Channel Noise Suppression)**.** It follows from (26) that the power of channel noise in AirComp, given by $\frac{1}{\gamma_{\mathsf{air}}} = \frac{1}{\gamma_{\mathsf{air}}/K} \cdot \frac{1}{K}$, is inversely proportional to K as the term γ_{air}/K is

independent of K.

D. Aggregation Gain

Given the preceding analysis, we are ready to quantify the view-and-channel aggregation gain under the joint effects of view aggregation that suppresses sensing noise and channel aggregation that suppresses channel noise. The core of the analysis is to derive a key variable – channel-induced loss on sensing performance as follows.

1) Channel Induced Performance: As shown in (24), multi-view aggregation with channel distortion can achieve sensing uncertainty with the same form as its noiseless counterpart, except for reducing the average class differentiability from $\bar{D}$ to $\tilde{D}(\gamma_{\mathsf{air}})$ defined in (24). It follows that their ratio can represent the channel-induced performance loss: $A_{\mathsf{loss}} \triangleq \frac{\tilde{D}(\gamma_{\mathsf{air}})}{\bar{D}}$. By using the Woodbury matrix identity [43], $\tilde{D}(\gamma_{\mathsf{air}})$ in (24) can be expressed as

$$\begin{aligned}
&\tilde{D}(\gamma_{\mathsf{air}}) \\
&= \mathsf{Tr}\left(\bar{\mathbf{P}}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D}\right) - \mathsf{Tr}(\bar{\mathbf{P}}\mathbf{C}^{-1}(\mathbf{C}^{-1} + \frac{K}{\gamma_{\mathsf{air}}}\mathbf{I}_M)^{-1}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D}), \\
&= \bar{D} - \mathsf{Tr}(\bar{\mathbf{P}}\mathbf{C}^{-1}(\mathbf{C}^{-1} + \frac{K}{\gamma_{\mathsf{air}}}\mathbf{I}_M)^{-1}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D}).
\end{aligned} \tag{27}$$

As a result,

$$A_{\mathsf{loss}} = 1 - \frac{\mathsf{Tr}(\bar{\mathbf{P}}\mathbf{C}^{-1}(\mathbf{C}^{-1} + \frac{K}{\gamma_{\mathsf{air}}}\mathbf{I}_M)^{-1}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D})}{\mathsf{Tr}\left(\bar{\mathbf{P}}\mathbf{C}^{-1}\bar{\mathbf{P}}\mathbf{D}\right)}. \tag{28}$$

$$H_{\mathsf{s}}(\gamma_{\mathsf{air}}) = \log\left[1 + (L-1)\exp\left(-\kappa\bar{D}KA_{\mathsf{loss}}\right)\right]. \tag{29}$$

2) Main Result: Based on (26), A_{loss} is independent of K as $K \to \infty$. Therefore, combining (24) and (28) yields the main result.

Main Result 2 (View-and-Channel Aggregation Gain)**.** Consider an ISEA system employing AirComp-based multi-view aggregation, the sensing uncertainty is exponentially decreasing w.r.t. K:

$$\boxed{H_{\mathsf{s}}(\gamma_{\mathsf{air}}) \approx (L-1)\exp\left(-\kappa\xi A_{\mathsf{loss}}K\right), \qquad K \gg 1,} \tag{30}$$

where κ and ξ are given in (21).

The above result shows that due to channel aggregation in AirComp, channel distortion does not change the exponential decay of sensing uncertainty but does reduce the exponential rate by

a factor of A_{loss}.

Last, we investigate the effects of several system parameters on the key variable A_{loss}. Define $r = 2\gamma\nu^{-2}(1+\sqrt{\omega})^2\lambda_{\mathbf{C},\min}$ for ease of notation, where γ, ν^2, ω and $\lambda_{\mathbf{C},\min}$ denote transmit SNR, the variance of transmit symbols, sensor-antenna number ratio, and the largest eigenvalue of the feature covariance matrix $\mathbf{C}$, respectively. Then, using the distribution of γ_{air} in (26), the expectation of A_{loss} is obtained as

$$\begin{aligned}\mathsf{E}_{\frac{K}{\gamma_{\mathsf{air}}}}\left[A_{\mathsf{loss}}\right] &\geq 1-\mathsf{E}_{\frac{K}{\gamma_{\mathsf{air}}}}\left[\frac{1}{1+\frac{K}{\gamma_{\mathsf{air}}}\lambda_{\mathbf{C},\min}}\right]\\ &\overset{(a)}{=} 1-\frac{\exp(r^{-1})}{r}\cdot\mathsf{E}_1\left(r^{-1}\right)\\ &\overset{(b)}{\geq} 1-\frac{\ln\left(1+r\right)}{r},\end{aligned} \tag{31}$$

where the inequality follows from the trace inequalities [43], the $\mathsf{E}_1\left(x\right)$ in step (a) denotes the exponential integral defined as $\mathsf{E}_1\left(x\right)=\int_x^\infty \frac{e^{-t}}{t}\mathrm{d}t$, and the inequality $\mathsf{E}_1\left(x\right)\leq e^{-x}\ln(1+1/x)$ is used in step (b). The effects of the system parameters on A_{loss} are then inferred from (31). Specifically, transmit SNR γ can linearly enlarge r and thereby increases A_{loss} with the scaling of $\frac{\ln(1+\gamma)}{\gamma}$. In view aggregation, letting the number of antennas up faster than K can give a large value of $\omega = N/K$. If $\omega \gg 1$, $(1+\sqrt{\omega})^2 \approx \omega$, A_{loss} increases w.r.t. ω at the rate of $\frac{\ln(1+\omega)}{\omega}$.

V. AirComp or Analog Orthogonal Access?

With limited receive antennas ($N \leq K$), AirComp enables simultaneous access while (spatial division) orthogonal access is infeasible. On the other hand, with a large receive array, both schemes are feasible. In this section, we benchmark AirComp against analog orthogonal access. The study leads to the development of a new scheme supporting dynamic access-mode switching.

A. Performance of Analog Orthogonal Access

The effective feature map received by using analog orthogonal access can be extracted from the real part of $\tilde{\mathbf{f}}$ in (10): $\tilde{\mathbf{f}}^{\mathsf{re}} = \Re\left\{\tilde{\mathbf{f}}\right\} = \bar{\mathbf{f}} + \frac{\nu}{K\sqrt{P}}\Re\left\{\mathbf{Z}^H\sum_k \mathbf{b}_k\right\}$. It shows that the analog orthogonal access introduces Gaussian channel noise into the ground-truth features with the covariance of $\frac{1}{\gamma_{\mathrm{aoa}}}\mathbf{I}_M$ and $\gamma_{\mathrm{aoa}} \triangleq \frac{\gamma K^2}{\nu^2\sum_k\|\mathbf{b}_k\|^2}$ being the effective receive SNR. The result is similar to the case

of AirComp except for the changed effective SNR. This allows the sensing uncertainty in (24) to be modified for analog orthogonal access as

$$H_{\mathsf{s}}(\gamma_{\rm aoa}) = \log\left[1 + (L-1)\exp\left(-\kappa\tilde{D}(\gamma_{\rm aoa})K\right)\right], \tag{32}$$

where $\tilde{D}(\gamma_{\rm aoa}) = \mathsf{Tr}\left(\bar{\mathbf{P}}\left(\mathbf{C} + \frac{K}{\gamma_{\rm aoa}}\mathbf{I}_M\right)^{-1}\bar{\mathbf{P}}\mathbf{D}\right)$.

B. Crossing Point and Access Mode Switching

Comparing the uncertainty functions in (32) and (24) and using their monotonicity (see Lemma 4), we can infer that AirComp outperforms analog orthogonal access if $\gamma_{\rm air} \geq \gamma_{\rm aoa}$, and vice versa. This motivates us to propose the scheme of adaptive access-mode switching, referring to as *adaptive access*, as

$$\text{Multi-access mode} = \begin{cases} \text{AirComp}, & \gamma_{\rm air} \geq \gamma_{\rm aoa}, \\ \text{Anal. orthog. access}, & \gamma_{\rm air} < \gamma_{\rm aoa}. \end{cases} \tag{33}$$

Given adaptive access, the dependence of optimal access mode on system parameters N and K is understood in the sequel. To this end, a useful result on the distribution of the square norm of the ZF beamformers in analog orthogonal access, $\|\mathbf{b}_k\|^2$, is derived as follows.

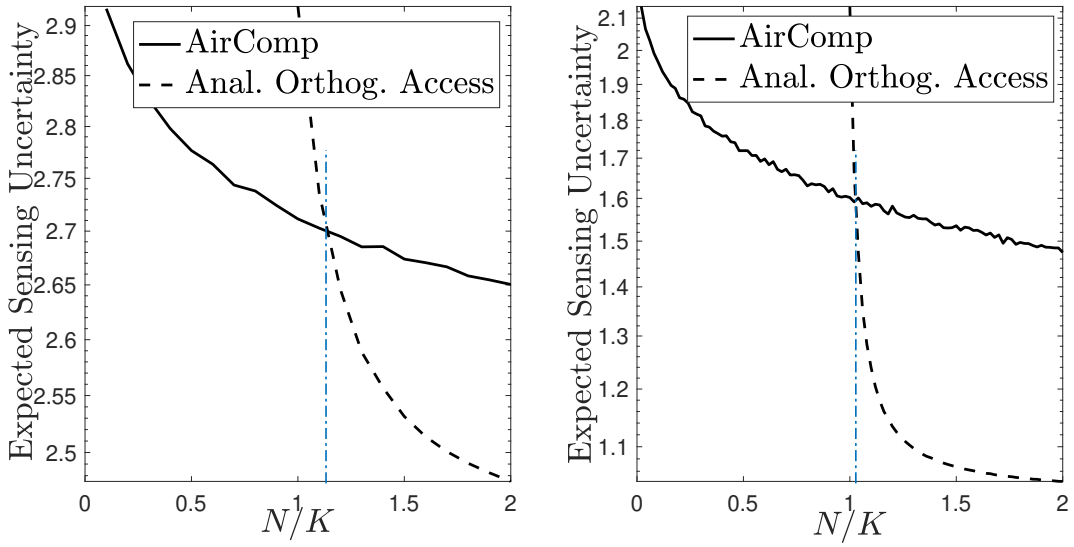
Lemma 6. Given $\mathbf{b}_k = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_k$ in analog orthogonal access, $\|\mathbf{b}_k\|^2$ follows an i.i.d. distribution of $\|\mathbf{b}_k\|^2 \sim 2 \cdot \text{Inv}{-}\chi^2_{2(N-K+1)}$, where $\text{Inv}{-}\chi^2_{2(N-K+1)}$ denotes the inverse chi-square distribution with $2(N-K+1)$ degrees of freedom.

Proof. (See Appendix G). □

Using Lemma 6 and the law of large numbers, the effective receive SNR for analog orthogonal access can be approximated as

$$\frac{\gamma_{\rm aoa}}{K} = \frac{\gamma}{\nu^2 \frac{1}{K}\sum_k \|\mathbf{b}_k\|^2} \approx \frac{\gamma(N-K)}{\nu^2}, \qquad K \gg 1. \tag{34}$$

It follows from (26) and (34) that if $N = K$, $\frac{\gamma_{\rm aoa}}{K} \approx 0$ and $\frac{\gamma_{\rm air}}{K} \geq 0$, leading to the event, $\gamma_{\rm air} \geq \gamma_{\rm aoa}$, occurring with probability 1. On the other hand, if $N > K$, the probability of this



(a) $K = 10$ (b) $K = 50$

Fig. 4. Numerical validation on the crossing point between AirComp and analog orthogonal access. The parameters are set as $M = L = 10$, $\mathbf{C} = 0.1\mathbf{I}_M$, $\mathsf{rank}(\mathbf{P}_k) = 1$, $\gamma = 10$ dB. The expectation is taken over channel distribution.

event is asymptotically close to 0 since using (26) and (34)

$$\begin{aligned}\Pr\left(\gamma_{\text{air}} \geq \gamma_{\text{aoa}}\right) &= \Pr\left(\frac{\gamma_{\text{air}}}{K} \geq \frac{K\gamma(\omega-1)}{\nu^2}\right) \\ &= \exp\left(-\frac{K}{2}\frac{\sqrt{\omega}-1}{\sqrt{\omega}+1}\right) \to 0, \text{ as } K \to \infty. \end{aligned} \tag{35}$$

Combining the above results gives the following conclusion.

Main Result 3 (Mode Switching Point)**.** There exists a crossing point of the sensing uncertainty between AirComp and analog orthogonal access. Given $N = \omega K$ and $K >> 1$, AirComp outperforms analog orthogonal access for the case of $\omega \leq 1$ (i.e., $N \leq K$); otherwise, the reverse holds. In other words, the crossing point is around $N = K$. This conclusion is numerically validated in Fig. 4.

VI. Experimental Results

A. Experimental Settings

Consider the ISEA system as shown in Fig. 1. Assuming frequency non-selective Rayleigh fading, the multi-access channel is composed of i.i.d. Gaussian $\mathcal{N}(0, 1)$ elements. The coherence

duration of the channel spreads over 256 symbol slots, supporting analog transmission of feature vector with the maximum length of 256. For the MVCNN architecture, we consider both the cases of linear classification on synthetic data and CNN-based classification on real-world data as follows.

- *Linear classification on synthetic data*: Local feature maps are drawn from the GMM in (3) and fed into the classifier given in (22) after over-the-air averaging via AirComp. The feature maps' dimensionality is $M = 100$, the number of classes is $L = 20$, and the covariance matrix $\mathbf{C} = 0.1\mathbf{I}_M$. The observation matrices $\{\mathbf{P}_k\}$ are randomly generated as the principal eigenspaces of random matrices with i.i.d. Gaussian entries. For instance, let $\mathbf{G}$ be a randomly generated $M \times M$ Gaussian matrix. Then, $\mathbf{P}_k = \mathbf{U}_\mathbf{G}\mathbf{U}_\mathbf{G}^\top$ with $\mathbf{U}_\mathbf{G}$ being the $\mathsf{rank}(\mathbf{P}_k)$-dimensional principal eigenspace of $\mathbf{G}$.
- *MVCNN-based classification on real-world data*: We consider the well-known *ModelNet* dataset which comprises multi-view images of objects (e.g., sofas and tables) and the popular VGG11 model for implementing the MVCNN architecture. The VGG11 is split before the linear classifier with the classifier employed at the server and the other components deployed at each sensor for feature extraction [6]. The resultant MVCNN architecture is trained for average pooling. Therein, we select a data subsect of ModelNet corresponding to $L = 10$ popular object classes for our experiments. The data entries for each target are captured by $K = 12$ sensors (i.e., cameras) with the angle between adjacent sensors being 30°. Each feature map output from an on-device model is described as a $512 \times 7 \times 7$ tensor, where the 512×7 slices of these feature tensors are transmitted and aggregated sequentially at the server for global classification.

Last, to evaluate the performance of AirComp, we adopt two benchmarking schemes, namely analog orthogonal access in (10) and the adaptive (dual-mode) access in (33), to support local feature uploading.

B. ISEA with Linear Classification

The curves of E2E sensing accuracy and uncertainty versus number of sensors, K, are plotted in Fig. 5. Different levels of local observation DoFs are considered: $\mathsf{rank}(\mathbf{P}_k) = \{0.5M, 0.7M\}$ for all k. First, it can be observed from Fig. 5 that the E2E sensing uncertainty diminishes at an exponential rate as K grows. This is consistent with the main analytical results in (30). On the other hand, the E2E sensing accuracy converges to the saturation level (i.e., maximum accuracy)

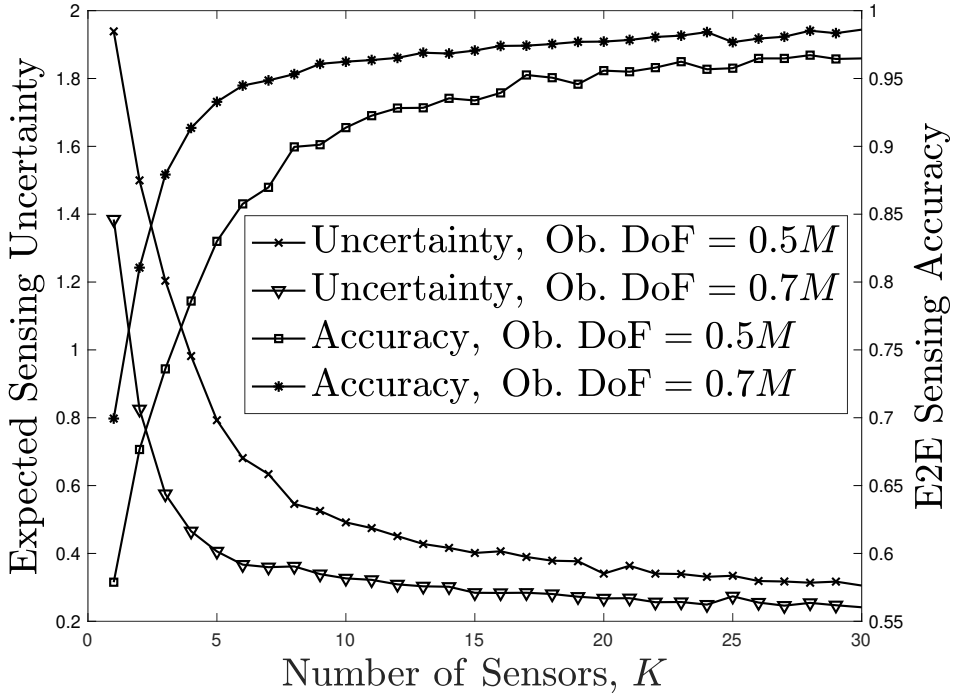
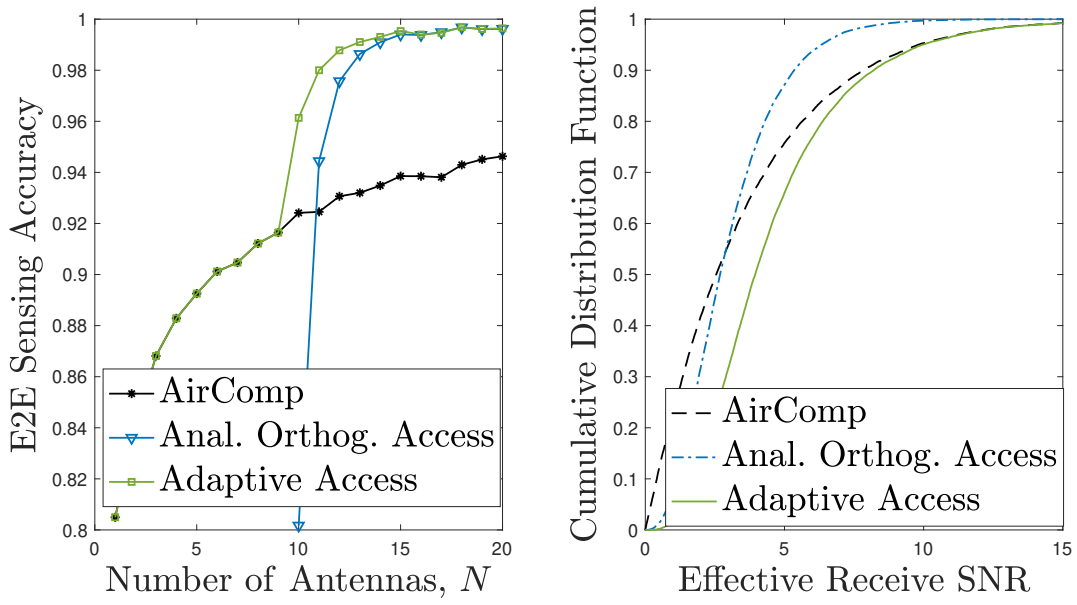


Fig. 5. (Linear classification) Comparison between E2E sensing uncertainty and accuracy for a variable number of sensors and different local observation DoFs.

also exponentially fast. The consistency of uncertainty and accuracy validates their duality. One can also observe the existence of a critical range (i.e., $K \leq 10$) where the sensing performance is sensitive to changes on sensor number. Last, increasing local observation DoFs is found to yield significant gains on sensing performance.

In Fig. 6, the performance of ISEA using AirComp is compared with that of counterparts employing benchmarking access schemes. Specifically, Fig. 6(a) depicts the curves of E2E sensing accuracy versus receive array size, N; the cumulative distribution function (CDF) curves of effective receive SNR are plotted in Fig. 6(b). The number of sensors is fixed as $K = 10$. The most important observation can be made from Fig. 6(a) that there exists a crossing point between AirComp and analog orthogonal access at around $N = K$. This validates the Main Result 3. Next, the adaptive access scheme designed in Section V is observed to be effective as it outperforms the other two underpinning schemes. The above observations are consistent with those from Fig. 6(b), i.e., the crossing point of SNR CDF curves and superiority of adaptive access in terms of effective receive SNR.



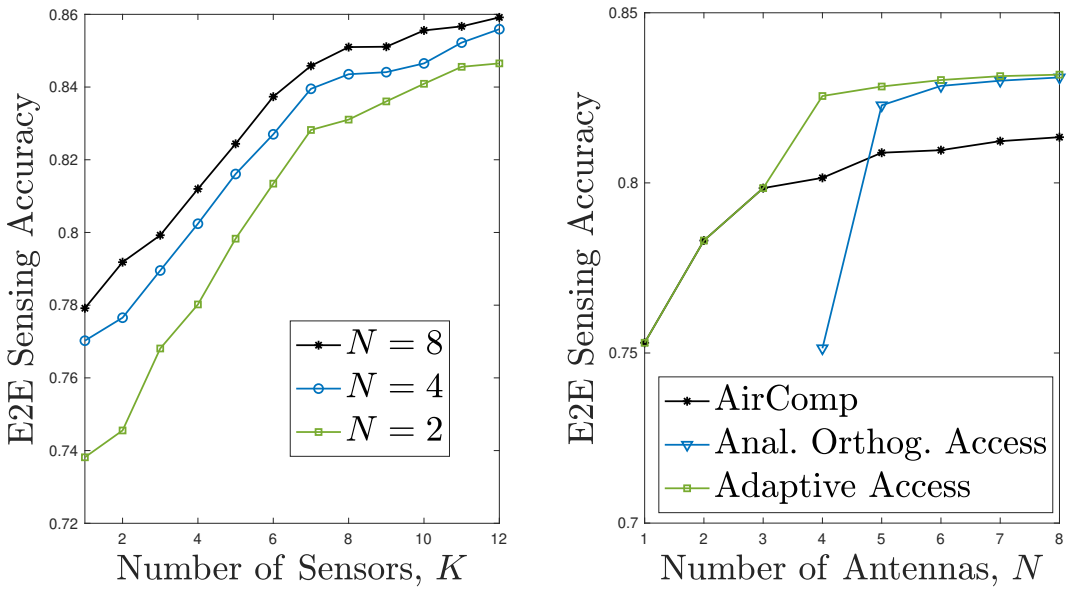
(a) (b)

Fig. 6. (Linear classification) Performance comparison between AirComp, analog orthogonal access, and adaptive access in terms of (a) E2E sensing accuracy and (b) effective receive SNR with $N = 12$.

C. MVCNN-based Classification

Experimental results for the case of MVCNN based classification are presented to validate the insights from our analysis based on linear classification. Specifically, the MVCNN counterparts of the E2E sensing accuracy curves in Fig. 5 and those in Fig. 6(a) are obtained as plotted in Fig. 7(a) and Fig. 7(b), respectively. The main observations from the MVCNN curves are lowly identical to those from their linear-classification counterparts. Specifically, the exponential convergence of E2E sensing accuracy is reflected in Fig. 7(a); the crossing point between AirComp and analog orthogonal access is found in Fig. 7(b) to be also around $N = K$.

Last, the curves of E2E sensing accuracy versus transmit SNR are plotted in Fig. 8 for different values of (N, K). The main observation is that the close-to-maximum accuracy is achievable even at very low transmit SNR (e.g., -5 dB). As mentioned early, the reason is view-and-channel aggregation gain in two aspects. First, the aggregation gain enhances the receive SNR by a factor approximately equal to K. In other words, $K = 10$ can achieve the effective receive SNR of 10 dB even given transmit SNR as low as 0 dB. The other aspect is that view aggregation improves the model's classification margin (see, e.g., [6]) to absorb channel distortion without compromising sensing accuracy. The above observation advocates the use of AirComp



(a) (b)

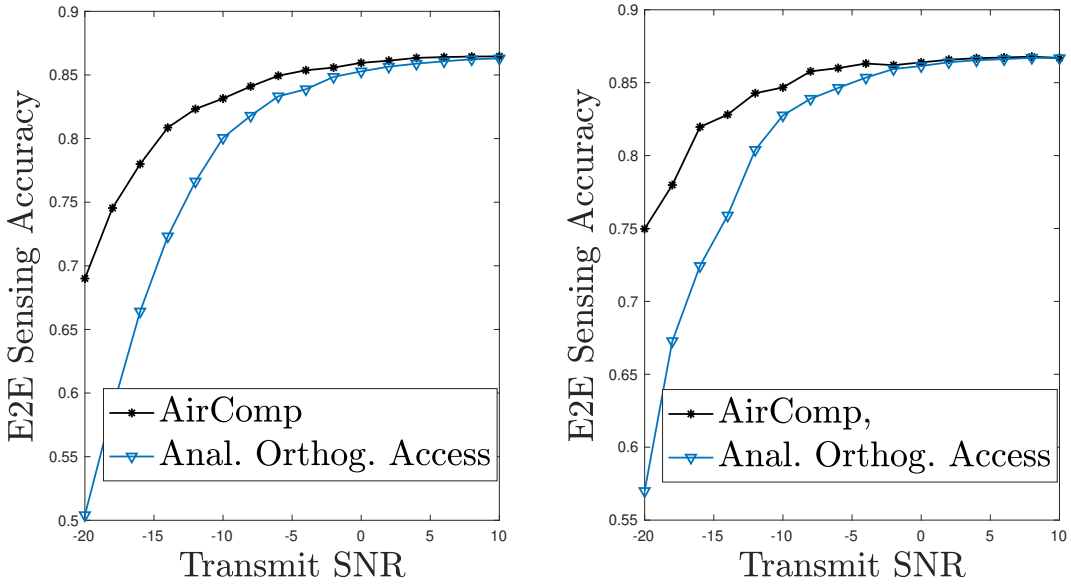
Fig. 7. (MVCNN classification) The dependence of E2E sensing accuracy on (a) the numbers of sensors and (b) receive antennas under transmit SNR $\gamma = -10$ dB.

and uncoded analog transmission at large to support fast ISEA.

VII. Conclusion

We have presented a theoretical framework for characterizing the performance gains from view-and-channel aggregation in an ISEA system. Our results reveal that the sensing/inference uncertainty decreases exponentially with the increasing number of views/sensors, with the rate being directly proportional to the global discriminant gain. Furthermore, it is demonstrated that the channel distortion resulting from aggregation via AirComp or analog orthogonal access does not alter this scaling law, except for a reduction in the exponential rate. Utilizing the end-to-end performance analysis, we have also developed a scheme for aggregation mode adaption that dynamically switches between AirComp and orthogonal analog access to achieve optimal system performance.

Expanding upon the insights from this study, we anticipate that distributed sensing, leveraging its multi-view aggregation gain, will emerge as a mainstream direction in the area of ISEA. Advancing this direction calls for novel quantitative analysis and protocol designs that strike a balance between access control, latency, and computation accuracy so as to improve the end-to-



(a) $K = 4$ (b) $K = 8$

Fig. 8. (MVCNN classification) The effect of transmit SNR on the E2E sensing accuracy for the number of sensors $K = N = \{4, 8\}$.

end performance. This study serves as an initial step in establishing a theoretical framework for the advancement of ISEA. Extending this framework to incorporate other wireless techniques (such as broadband transmission and random access) and other sensing scenarios (including sensing via point clouds and multi-modal fusion) warrant further investigation.

APPENDIX

A. *Proof of Lemma 1*

Using (2), the aggregated feature map can be rewritten as

$$\bar{\mathbf{f}} = \frac{1}{K}\sum_k (\mathbf{P}_k\mathbf{g} + \mathbf{w}_k) = \bar{\mathbf{P}}\mathbf{g} + \frac{1}{K}\sum_k \mathbf{w}_k,$$

where $\bar{\mathbf{P}} = \frac{1}{K}\sum_k \mathbf{P}_k$. According to (1), the first term $\bar{\mathbf{P}}\mathbf{g}$ has a distribution of $\Pr(\bar{\mathbf{P}}\mathbf{g} = \bar{\mathbf{P}}\boldsymbol{\mu}_\ell) = \frac{1}{L},\ \forall \ell$. At the same time, the summation of the i.i.d. Gaussian sensing noise, $\frac{1}{K}\sum_k \mathbf{w}_k$, follows a Gaussian distribution, say $\frac{1}{K}\sum_k \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \frac{1}{K}\mathbf{C})$. Hence, the overall distribution of $\bar{\mathbf{f}}$ is a Gaussian mixture with the same priors and the ℓ-th Gaussian component has the mean $\bar{\mathbf{P}}\boldsymbol{\mu}_\ell$ and covariance $\mathbf{C}/K$, which completes the proof.

B. Proof of Proposition 1

Based on the definition given in (11), the sensing uncertainty using the aggregated feature $\bar{\mathbf{f}}$ is given as

$$H = \mathsf{E}\left[-\sum_{\ell=1}^{L}\int \Pr\left(\boldsymbol{\mu}_\ell|\bar{\mathbf{f}}\right)\log\Pr\left(\boldsymbol{\mu}_\ell|\bar{\mathbf{f}}\right)p(\bar{\mathbf{f}})\mathrm{d}\bar{\mathbf{f}}\right].$$

Given the output uniqueness of the linear classifier in (5), there is a one-to-one mapping between $\boldsymbol{\mu}_\ell$ and $\bar{\mathbf{P}}\boldsymbol{\mu}_\ell$, leading to $\Pr\left(\boldsymbol{\mu}_\ell|\bar{\mathbf{f}}\right) = \Pr\left(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell|\bar{\mathbf{f}}\right)$. Then, using the Bayes' theorem, the probability $\Pr\left(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell|\bar{\mathbf{f}}\right)$ can be expressed as $\Pr\left(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell|\bar{\mathbf{f}}\right) = p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)p(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell)/p(\bar{\mathbf{f}})$, which is used to rewrite H as

$$\begin{aligned}
H &= -\sum_{\ell=1}^{L}\int p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)p(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell)\log\frac{p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)p(\bar{\mathbf{P}}\boldsymbol{\mu}_\ell)}{p(\bar{\mathbf{f}})}\mathrm{d}\bar{\mathbf{f}} \\
&\overset{(a)}{=} \frac{1}{L}\sum_{\ell=1}^{L}\int p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)\log\frac{\sum_{\ell'}p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_{\ell'}\right)}{p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)}\mathrm{d}\bar{\mathbf{f}} \\
&\overset{(b)}{=} \frac{1}{L}\sum_{\ell=1}^{L}\int p\left(\mathbf{x}\right)\log\frac{\sum_{\ell'}e^{\frac{-K}{2}(\mathbf{x}+\bar{\mathbf{P}}\boldsymbol{\phi}_{\ell,\ell'})^\top\mathbf{C}^{-1}(\mathbf{x}+\bar{\mathbf{P}}\boldsymbol{\phi}_{\ell,\ell'})}}{e^{\frac{-K}{2}\mathbf{x}^\top\mathbf{C}^{-1}\mathbf{x}}}\mathrm{d}\mathbf{x},
\end{aligned}$$

where the PDF $p\left(\bar{\mathbf{f}}|\bar{\mathbf{P}}\boldsymbol{\mu}_\ell\right)$ is used in step (a), the integration variable is changed as $\mathbf{x} = \bar{\mathbf{f}} - \boldsymbol{\mu}_\ell$ with the resulting $p\left(\mathbf{x}\right) = \mathcal{N}\left(\mathbf{0}, \frac{\mathbf{C}}{K}\right)$, $\boldsymbol{\phi}_{\ell,\ell'} = \boldsymbol{\mu}_\ell - \boldsymbol{\mu}_{\ell'}$ in step (b). Then, the lower and upper bounds of H are given respectively as follows.

1) Lower bound: First, using the definition of $G_{\ell,\ell'}$ in Lemma 2, the sensing uncertainty is given as $H = \frac{1}{L}\sum_{\ell=1}^{L}\int p\left(\mathbf{x}\right)\log\left[1+\sum_{\ell'\neq\ell}e^{-K\mathbf{x}^\top\mathbf{C}^{-1}\bar{\mathbf{P}}\boldsymbol{\phi}_{\ell,\ell'}}e^{-\frac{1}{2}G_{\ell,\ell'}}\right]\mathrm{d}\mathbf{x}$. Then, based on the convexity of log-sum-exp functions, using Jensen's inequality gives a lower bound of H as

$$H \geq \frac{1}{L}\sum_{\ell=1}^{L}\log\left[1+\sum_{\ell'\neq\ell}e^{-\int p(\mathbf{x})K\mathbf{x}^\top\mathbf{C}^{-1}\bar{\mathbf{P}}\boldsymbol{\phi}_{\ell,\ell'}\mathrm{d}\mathbf{x}}e^{-\frac{1}{2}G_{\ell,\ell'}}\right],$$

where the integral in the exponential term is computed as zero and the final lower bound is obtained.

2) Upper bound: To derive the upper bound of the sensing uncertainty, we first define rewrite H derived before as

$$\begin{aligned} H =& \frac{1}{L}\sum_{\ell=1}^{L}\int p\left(\mathbf{x}\right)\log\frac{\sum_{\ell'} e^{\frac{-K}{2}(\mathbf{x}+\bar{\mathbf{P}}\boldsymbol{\phi}_{\ell,\ell'})^{\top}\mathbf{C}^{-1}(\mathbf{x}+\bar{\mathbf{P}}\boldsymbol{\phi}_{\ell,\ell'})}}{e^{\frac{-K}{2}\mathbf{x}^{\top}\mathbf{C}^{-1}\mathbf{x}}}\mathrm{d}\mathbf{x} \\ \leq& \frac{1}{L}\sum_{\ell=1}^{L}\log\left[1+\sum_{\ell'\neq\ell}\exp\left(-\frac{1}{a+2}G_{\ell,\ell'}\right)\right] \\ &+\frac{M}{a}\log e-\frac{M}{2}\log(1+2/a), \end{aligned}$$

where a is a positive constant and the inequality is obtained by using the log-concavity and Jensen's inequality. Then, let $a = cM$ with $c > 0$, there is $\frac{M}{a}\log e - \frac{M}{2}\log(1+2/a) \leq \log\frac{e^{\frac{1}{c}}}{1+\frac{2}{cM}\frac{M}{2}} = \log\frac{e^{\frac{1}{c}}}{1+\frac{1}{c}}$, where the second step is based on the well-known Bernoulli's inequality. This gives the upper bound of H.

C. Proof of Proposition 2

Let $\{D_{\ell,\ell'}\}$ be aggregated into a $L(L-1)$-dimensional vector $\mathbf{u}$ and define the function $U_\kappa(K,\mathbf{u}) = \frac{1}{L}\sum_{\ell=1}^{L}\log\left[1+\sum_{\ell'\neq\ell}\exp\left(-\kappa D_{\ell,\ell'}K\right)\right]$ for simplicity. Using the first-order approximation, $U_\kappa(K,\mathbf{u})$ can be rewritten as $U_\kappa(K,\mathbf{u}) = U_\kappa(K,\mathbf{c}) + (\mathbf{u}-\mathbf{c})^{\top}\nabla_{\mathbf{u}}U_\kappa(K,\mathbf{u})|_{\mathbf{u}=\mathbf{c}} + \mathcal{O}\left((\mathbf{u}-\mathbf{c})^{\top}\mathcal{H}_{\mathbf{u}}U_\kappa(K,\mathbf{u})|_{\mathbf{u}=\mathbf{c}}(\mathbf{u}-\mathbf{c})\right)$, where $\mathbf{c}$ denotes an arbitrary constant vector, $\nabla_{\mathbf{u}}U_\kappa(K,\mathbf{u})$ and $\mathcal{H}_{\mathbf{u}}U_\kappa(K,\mathbf{u})$ denote the derivative and Hessian of $U_\kappa(K,\mathbf{u})$ w.r.t. $\mathbf{u}$, respectively. Then, the gradient values of $U_\kappa(K,\mathbf{u})$ are expressed as

$$\frac{\partial U_\kappa(K,\mathbf{u})}{\partial D_{\ell,\ell'}} = -\frac{\kappa K}{L}\frac{\exp\left(-\kappa D_{\ell,\ell'}K\right)}{1+\sum_{l\neq\ell}\exp\left(-\kappa D_{\ell,l}K\right)}.$$

Hence, let the constant vector $\mathbf{c}$ be $\mathbf{c} = \bar{D}\mathbf{1}$ with $\bar{D}$ denoting the average version of $\{D_{\ell,\ell'}\}$ as stated before. There is then $\frac{\partial U_\kappa(K,\mathbf{u})}{\partial D_{\ell,\ell'}}|_{D_{\ell,\ell'}=\bar{D}} = -\frac{\kappa K}{L}\frac{\exp\left(-\kappa\bar{D}K\right)}{1+\sum_{l\neq\ell}\exp\left(-\kappa\bar{D}K\right)}$, based on which the residual term can be further written as $\mathcal{O}\left((\mathbf{u}-\mathbf{c})^{\top}\mathcal{H}_{\mathbf{u}}U_\kappa(K,\mathbf{u})|_{\mathbf{u}=\mathbf{c}}(\mathbf{u}-\mathbf{c})\right) = \mathcal{O}(\frac{1}{L(L-1)}\sum_{\ell}\sum_{\ell\neq\ell'}(D_{\ell,\ell'}-\bar{D})^2)$. It then follows that $U_\kappa(K,\mathbf{u})$ can be first-order approximated as

$$\begin{aligned} U_\kappa(K,\mathbf{u}) =& U_\kappa(K,\bar{D}\mathbf{1}) + \underbrace{(\mathbf{u}-\bar{D}\mathbf{1})^{\top}\nabla_{\mathbf{u}}U_\kappa(K,\mathbf{u})|_{\mathbf{u}=\bar{D}\mathbf{1}}}_{=0} \\ &+\mathcal{O}(\frac{1}{L(L-1)}\sum_{\ell}\sum_{\ell\neq\ell'}(D_{\ell,\ell'}-\bar{D})^2). \end{aligned}$$

Based on the definition, $U_\kappa(K, \bar{D}\mathbf{1})$ can be expressed as $U_\kappa(K, \bar{D}\mathbf{1}) = \frac{1}{L}\sum_{\ell=1}^{L} \log\left[1 + \sum_{\ell' \neq \ell} \exp\left(-\kappa \bar{D} K\right)\right]$ that gives the final result.

D. Proof of Lemma 3

Using (2) and (8), the efficient aggregated feature map can be rewritten as

$$\tilde{\mathbf{f}} = \bar{\mathbf{P}}\mathbf{g} + \frac{1}{K}\sum_{k} \mathbf{w}_k + \tilde{\mathbf{z}},$$

where the noise term $\tilde{\mathbf{z}} = \frac{1}{K}\sum_k \mathbf{w}_k + \frac{1}{K}\Re\{\mathbf{Z}^H\mathbf{b}\}$ is a summation of real independent Gaussian vectors and thus has a Gaussian distribution. Since both $\mathbf{w}_k$ and $\mathbf{Z}^\top$ are zero-mean, there is $\mathsf{E}\left[\tilde{\mathbf{z}}\right] = \mathbf{0}$. Furthermore, $\mathbf{w}_k$ and $\mathbf{Z}^\top$ are mutually independent, leading the covariance matrix of $\tilde{\mathbf{z}}$ computed as

$$\begin{aligned}\mathsf{E}\left[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^\top\right] &= \frac{1}{K^2}\mathsf{E}\left[\sum_{k,j}\mathbf{w}_k\mathbf{w}_j^\top\right] + \frac{1}{K^2}\mathsf{E}\left[\Re\{\mathbf{Z}^H\mathbf{b}\}\Re\{\mathbf{b}^\top\mathbf{Z}^*\}\right] \\ &= \frac{1}{K}\mathbf{C} + \frac{\sigma^2}{2K^2}\|\mathbf{b}\|^2.\end{aligned}$$

Still, there is $p(\bar{\mathbf{P}}\mathbf{g} = \bar{\mathbf{P}}\boldsymbol{\mu}_\ell) = \frac{1}{L},\ \forall \ell$, meaning that $\tilde{\mathbf{f}}^{\mathsf{re}}$ follows a GMM with the uniform priors and the ℓ-th Gaussian component having the mean $\bar{\mathbf{P}}\boldsymbol{\mu}_\ell$ and covariance $\frac{1}{K}\mathbf{C} + \frac{\sigma^2\|\mathbf{b}\|^2}{2K^2}\mathbf{I}_M$.

E. Proof of Lemma 4

Straightforwardly, $H_{\mathsf{s}}(\gamma_{\mathsf{air}})$ is monotonically decreasing w.r.t. $\tilde{D}(\gamma_{\mathsf{air}})$. Hence, we prove the Lemma 4 by showing that the derivative of $\tilde{D}(\gamma_{\mathsf{air}})$ w.r.t. γ_{air} is always positive given any $\mathbf{b}$. To this end, the chain rule of matrix derivative is used as

$$\frac{\partial \tilde{D}(\gamma_{\mathsf{air}})}{\partial \gamma_{\mathsf{air}}} = \mathsf{Tr}\left(\frac{\partial \tilde{D}(\gamma_{\mathsf{air}})}{\partial \mathbf{A}} \cdot \left(\frac{\partial \mathbf{A}}{\partial \gamma_{\mathsf{air}}}\right)^\top\right),$$

where the matrix $\mathbf{A} = \left(\mathbf{C} + \frac{K}{\gamma_{\mathsf{air}}}\mathbf{I}_M\right)^{-1}$ for ease of notation. Therein, using the trace property, $\frac{\partial \tilde{D}(\gamma_{\mathsf{air}})}{\partial \mathbf{A}}$ is computed as $\frac{\partial \tilde{D}(\gamma_{\mathsf{air}})}{\partial \mathbf{A}} = \bar{\mathbf{P}}\mathbf{D}\bar{\mathbf{P}}$. The other derivative term, $\frac{\partial \mathbf{A}}{\partial \gamma_{\mathsf{air}}}$, is computed by using the matrix inverse' derivative as $\frac{\partial \mathbf{A}}{\partial \gamma_{\mathsf{air}}} = -\mathbf{A}\frac{\partial \mathbf{A}^{-1}}{\partial \gamma_{\mathsf{air}}}\mathbf{A} = \frac{K}{\gamma_{\mathsf{air}}^2}\mathbf{A}^2$. As a result, we have the derivative $\frac{\partial \tilde{D}(\gamma_{\mathsf{air}})}{\partial \gamma_{\mathsf{air}}} = \frac{K}{\gamma_{\mathsf{air}}^2}\mathsf{Tr}\left(\bar{\mathbf{P}}\mathbf{D}\bar{\mathbf{P}}\mathbf{A}^2\right) = \frac{K}{\gamma_{\mathsf{air}}^2}\mathsf{Tr}\left(\mathbf{A}\bar{\mathbf{P}}\mathbf{D}\bar{\mathbf{P}}\mathbf{A}\right)$, where the argument of the trace function is a semi-positive Hermitian matrix, leading to $\frac{\partial \tilde{D}(\gamma_{\mathsf{air}})}{\partial \gamma_{\mathsf{air}}} \geq 0$. This completes the proof.

F. Proof of Lemma 5

To obtain the distribution of ζ_{air}, we first rewrite the expression of the data vector $\mathbf{s}$ in (8) to rewrite $\|\mathbf{b}^\star\|$. Specifically, since $\mathbf{b}^\star = \|\mathbf{b}^\star\|\mathbf{v}$ with $\mathbf{v}$ being the first eigenvector of the channel matrix $\mathbf{H}$, there is

$$\begin{aligned}\mathbf{s} &= \mathbf{Y}^H\mathbf{b}^\star \\ &= \mathbf{Z}^H\mathbf{b}^\star + \mathbf{X}\mathsf{diag}(\rho_1^*,\cdots,\rho_K^*)\mathbf{q}_1\sqrt{\lambda_1}\|\mathbf{b}^\star\|,\end{aligned}$$

where $\mathbf{X} = [\mathbf{x}_1^H,\cdots,\mathbf{x}_K^H]$, $\mathbf{q}_1$ and λ_1 denotes the first eigenvector and the first eigenvalue of $\mathbf{H}^H\mathbf{H}$. Therefore, to realize AirComp, the transmit power control of sensor k can be rewritten as $\rho_k = \frac{1}{q_{k,1}^*\sqrt{\lambda_1}\|\mathbf{b}^\star\|}$, where $q_{k,1}$ denotes the k-th element of $\mathbf{q}_1$. To minimize $\|\mathbf{b}^\star\|$ under the power constraint $\rho_k^2 \le \frac{P}{\nu^2}$, $\forall k$, there is $\|\mathbf{b}^\star\| = \max_k \ \frac{\nu^2}{P}\frac{1}{q_{k,1}^2\lambda_1} = \frac{\nu^2}{P}\frac{1}{\lambda_1\min_k \ q_{k,1}^2}$. Comparing the above result, the scaled effective channel gain can be re-expressed as $\zeta_{\mathsf{air}} = K\lambda_1\min_k \ q_{k,1}^2$. Hence, ζ_{air} can be characterized by investigating the asymptotic values of λ_1 and $\min_k \ q_{k,1}^2$. Specifically, using the asymptotic spectrum of random matrices [44], given $N = \omega K$, there is

$$\frac{1}{K}\lambda_1 \to (1+\sqrt{\omega})^2, \ K\to\infty.$$

At the same time, since $\mathbf{H}$ is composed of i.i.d. Gaussian elements, the eigenvector $\mathbf{q}_1$ is isotropically distributed on a K-dimensional complex hypersphere and independent of the eigenvalue λ_1. Furthermore, it has the same distribution as a normalized complex Gaussian vector $\mathbf{h} = [h_1,\cdots,h_K]$ with $h_k \sim \mathcal{CN}(0,1)$, say $\mathbf{q}_1 \overset{d}{=} \frac{\mathbf{h}}{\|\mathbf{h}\|}$. Therefore, we have

$$K\cdot\min_k \ q_{k,1}^2 \overset{d}{=} \frac{K}{\|\mathbf{h}\|^2}\min_k \ h_k^2,$$

where $\frac{K}{\|\mathbf{h}\|^2} = 1$ as $K\to\infty$, which means that $K\cdot\min_k \ q_{k,1}^2$ asymptotically equals to $\min_k \ h_k^2$. Clearly, $2h_k^2$ is a chi-square random variable with the degrees of freedom 2 and is also exponentially distributed with the parameter of $\frac{1}{2}$, say $h_k^2 \sim \mathrm{Exp}(\frac{1}{2})$. Using the order statistics of exponential distributions, we have $\min_k \ 2h_k^2 \sim \mathrm{Exp}(\frac{K}{2})$. Concluding the above results gives the asymptotic distribution of $\zeta_{\mathsf{air}} \to \frac{1}{2}K(1+\sqrt{\omega})^2\zeta'$ with $\zeta' \sim \mathrm{Exp}\left(\frac{K}{2}\right)$. Finally, using the scaling property of exponential distributions obtains the distribution of ζ_{air}.

G. Proof of Lemma 6

To obtain the distribution of the norm of the k-th ZF receiver, let $\mathbf{U}_k$ be the $(K-1)$-dimensional principal eigenspace of the matrix $[\mathbf{h}_1,\cdots,\mathbf{h}_{k-1},\mathbf{h}_{k+1},\cdots,\mathbf{h}_K]$ and $\|\mathbf{b}_k\|^2$ can be rewritten as

$\|\mathbf{b}_k\|^2 = \mathbf{e}_k^H(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{e}_k = \frac{1}{\mathbf{h}_k\mathbf{U}_k^{\perp}(\mathbf{U}_k^{\perp})^H\mathbf{h}_k}$, where $\mathbf{U}_k^{\perp}$ denotes the orthogonal compliment of $\mathbf{U}_k$ and $(\mathbf{U}_k^{\perp})^H\mathbf{U}_k^{\perp} = \mathbf{I}_{N-K+1}$. Since $\mathbf{h}_k$ is composed of i.i.d. $\mathcal{CN}(0,1)$ elements, it is isotropically distributed on the N-dimensional complex unit hypersphere, i.e., $p(\mathbf{h}_k) = p(\mathbf{Q}\mathbf{h}_k)$ for any $\mathbf{Q}$ satisfying $\mathbf{Q}^H\mathbf{Q} = \mathbf{Q}\mathbf{Q}^H = \mathbf{I}_N$. Hence, let $\mathbf{Q} = [\mathbf{U}_k, \mathbf{U}_k^{\perp}]$ and

$$\mathbf{h}_k\mathbf{U}_k^{\perp}(\mathbf{U}_k^{\perp})^H\mathbf{h}_k \overset{d}{=} \mathbf{h}_k^H\mathbf{Q}^H\mathbf{U}_k^{\perp}(\mathbf{U}_k^{\perp})^H\mathbf{Q}\mathbf{h}_k = \sum_{k=K}^{N} h_{k,n}^2.$$

Since $h_{k,n} \sim \mathcal{CN}(0,1)$, the summation $2\sum_{k=K}^{N} h_{k,n}^2$ follows a chi-square distribution with $2(N-K+1)$ degrees of freedom, which completes the proof.

REFERENCES

[1] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6G: Vision, enabling technologies, and applications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 5–36, Jan. 2022.

[2] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Toward an intelligent edge: Wireless communication meets machine learning," *IEEE Commun. Mag.*, vol. 58, no. 1, p. 19–25, 2020.

[3] Z. Feng, Z. Wei, X. Chen, H. Yang, Q. Zhang, and P. Zhang, "Joint communication, sensing, and computation enabled 6G intelligent machine system," *IEEE Netw.*, vol. 35, no. 6, pp. 34–42, 2021.

[4] L. Xie, S. Song, Y. C. Eldar, and K. B. Letaief, "Collaborative sensing in perceptive mobile networks: Opportunities and challenges," *IEEE Wireless Commun.*, vol. 30, no. 1, pp. 16–23, 2023.

[5] S. Duan, D. Wang, J. Ren, F. Lyu, Y. Zhang, H. Wu, and X. Shen, "Distributed artificial intelligence empowered by end-edge-cloud computing: A survey," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 591–624, Firstquarter 2023.

[6] Z. Liu, Q. Lan, A. E. Kalør, P. Popovski, and K. Huang, "Over-the-air multi-view pooling for distributed sensing," [Online] https://arxiv.org/pdf/2302.09771.pdf, 2023.

[7] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, Santiago, Chile, Dec. 7–13 2015.

[8] L. Bai, Y. Yang, M. Chen, C. Feng, C. Guo, W. Saad, and S. Cui, "Computer vision-based localization with visible light communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 2051–2065, Mar. 2022.

[9] Y. Li, S. Ren, P. Wu, S. Chen, C. Feng, and W. Zhang, "Learning distilled collaboration graph for multi-agent perception," in *Proc. Adv. in Neural Inf. Process. Syst. (NeurIPS)*, Virtual-only, Dec. 6-12 2021.

[10] J. Shao, Y. Mao, and J. Zhang, "Task-oriented communication for multi-device cooperative edge inference," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, p. 73–87, Jan. 2023.

[11] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge AI: On-demand accelerating deep neural network inference via edge computing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, p. 447–457, 2020.

[12] W. Shi, S. Zhou, Z. Niu, M. Jiang, and L. Geng, "Multiuser co-inference with batch processing capable edge server," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 286–300, 2023.

[13] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, 2019.

[14] X. Wei, R. Yu, and J. Sun, "Learning view-based graph convolutional network for multi-view 3D shape analysis," *IEEE Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7525–7541, 2023.

[15] K. Senel and E. G. Larsson, "Grant-free massive MTC-enabled massive MIMO: A compressive sensing approach," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6164–6175, 2018.

[16] L. Liu, E. G. Larsson, W. Yu, P. Popovski, C. Stefanovic, and E. de Carvalho, "Sparse signal processing for grant-free massive connectivity: A future paradigm for random access protocols in the internet of things," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 88–99, 2018.

[17] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016.

[18] 3GPP, "Release 18: Study on traffic characteristics and performance requirements for AI/ML model transfer," *3GPP TR 22.874, version 18.2.0*, May 2021.

[19] Y.-C. Liu, J. Tian, N. Glaser, and Z. Kira, "When2com: Multi-agent perception via communication graph grouping," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recogn. (CVPR)*, Seattle, WA, USA, Jun. 26 – 29 2020.

[20] M. Singhal, V. Raghunathan, and A. Raghunathan, "Communication-efficient view-pooling for distributed multi-view neural networks," in *Proc. Des. Autom. Test Eur. Conf. & Exh. (DATE)*, Grenoble, France, Mar. 9–13 2020.

[21] G. Zhu and K. Huang, "MIMO over-the-air computation for high-mobility multimodal sensing," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6089–6103, 2019.

[22] M. Mohammadi Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," *IEEE Trans. Signal Process.*, vol. 68, pp. 2155–2169, 2020.

[23] S. F. Yilmaz, B. Hasırcıoglu, and D. Gündüz, "Over-the-air ensemble inference with model privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Espoo, Finland, Jun. 26 – Jul. 1 2022.

[24] D. Wen, X. Jiao, P. Liu, G. Zhu, Y. Shi, and K. Huang, "Task-oriented over-the-air computation for multi-device edge AI," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[25] X. Chen, E. G. Larsson, and K. Huang, "Analog MIMO communication for one-shot distributed principal component analysis," *IEEE Trans. Signal Process.*, vol. 70, pp. 3328–3342, 2022.

[26] T. Marzetta and B. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, 2006.

[27] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 3579–3605, 2021.

[28] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, 2020.

[29] T. Sery, N. Shlezinger, K. Cohen, and Y. C. Eldar, "Over-the-air federated learning from heterogeneous data," *IEEE Trans. Signal Process.*, vol. 69, pp. 3796–3811, 2021.

[30] N. Zhang and M. Tao, "Gradient statistics aware power control for over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5115–5128, 2021.

[31] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 491–506, 2020.

[32] X. Li, F. Liu, Z. Zhou, G. Zhu, S. Wang, K. Huang, and Y. Gong, "Integrated sensing, communication, and computation over-the-air: MIMO beamforming design," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5383–5398, 2023.

[33] D. Wen, P. Liu, G. Zhu, Y. Shi, J. Xu, Y. C. Eldar, and S. Cui, "Task-oriented sensing, computation, and communication integration for multi-device edge AI," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[34] S. Tang, P. Popovski, C. Zhang, and S. Obana, "Multi-slot over-the-air computation in fading channels," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6766–6777, 2023.

[35] Z. Ding and Y. Fu, "Robust multiview data analysis through collective low-rank subspace," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1986–1997, 2018.

[36] P. Hu, D. Peng, Y. Sang, and Y. Xiang, "Multi-view linear discriminant analysis network," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5352–5365, 2019.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data mining, Inference, and Prediction*. New York, NY, USA: Springer Science & Business Media, 2009.

[38] M. Ye and R. Yang, "Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 23-28 2014.

[39] Q. Lan, Q. Zeng, P. Popovski, D. Gündüz, and K. Huang, "Progressive feature transmission for split classification at the wireless edge," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3837–3852, 2023.

[40] M. Feder and N. Merhav, "Relations between entropy and error probability," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 259–266, 1994.

[41] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.

[42] P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. National Inst. Sci. India*, vol. 2, no. 1, pp. 49–55, 1936.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge Univ. Press, 2012.

[44] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.