Semantic Communications for Wireless Sensing: RIS-Aided Encoding and Self-Supervised Decoding

Hongyang Du, *Graduate Student Member, IEEE*, Jiacheng Wang, Dusit Niyato, *Fellow, IEEE*, Jiawen Kang, Zehui Xiong, *Member, IEEE*, Junshan Zhang, *Fellow, IEEE*, and Xuemin Shen, *Fellow, IEEE*

Abstract— Semantic communications can reduce the resource consumption by transmitting task-related semantic information extracted from source messages. However, when the source messages are utilized for various tasks, e.g., wireless sensing data for localization and activities detection, semantic communication technique is difficult to be implemented because of the increased processing complexity. In this paper, we propose the *inverse semantic communications* as a new paradigm. Instead of extracting semantic information from messages, we aim to encode the task-related source messages into a hyper-source message for data transmission or storage. Following this paradigm, we design an inverse semantic-aware wireless sensing framework with three algorithms for data sampling, reconfigurable intelligent surface (RIS)-aided encoding, and self-supervised decoding, respectively. Specifically, on the one hand, we propose a novel RIS hardware design for encoding several signal spectrums into one *MetaSpectrum*. To select the task-related signal spectrums for achieving efficient encoding, a semantic hash sampling method is introduced. On the other hand, we propose a self-supervised learning method for decoding the *MetaSpectrums* to obtain the original signal spectrums. Using the sensing data collected from real-world, we show that our framework can reduce the data volume by 95% compared to that before encoding, without affecting the accomplishment of sensing tasks. Moreover, compared with the typically used uniform sampling scheme, the proposed semantic hash sampling scheme can achieve 67% lower mean squared error in recovering the sensing parameters. In addition, experiment results demonstrate that the amplitude response matrix of the RIS enables the encryption of the sensing data. The code for this paper is available at https://github.com/HongyangDu/SemSensing.

Index Terms— Semantic communications, reconfigurable intelligent surface, wireless sensing, self-supervised learning.

Hongyang Du, Jiacheng Wang, and Dusit Niyato are with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hongyang001@e.ntu.edu.sg; jcwang_cq@foxmail.com; dniyato@ntu.edu.sg).

Jiawen Kang is with the School of Automation, Guangdong University of Technology, Guangdong 510006, China (e-mail: kavinkang@gdut.edu.cn).

Zehui Xiong is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372 (e-mail: zehui_xiong@sutd.edu.sg).

Junshan Zhang is with the Department of Electrical and Computer Engineering, University of California at Davis, Davis, CA 95616 USA (e-mail: jazh@ucdavis.edu).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSAC.2023.3288231.

Digital Object Identifier 10.1109/JSAC.2023.3288231

I. INTRODUCTION

WITH the evolution of the next-generation Internet and the proliferation of wireless applications, the demand of network resources for data transmission, storage, and computation has been increasing rapidly. Specifically, the advancement of technologies like extended reality and digital twins is driving the development of the Metaverse and Web 3.0 concepts [1]. As a result, there is an increasing demand for robust communication and computing support. To address the strict requirements of next-generation Internet applications, such as low latency, high reliability, and immersive experiences, semantic communication has been proposed as a key approach in the context of sixth-generation wireless communications [2]. By transmitting only task-related semantic information extracted from source messages, semantic communications are believed to extend the conventional Shannon communication paradigm and bring higher quality of experience to users [3], [4].

While semantic communication techniques have demonstrated their significant effectiveness in processing source data across multiple modalities [5], such as audio [6], image [7], video [8], and text [9], the application of semantic communications to wireless sensing data processing remains a promising area with limited exploration thus far [4]. The sensing data is important because that wireless signals are ubiquitous in our daily life, and can be used to accomplish various tasks requested by service providers. Specifically, wireless signals not only help users access the Internet more efficiently, e.g., Metaverse, but also enable indoor positioning and activities detection more effectively. The wireless sensing data also facilitates the construction of virtual worlds such as digital twins. Unlike on-body sensor-based solutions [10], wireless sensing does not require the user to carry any devices and
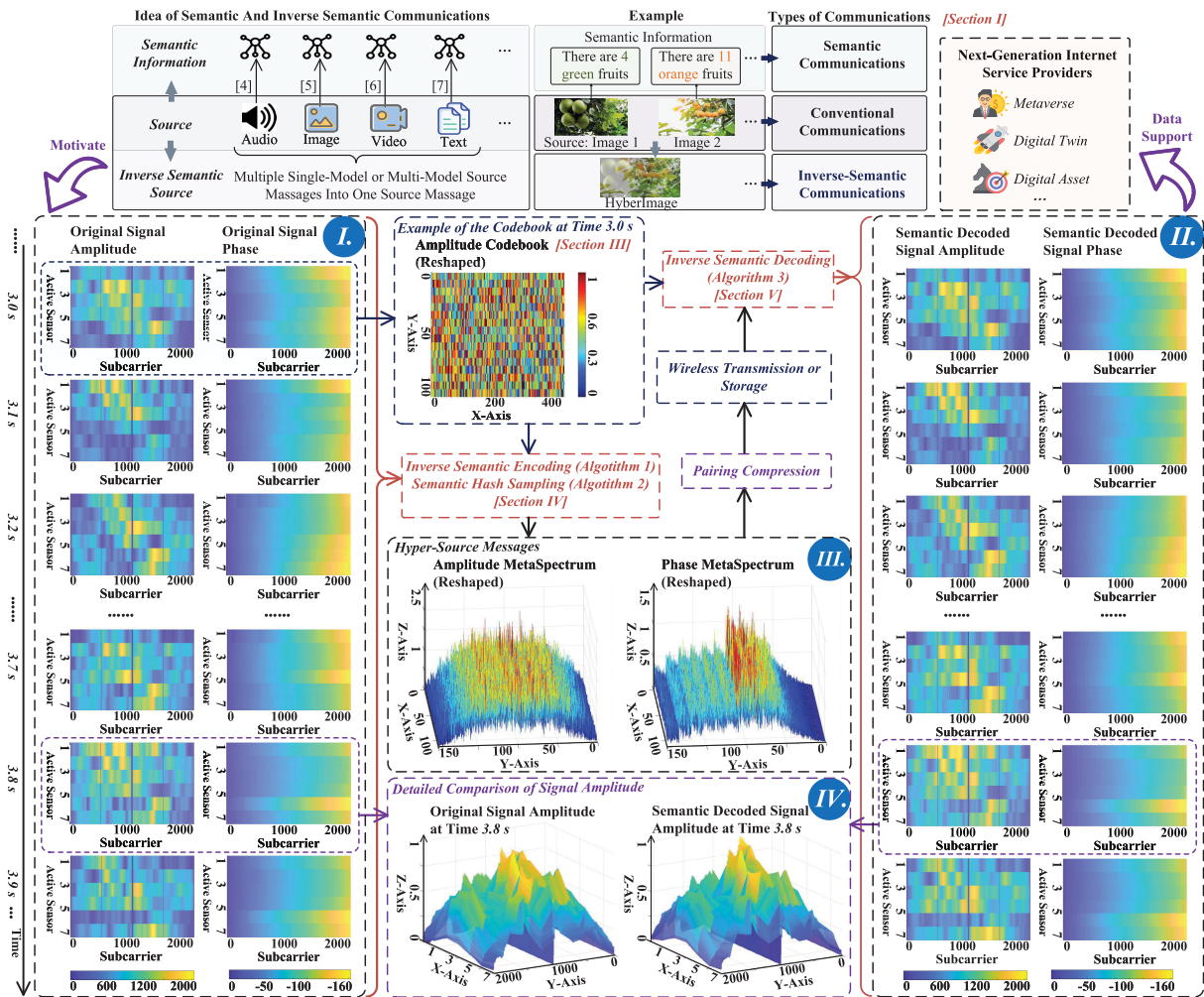
Fig. 1. The ideas of conventional, semantic, and inverse semantic communications. Motivated by the inverse semantic-aware communication, we propose an inverse semantic-aware encoding and decoding framework and show the results. Specifically, we select 10 original signal amplitude/phase spectrum (Part I) by using our proposed **Algorithm 2**, and encode them into one MetaSpectrum by using the RIS and **Algorithm 1**. After wireless transmission, the reconstruction results (Part II) are obtained by decoding the MetaSpectrum using **Algorithm 3**. The sensing data is collected by real experiments with an IEEE 802.11ax based test platform [15].

equipment, which is more practical and convenient. Additionally, the wireless sensing method is more robust than camera-based methods particularly in cases of occlusion or inadequate illumination, while causing fewer privacy issues.

However, the wireless sensing technique has one major limitation. The transmission and storage of the sensing data, such as signal amplitude and phase spectrums, consumes a large number of resources [11]. In particular, the development of communication technologies such as multiple-input multiple-output and orthogonal frequency-division multiplexing (OFDM) improve the sensing resolution in the spatial and time-frequency domains, which, however, further increases the sensing data volume. Therefore, the semantic communication technique is expected to achieve efficient sensing data transmission or storage while achieving sensing tasks. This vision is more meaningful for applications that require long-term storage of sensing data, such as incremental learning for recognition [12], healthcare services [13] and Internet-of-Things (IoT) applications [14]. The reason that semantic communications "exceed" the Shannon limit is the "impairment" of the transmitted data, i.e., an effective semantic encoder extracts

only task-independent semantic information from the source messages. However, a potential pitfall is that the well-trained semantic encoding and decoding models for one specific task may fail when the source messages are needed to accomplish several different tasks. As shown at the top of Fig. 1, instead of transmitting an image, the semantic encoder can extract sentences describing the image. This greatly reduces the number of bits that are required to be transmitted. However, semantic communications would not work well when the task is not only to know the type and number of fruits, but also to know the spatial location. In this case, updated semantic models are required to be re-trained. In a word, semantic communications achieve efficiency transmission while introducing limitations. For the wireless sensing data, if we extract only the semantic information used for localization, the gesture detection task might not be accomplished.

To fill this gap in semantic communications, we propose an inverse semantic-aware approach by treating the source messages as semantic information of a *hyper-source message*. As shown in Fig. 1, the "inverse" means that the processing of source messages is no longer to extract semantic information,

but to combine multiple source messages (Part I) into one hyper-source message (Part III) for transmission or storage. Subsequently, by decoding, the semantic information of the hyper-source message (Part II), i.e., source messages, can be obtained to support multiple different tasks. Using the inverse semantic-aware approach, we reduce the data volume for transmission or storage, while avoiding the task limitations brought by semantic communications. For the wireless sensing, the source messages are signal amplitude and phase spectrums, and we call the hyper-source messages as amplitude and phase MetaSpectrums, respectively. We use the reconfigurable intelligent surface (RIS) to ensure efficient inverse semantic-aware encoding and decoding.[1] With the RIS's superior ability to modulate signals, our scheme can be implemented effectively by modifying a small number of elements on the RIS without affecting the RIS-aided communications. *Unlike most RIS research works that consider only the phase response matrix of RIS, to the best of our knowledge, this is the first paper to make full use of the amplitude response matrix of RIS to help the system design for wireless sensing.* The amplitude response matrix is not only used to reduce the sensing data volume significantly, but also to encrypt the sensing data because the amplitude response matrix is inevitable in the decoding process. Visual representation of the contributions of this paper is shown in Fig. 1, which are summarized as follows:

- We propose a novel RIS hardware design following the paradigm of inverse semantic communications. The design features L-shaped active sensors placed behind transmissive elements in RIS, enabling inverse semantic-aware wireless sensing that significantly reduces the sensing data volume to 5% of the original data volume, leading to improved efficiency and resource utilization.
- We develop the inverse semantic-aware encoding and decoding methods that leverage the amplitude response matrix of the RIS to embed prior knowledge in the sensing signals. The self-supervised learning-based decoding method requires no pre-training resource consumption and allows for the recovery of source messages to support multiple different tasks, overcoming the limitations encountered in semantic communications.
- We introduce an effective semantic hash sampling algorithm for selecting task-related sensing signal spectrums for decoding. Our approach achieves a mean squared error (MSE) between the ground truth and the 2D angles-of-arrival (AoA) estimation results that is 67% lower than that of typically used uniform sampling schemes, and enhances security by utilizing the amplitude response matrix of the RIS for data encryption.
- We build an IEEE 802.11ax based test platform [15] to collect real-world sensing data and perform experiments to show the effectiveness of our proposed framework.

The remainder of the paper is organized as follows. In Section II, we review the related work in the literature. Section III introduces the system model, which contains the novel RIS hardware and the sensing signal model. The inverse semantic-aware encoding and decoding methods are proposed in Section IV and Section V, respectively. Section VI presents the experiment results. In Section VII, we present the conclusion and discuss some potential research directions.

II. RELATED WORK

In this section, we provide a brief review of three related techniques, i.e., wireless sensing, RIS, and spectral snapshot compressive imaging.

A. Wireless Sensing

Wireless signals, such as WiFi [17], have been widely used for sensing tasks ranging from large-scale intrusion detection to small-scale gesture recognition and breathing monitoring. With the rapid advancement of wireless sensing techniques, next-generation internet service providers can construct digital models of the physical world (for digital twin service) or conduct analysis of users' behaviors (for Metaverse services) [13], [18], [19]. Wireless IoT devices collect the sensing data and use channel state information (CSI), which can be obtained as a sampled channel frequency response (CFR), for sensing tasks such as human activities detection [20] and passive localization [21]. The CFR can be expressed as a complex matrix and decomposed into an amplitude spectrum and a phase spectrum for easy transmission and storage. With a three-dimensional multiple signal classification (3D-MUSIC) algorithm, the 3D spectrum can be obtained using the amplitude and phase spectra, which contain the time of flight (ToF) information. The obtained 3D spectrum can achieve several purposes, such as user localization [22] or activity detection in the physical world. A challenge in this process is that the storage or transmission can cause excessive network resource consumption due to a large amount of sensing data.

B. Reconfigurable Intelligent Surface

Significant developments in RIS-aided wireless communications have been witnessed over the past 3 years, from hardware and algorithms design to deep integration with various technologies [23], [24]. One of the most important application scenarios is to enhance wireless sensing [25], such as indoor localization [26] and direction-of-arrival estimation [27]. However, the existing methods typically aim to improve the sensing accuracy through signal enhancement by the RIS. The signal control capability of the RIS is not fully utilized, and most literature is limited in the study of reflective RIS that cannot achieve complete coverage. As our understanding of reconfigurable intelligent surface (RIS) hardware deepens, there has been a growing focus on transmissive and refractive RISs in the research community [28], [29], [30]. Recently, novel RIS architectures such as simultaneously transmitting and reflecting (STAR) RIS [29] and intelligent omni-surface (IOS) [30] have been proposed to facilitate full-dimensional communications. We believe that the implementation of STAR RIS or IOS has the potential to significantly enhance wireless sensing capabilities. In addition to the direct benefit of improving sensing performance through signal enhancement, the ability to adjust the amplitude of transmissive signals can be leveraged as prior knowledge for the efficient compression of wireless sensing data. This paper will delve further into this concept and its implications for the field of wireless sensing.

---

[1]Our scheme can alternatively be achieved by using active antennas and processors [16] to simulate the same signal processing as the RIS. However, higher hardware costs are introduced compared to the scheme using RIS.

C. Spectral Snapshot Compressive Imaging

Capturing high dimension (HD) data is a long-term challenge in signal processing and related fields [31]. With theoretical guarantees, snapshot compressed imaging (SCI) uses two-dimensional (2D) detectors to capture HD, e.g., 3D, data in snapshot measurements using novel optical design. Then, reconstruction algorithms are applied to obtain the required HD data cubes [32], [33]. SCI has been used in many fields such as hyper-spectral imaging, video, holography, tomography, focal depth imaging, polarization imaging, and microscopy [34]. However, there is no prior work discussing how to apply SCI to compressed sensing signals in the time dimension. The reason is that the highly dynamic nature of sensing signals brings difficulties to detector hardware design, coded aperture structure, and decompression algorithms. To fill this gap, in Section IV-A, we use the novel RIS hardware to perform one kind of special SCI to the sensing data. Using our proposed inverse semantic-aware encoding and decoding methods, the compression and self-supervised decompression of the sensing data can be achieved on time scale. Note that our design is different from compressive sensing (CS) methods in wireless sensing, and in fact can be used to further improve the performance of wireless CS systems.

One primary objective of this study is to solve an important problem of overwhelming storage or transmission resources consumption in the wireless sensing. Inspired by the SCI system, we propose an encoding and decoding framework using the RIS to achieve inverse semantic-aware sensing, which significantly reduces the data volume and does not affect the accomplishment of various sensing tasks.

III. SYSTEM MODELS

Wireless signals contain user information such as activities and walking trajectories, and can preserve user privacy better than camera-based methods. Thus, mobile application SPs can use wireless signals to provide better services to users. For example, healthcare SPs can provide medical advice by analyzing the user's sleeping postures, and Metaverse SPs can customize virtual traveling scenes by positioning the users. To meet the needs of ubiquitous sensing data collection, we consider a 3D wireless indoor communications scenario as an example. As shown in Fig. 2 (Part I), a multi-antenna transmitter, e.g., IoT devices or WiFi router, transmits signals to multiple users with the help of an RIS. Different from the conventional scheme that uses RIS to improve sensing accuracy by enhancing signal strength, in this section, we propose a novel RIS hardware design to enable RIS with wireless sensing capability. Then, we analyze the mathematical formulas of the received sensing signals.

A. Novel Hardware of Reconfigurable Intelligent Surface

To enable RIS to sense the environment, a widely used solution is to replace some reflecting elements on the RIS with active sensors, e.g., for channel estimation using CS [35]. Thus, a part of the RIS elements can switch between two operation modes, i.e., i) channel sensing mode that is used to estimate the channels, ii) reflection mode that reflects the signal. However, we see that the RIS cannot assist communications in mode 1. We do not adopt directly the aforementioned
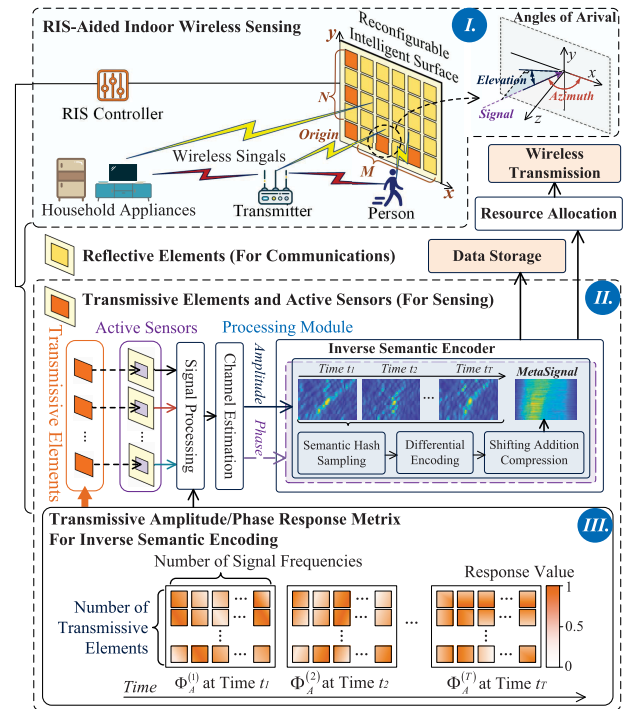

Fig. 2. The framework of the proposed inverse semantic-aware wireless sensing system.

solution since our goal is not merely to estimate the channel, but also constantly to sense the environment for localizing and detecting user activities. To facilitate the sensing capabilities of the RIS without compromising its auxiliary communication functions, we initially integrate the RIS with a limited number of simultaneous transmitting and reflecting patches [29], which are called transmissive elements in this paper for convenience. Specifically, as shown in Fig. 2 (Part II), the RIS is equipped with L-shaped $(M + N + 1)$ transmissive elements, and active sensors are strategically positioned behind them to receive signals modulated by the RIS.

Remark 1: The reason for using the L-shaped array is that such a structure has more accurate 2D AoA estimation results than other structures, e.g., cross, linear, and rectangular arrays. This conclusion can be obtained by comparing the Cramer-Rao Bound metrics of different structures [36], [37].

Accordingly, the signal incident on the qth transmissive element can be transmitted and reflected as [29]

$$\beta_{i,q}\exp(j\delta_{i,q}), \qquad i \in \{T, R\}, \tag{1}$$

where $i = T$ is for transmission coefficients and $i = R$ is for reflection coefficients. Note that, for each element, the responses of the RIS for transmission and reflection modes can be designed independently from each other [38]. In the following, we focus on the sensing function that only uses the transmitted signals. The reflection coefficients can be designed independently, which is outside the scope of this paper. Thus, after one path signal penetrates the qth transmissive element on the RIS, the amplitude of the signal is multiplied by $\beta_{T,q}$, and the phase is added by $\delta_{T,q}$.

A common assumption in much of the existing literature is that the amplitude and phase response of each element on the RIS remains constant throughout the signal bandwidth [39], [40]. While this assumption is generally acceptable for narrow

bandwidths, it may become less accurate when dealing with multiple sub-carriers at varying frequencies within a broader range [39], [41]. In our system model, we consider that the transmitter sends the wireless signals modulated by OFDM technology into K sub-carriers.[2] Because that K might be large, e.g., 2048 OFDM sub-carriers are used to transmit data in the IEEE 802.11ax protocol, we consider the practical case in which the element on the RIS has different responses to signals with different frequencies. Thus, the amplitude and phase response matrices of the L-shaped transmissive elements to K sub-carriers at time t can be expressed as

$$\mathbf{\Phi}_A^{(t)} = \begin{bmatrix} \beta_{f_1}^{[1,0]} & \cdots & \beta_{f_1}^{[M,0]} & \beta_{f_1}^{[0,0]} & \beta_{f_1}^{[0,1]} & \cdots & \beta_{f_1}^{[0,N]} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{f_K}^{[1,0]} & \cdots & \beta_{f_K}^{[M,0]} & \beta_{f_K}^{[0,0]} & \beta_{f_K}^{[0,1]} & \cdots & \beta_{f_K}^{[0,N]} \end{bmatrix}, \quad (2)$$

and

$$\mathbf{\Phi}_P^{(t)} = \begin{bmatrix} \delta_{f_1}^{[1,0]} & \cdots & \delta_{f_1}^{[M,0]} & \delta_{f_1}^{[0,0]} & \delta_{f_1}^{[0,1]} & \cdots & \delta_{f_1}^{[0,N]} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{f_K}^{[1,0]} & \cdots & \delta_{f_K}^{[M,0]} & \delta_{f_K}^{[0,0]} & \delta_{f_K}^{[0,1]} & \cdots & \delta_{f_K}^{[0,N]} \end{bmatrix}, \quad (3)$$

respectively, where f_i denotes the frequency of the ith sub-carrier, and $[x, y]$ denotes the location of the activate element. As shown in Fig. 2 (Part I), $0 \leqslant x \leqslant M$ and $0 \leqslant y \leqslant N$ indicate the locations of transmissive elements that are in the X-direction and Y-direction of the L-shape array, respectively. In the following, we analyze the sensing signal model, and propose the amplitude and phase response matrices design.

B. Sensing Signal Model

To better understand the phase differences of the incident signals from different directions, we analyze the signals that impact the transmissive elements at different positions separately, i.e., the origin, the X-direction, and the Y-direction elements. At time t, the CFR of the multipath signals corresponding to the k-th OFDM sub-carrier obtained by the transmissive element located at the origin of the L-shaped array, can be expressed as

$$h_{f_k}^{[0,0]}(t) = \sum_{i=1}^{I} \alpha_i^{[0,0]} e^{-j2\pi f_k \tau_i}, \quad (4)$$

where $[0, 0]$ represents the origin point, α_i and τ_i are the complex signal amplitude attenuation and time delay of the i-th propagation path, respectively, f_k is the frequency of the k-th sub-carrier, and I is the total number of multipaths. Since different elements are placed at different positions in L-shaped array, the signal needs to travel different distances to arrive each transmissive element. Taking $h_{f_k}^{[0,0]}$ as a reference, therefore, the CFR obtained by the m-th X-direction transmissive element can be expressed as

$$h_{f_k}^{[m,0]}(t) = \sum_{i=1}^{I} \alpha_i^{[m,0]} e^{-j2\pi f_k \left(\tau_i + m \frac{d\cos(\theta_i)\sin(\varphi_i)}{c}\right)}, \quad (5)$$

where $\alpha_i^{[m]}$ is the signal amplitude attenuation, d is the antenna spacing equals half wave length, θ_i and φ_i represent the

elevation angle and azimuth angle of the incident signal, respectively, as shown in Fig. 2 (Part I), and c is the signal propagation speed in the air. Similarly, we can obtain the CFR of the n-th Y-direction transmissive element as

$$h_{f_k}^{[0,n]}(t) = \sum_{i=1}^{I} \alpha_i^{[0,n]} e^{-j2\pi f_k \left(\tau_i + n \frac{d\sin(\theta_i)\sin(\varphi_i)}{c}\right)}. \quad (6)$$

Therefore, at time t, the overall CFR obtained by the L-shaped transmissive element array on the RIS can be expressed as a CFR matrix as follows:

$$\mathbf{H}_{xoy}^{(t)} = \begin{bmatrix} \mathbf{H}_x^{(t)} & \mathbf{H}_0^{(t)} & \mathbf{H}_y^{(t)} \end{bmatrix}$$

$$= \begin{bmatrix} h_{f_1}^{[1,0]} & \cdots & h_{f_1}^{[M,0]} & h_{f_1}^{[0,0]} & h_{f_1}^{[0,1]} & \cdots & h_{f_1}^{[0,N]} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \underbrace{h_{f_K}^{[1,0]} \cdots h_{f_K}^{[M,0]}}_{\mathbf{H}_x} & \underbrace{h_{f_K}^{[0,0]}}_{\mathbf{H}_0} & \underbrace{h_{f_K}^{[0,1]} \cdots h_{f_K}^{[0,N]}}_{\mathbf{H}_y} \end{bmatrix}. \quad (7)$$

Note that $\mathbf{H}_{xoy}^{(t)}$ is the original sensing data that can support various sensing tasks. Because of the high available sampling frequency of the sensing device, e.g., 300 times in one second [19], and novel services that require long-term sensing, a large amount of $\mathbf{H}_{xoy}^{(t)}$ would be collected. To reduce the resources costed to store and transmit $\mathbf{H}_{xoy}^{(t)}$, we propose the inverse semantic-aware encoding and decoding methods in the following sections, respectively.

IV. INVERSE SEMANTIC-AWARE ENCODING

In this section, we introduce the inverse semantic-aware RIS-aided encoding method to compress multiple signal spectrums into one. Two steps, i.e., differential encoding and shifting addition compression, are discussed. Moreover, we propose a semantic hash sampling method to select the task-related signal spectrum to record.

A. Encoding Method

One can observe from (7) that every element in the CFR matrix is a complex number, which denotes the amplitude and phase of the CFR. Taking $\mathbf{H}_x^{(t)}$ as an example, it can be further decomposed into the amplitude and phase spectrums as

$$\mathbf{H}_x^{(t)} \rightarrow \left\{ \mathbf{H}_{x_a}^{(t)}, \mathbf{H}_{x_p}^{(t)} \right\}$$

$$= \left\{ \underbrace{\begin{bmatrix} \left\| h_{f_1}^{[1,0]} \right\| & \cdots & \left\| h_{f_1}^{[M,0]} \right\| \\ \vdots & \ddots & \vdots \\ \left\| h_{f_K}^{[1,0]} \right\| & \cdots & \left\| h_{f_K}^{[M,0]} \right\| \end{bmatrix}}_{\text{amplitude matrix}}, \underbrace{\begin{bmatrix} \angle h_{f_1}^{[1,0]} & \cdots & \angle h_{f_1}^{[M,0]} \\ \vdots & \ddots & \vdots \\ \angle h_{f_K}^{[0,1]} & \cdots & \angle h_{f_K}^{[M,0]} \end{bmatrix}}_{\text{phase matrix}} \right\}, \quad (8)$$

where $\cdot \rightarrow \cdot$ denotes the amplitude and phase extraction operation, $\{\cdot\}$ represents the set of matrices, $\|\cdot\|$ is the Euclidean norm operator, and $\angle h_{f_k}^{[m,0]}$ denotes the signal phase of $h_{f_k}^{[m,0]}$. Through the same way, $\mathbf{H}_0$ and $\mathbf{H}_y$ can be expressed as amplitude and phase spectrums, respectively. Hence, the CFR
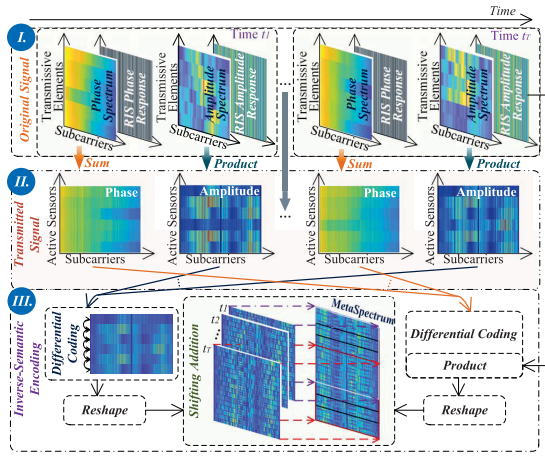
Fig. 3. The process of modulating the amplitude and phase spectrums through RIS, and then performing differential encoding and shifting addition.

extracted from the L-shaped transmissive element array on the RIS can be expressed as

$$\mathbf{H}_{xoy}^{(t)} = \left[\mathbf{H}_x^{(t)} \; \mathbf{H}_0^{(t)} \; \mathbf{H}_y^{(t)} \right] \rightarrow \left\{ \mathbf{H}_A^{(t)}, \mathbf{H}_P^{(t)} \right\}, \qquad (9)$$

where $\mathbf{H}_A^{(t)} = \left[\boldsymbol{H}_{x_a} \; \boldsymbol{H}_{0_a} \; \boldsymbol{H}_{y_a} \right]$ and $\mathbf{H}_P^{(t)} = \left[\boldsymbol{H}_{x_p} \boldsymbol{H}_{0_p} \; \boldsymbol{H}_{y_p} \right]$ denote the overall amplitude and phase at time t, respectively. As shown in Fig. 3 (Parts I and II), after being modulated by the transmissive elements, we can express T received amplitude and phase spectrums by the L-shaped active sensor array in two sets as

$$\boldsymbol{Y}_A = \left\{ \boldsymbol{H}_A^{(1)} \circ \boldsymbol{\Phi}_A^{(1)}, \ldots, \boldsymbol{H}_A^{(T)} \circ \boldsymbol{\Phi}_A^{(T)} \right\}, \qquad (10)$$

and

$$\boldsymbol{Y}_P = \left\{ \boldsymbol{H}_P^{(1)} + \boldsymbol{\Phi}_P^{(1)}, \ldots, \boldsymbol{H}_P^{(T)} + \boldsymbol{\Phi}_P^{(T)} \right\}, \qquad (11)$$

respectively, where $\circ$ is the Hadamard product calculator, a.k.a., element-wise product. In the following, we encode the 3D data $\boldsymbol{Y}_A$ and $\boldsymbol{Y}_P$ onto 2D measurements, respectively. The encoding idea is inspired by the SCI system that compresses several optical spectrums of an object over multiple wavelengths into one spectrum, or several frames of a high-speed video into one frame. Specifically, the 3D data is first modulated by a coded aperture, and then spectrally dispersed by the dispersing element, and finally integrated across the spectral dimension to a 2D measurement. For the 3D sensing data $\boldsymbol{Y}_A$ and $\boldsymbol{Y}_P$, although the spectral dispersion process can be performed by low-power computing elements, there are several difficulties in adopting the compression scheme as in the SCI system:

D1) The fixed coded aperture in the SCI system is hard to be used in encoding signal spectrums that change dramatically on the time scale. However, the time-varying coded aperture scheme [42] increases the hardware cost and consume more storage space to record the patterns.

D2) It is difficult for system designers to strike the balance between decoding performance and resource consumption. The inverse decoding problem is hard to be solved by traditional methods. The deep learning approach, such as convolutional neural networks, necessitates the use of extensive, well-labeled datasets and prolonged

training periods [43], which restricts the frequency of aperture pattern updates.

D3) Signal spectrums are more sensitive than spectral images or video frames. We find from the experiments that the decoded sensing signal spectrums may lead to errors when performing some sensing tasks that are sensitive to the deviations in signal phase values, e.g., localization.

To overcome the aforementioned difficulties, we rethink the SCI system from hardware design to software algorithms. For (D1) and (D2), we can observe from (10) that the amplitude response matrix of the RIS has potential to perform a similar function as the coded aperture in the SCI system. It has been shown that the reconfiguration time for the RIS to change the response matrix is around 33 ns [44]. Therefore, by changing the response matrix over time, the low-cost transmissive elements on the RIS can encode the sensing signals. In addition, the response values can be obtained from the hardware design parameters. This saves the storage resources to record a large number of original response values. Moreover, the amplitude and phase response values are discrete numbers, which can be determined by the number of coding bits. For example, 4-bit coding bringse 16 different available response values. Following that, we propose a self-supervision decoding algorithm for arbitrary RIS response matrices, which is discussed in Section V. Error negligible decoding results are achieved without pre-training resource consumption.

To solve (D3), we compress the differential matrices of the amplitude and phase spectrums instead of the original spectrums to ensure the sensing performance. Unlike channel estimation, which focuses on accurately obtaining the CSI to better perform channel equalization, wireless sensing focuses on extracting information describing the physical environment from the CSI, e.g., 2D AoA and time of flight. This information is hidden in the value difference of amplitude and phase spectrums obtained by the sensors at different locations. For example, the phase difference between active sensors supports the signal AoA estimation. Another advantage to encode the differential spectrum is that the differential spectrum tends to be smoother than the original spectrum, due to the existence of correlation. This results in improved decoding performance. We show that real images can also benefit from the differential encoding in Fig. 10 using the dataset [45].

The differential encoding and shifting addition compression methods are presented in the following.

1) Differential Encoding: We first focus on the amplitude spectrum set $\boldsymbol{Y}_A$. Let $\boldsymbol{Y}_A\{i\}$ denote the ith matrix in $\boldsymbol{Y}_A$. Each column in $\boldsymbol{Y}_A\{i\}$ represents the amplitude values of received signals at different frequencies by an active sensor, after amplitude modulation by the transmissive element on the RIS. Let $\boldsymbol{Y}_{A'}\{i\}$ denote the $\boldsymbol{Y}_A\{i\}$ after the differential encoding. Specifically, we let the jth column in $\boldsymbol{Y}_{A'}\{i\}$ store the difference values of the jth column and $(j-1)^{\text{th}}$ column in $\boldsymbol{Y}_A\{i\}$ as

$$\boldsymbol{Y}_{A'}\{i\}[j,:] = \boldsymbol{Y}_A\{i\}[j,:] - \boldsymbol{Y}_A\{i\}[j-1,:], \qquad (12)$$

where $j = 2, \ldots, L$. The first columns in $\boldsymbol{Y}_{A'}\{i\}$ and $\boldsymbol{Y}_A\{i\}$ are the same. Then, we have

$$\boldsymbol{Y}_{A'}\{i\}[1,:] = \boldsymbol{Y}_A\{i\}[1,:]. \qquad (13)$$

Similar differential encoding method can be used for the received phase spectrum set $\boldsymbol{Y}_P$. For the ith matrix in $\boldsymbol{Y}_P$, i.e., $\boldsymbol{Y}_P\{i\}$, we obtain the differential encoded matrix $\boldsymbol{Y}_{P'}\{i\}$ by

$$\boldsymbol{Y}_{P'}\{i\}[j,:] = \boldsymbol{Y}_P\{i\}[j,:] - \boldsymbol{Y}_P\{i\}[j-1,:], \quad (14)$$

where $j = 2, \ldots, L$. Considering that the phase response value of the RIS is added to the signal phase value, we let the first column in $\boldsymbol{Y}_{P'}\{i\}$ be the first column in $\boldsymbol{Y}_P\{i\}$ minus the phase response of the first transmissive element as

$$\boldsymbol{Y}_{P'}\{i\}[1,:] = \boldsymbol{Y}_P\{i\}[1,:] - \boldsymbol{\Phi}_P^{(i)}[1,:]. \quad (15)$$

To use the amplitude response matrix of the RIS as the prior knowledge, we multiply the amplitude response matrix of the RIS at the ith moment and $\boldsymbol{Y}_{P'}\{i\}$ by elements as

$$\boldsymbol{Y}_{P'}\{i\} = \boldsymbol{Y}_{P'}\{i\} \circ \boldsymbol{\Phi}_A^{(i)}. \quad (16)$$

In addition to the steps of (12), (13), (14), (15), and (16), the transmissive elements on the RIS should be designed by following Remark 2 to make the amplitude response matrix of the RIS available as a special coded aperture, i.e., prior knowledge used in decoding.

Remark 2: To achieve differential encoding, we should let every transmissive element on the RIS have the same hardware structure. Thus, different transmissive elements have the same amplitude and phase response to the signals with the same frequency, as shown in Fig. 3 (Part I). Specifically, every column in (2) and (3) is the same. This ensures that each column of $\boldsymbol{Y}_{A'}\{i\}$ can be represented as the signal amplitude difference values multiplied by the amplitude response values of the RIS as in (17).

Then, we can express $\boldsymbol{Y}_{A'}\{i\}$ and $\boldsymbol{Y}_{P'}\{i\}$ as

$$\boldsymbol{Y}_{A'}\{i\} = \boldsymbol{H}_{A'}^{(i)} \circ \boldsymbol{\Phi}_A^{(i)}, \quad (17)$$

and

$$\boldsymbol{Y}_{P'}\{i\} = \boldsymbol{H}_{P'}^{(i)} \circ \boldsymbol{\Phi}_A^{(i)}, \quad (18)$$

where $\boldsymbol{H}_{A'}^{(i)}$ and $\boldsymbol{H}_{P'}^{(i)}$ are the ith differential encoded amplitude and phase spectrums, respectively, and $\boldsymbol{\Phi}_A^{(i)}$ can be regarded as the corresponding codebook.

2) Shifting Addition: To replace the spatial shifting operation to the object spectrum that is performed by a dispersing lens in the SCI system, we perform zero compensation processing to the amplitude and phase spectrums as follows:

$$\boldsymbol{X}_A = \left\{ \begin{bmatrix} \boldsymbol{Q}_1(1) \\ \boldsymbol{Y}_{A'}\{1\} \\ \boldsymbol{Q}_2(1) \end{bmatrix}, \ldots, \begin{bmatrix} \boldsymbol{Q}_1(i) \\ \boldsymbol{Y}_{A'}\{i\} \\ \boldsymbol{Q}_2(i) \end{bmatrix}, \ldots, \begin{bmatrix} \boldsymbol{Q}_1(T) \\ \boldsymbol{Y}_{A'}\{T\} \\ \boldsymbol{Q}_2(T) \end{bmatrix} \right\}, \quad (19)$$

where $\boldsymbol{Q}_1(i) \in \mathbb{R}^{(i-1)D \times L}$, $\boldsymbol{Q}_2(i) \in \mathbb{R}^{(T-i)D \times L}$, $\boldsymbol{X}_A \in \mathbb{R}^{(D(T-1)+K) \times L}$, every elements in both $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ is zero, and D is the unit displacement step.

Thus, the amplitude MetaSpectrum, $\boldsymbol{Z}_A$, can be obtained by

$$\boldsymbol{Z}_A = \sum_{i=1}^{T} \boldsymbol{X}_A\{i\}, \quad (20)$$

Algorithm 1 The Algorithm for Inverse Semantic-Aware Encoding

Input: The received amplitude and phase spectrums in the active sensors: $\boldsymbol{Y}_A$ and $\boldsymbol{Y}_P$

Output: The amplitude and phase MetaSpectrums: $\boldsymbol{Z}_A$ and $\boldsymbol{Z}_P$

1: *## Achieve differential encoding*
2: **for** Every $\boldsymbol{Y}_A\{i\}$ in $\boldsymbol{Y}_A$ **do**
3:     Obtain $\boldsymbol{Y}_{A'}\{i\}$ according to (12) and (13)
4: **for** Every $\boldsymbol{Y}_P\{i\}$ in $\boldsymbol{Y}_P$ **do**
5:     Obtain $\boldsymbol{Y}_{P'}\{i\}$ according to (14), (15), and (16)
6: *## Achieve shifting addition compression*
7: Use $\boldsymbol{Y}_{A'}$ to obtain $\boldsymbol{X}_A$ according to (19)
8: Use $\boldsymbol{Y}_{P'}$ to obtain $\boldsymbol{X}_P$ according to (22)
9: Obtain amplitude MetaSpectrum $\boldsymbol{Z}_A$ according to (20)
10: Obtain phase MetaSpectrum $\boldsymbol{Z}_P$ according to (21)
11: **return** $\boldsymbol{Z}_A$ and $\boldsymbol{Z}_P$

---

where $\boldsymbol{Z}_A \in \mathbb{R}^{(K+(T-1)D) \times L}$ can be transmitted or stored. Similarly, the phase MetaSpectrum, $\boldsymbol{Z}_P$, can be expressed as

$$\boldsymbol{Z}_P = \sum_{i=1}^{T} \boldsymbol{X}_P\{i\}, \quad (21)$$

where

$$\boldsymbol{X}_P = \left\{ \begin{bmatrix} \boldsymbol{Q}_1(1) \\ \boldsymbol{Y}_{P'}\{1\} \\ \boldsymbol{Q}_2(1) \end{bmatrix}, \cdots, \begin{bmatrix} \boldsymbol{Q}_1(i) \\ \boldsymbol{Y}_{P'}\{i\} \\ \boldsymbol{Q}_2(i) \end{bmatrix}, \cdots, \begin{bmatrix} \boldsymbol{Q}_1(T) \\ \boldsymbol{Y}_{P'}\{T\} \\ \boldsymbol{Q}_2(T) \end{bmatrix} \right\}. \quad (22)$$

The overall RIS-aided encoding method is in **Algorithm 1**, which has polynomial complexity. After the RIS-aided encoding, we observe that the sensing data volume is significantly reduced. To demonstrate the efficiency of the inverse semantic communications approach, we define the data compression ratio, i.e., ρ, as the ratio between the number of elements in the received amplitude and phase spectrums and the number of elements in the coded MetaSpectrums. We provide the analysis of ρ in **Proposition 1**.

Proposition 1: The data compression ratio ρ of our proposed inverse semantic-aware coding method is approximately $1/T$. Proof: The total number of elements in the T recorded signal amplitude and phase spectrum is $2KLT$. Meanwhile, the total number of elements in $\boldsymbol{Z}_A$ and $\boldsymbol{Z}_P$ is $2(K + (T-1)D) \times L$. As a result, ρ can be calculated as:

$$\rho = \frac{2(K + (T-1)D) \times L}{2KLT} = \frac{1}{T} + \left(1 - \frac{1}{T}\right)\frac{D}{K}. \quad (23)$$

Given that D is generally small compared to K, for instance, $D = 1$ in [33] and $K = 2048$ in the IEEE 802.11ax protocol, the term $\left(1 - \frac{1}{T}\right)\frac{D}{K}$ in (23) can be negligible. Consequently, the value of ρ is approximately $1/T$, which completes the proof. ∎

Note that we encode T amplitude or phase spectrums into one spectrum. However, the T spectrums does not need and should not be sensed continuously in time. The reason is that the wireless channel remains stable during the channel coherence time. Specifically, as the moving or action speed of

people is limited, the CSI within the channel coherence time can be considered as constant without loss of precision [46]. Considering that the available maximal sensing frequency of the active sensors is much higher than the required frequency, we next propose a sampling scheme that selects the most relevant spectrum for the completion of sensing tasks over the channel coherence time for recording and encoding.

B. Semantic Hash Sampling

We divide the time into segments. Without loss of generality, we consider that the active sensors can perform T_N times sensing in one time segment. From each time segment, one pair of amplitude and phase spectrums is selected to record. In the ith time segment, we express T_N received amplitude and T_N phase spectrums as two sets, i.e., $\boldsymbol{S}_{A_i} = \left\{ \boldsymbol{H}_{A_i}^{(k)} \circ \boldsymbol{\Phi}_{A_i}^{(k)} \right\}$ $(k = 1, \ldots, T_N)$ and $\boldsymbol{S}_{P_i} = \left\{ \boldsymbol{H}_{P_i}^{(k)} \circ \boldsymbol{\Phi}_{P_i}^{(k)} \right\}$, respectively. The recorded amplitude and phase spectrums that are selected from the ith time segment are $\boldsymbol{Y}_A\{i\}$ $(i = 1, \ldots, T)$ and $\boldsymbol{Y}_P\{i\}$, respectively.

To remove information that is not relevant to the task, the traditional method is uniform sampling that selects the first pair of amplitude and phase spectrums in each time segment to record. However, we cannot guarantee that the first pair in every segment is always the most informative pair. Therefore, a better solution is to use one indicator to judge the semantic information richness of the pair of spectrums. As we discussed in Section IV-A, the information related to sensing tasks is contained in the changes of amplitude and phase spectrums. Therefore, we can select the pair of spectrums that has the largest change compared to the previous signal spectrums in each time segment. Note that the mean square error (MSE) is not recommended to be used as the indicator to compare the difference between spectrums. The reasons are given as follows:

- The MSE is calculated using the absolute values of the signal amplitude. However, the absolute values are not important for sensing tasks. The critical information is in the changing process of the signal amplitude over time [17].
- The results of MSE may be affected by several outliers, i.e., signal amplitude fluctuation at a certain time caused by the interference.
- Because the number of elements in the signal spectrum is large, calculating MSE brings large resource consumption.

Therefore, we have to propose a new indicator to characterize the semantic information richness in signal spectrums. Considering the success of the perceptual image hashing method [47] in the field of image retrieval, we aim to use a string of characters, i.e., fingerprints, to characterize the amplitude and phase spectrums. Perceptual image hashing [47] is a family of algorithms that generate content-based image hash fingerprints. Then, the Hamming distance between two fingerprints can be used to quantify the similarity of two images. The larger the Hamming distance is, the smaller the similarity of the images have. Although the hash fingerprints can be calculated efficiently with low energy cost, it cannot be applied directly to the similarity detection of sensing data. The reason
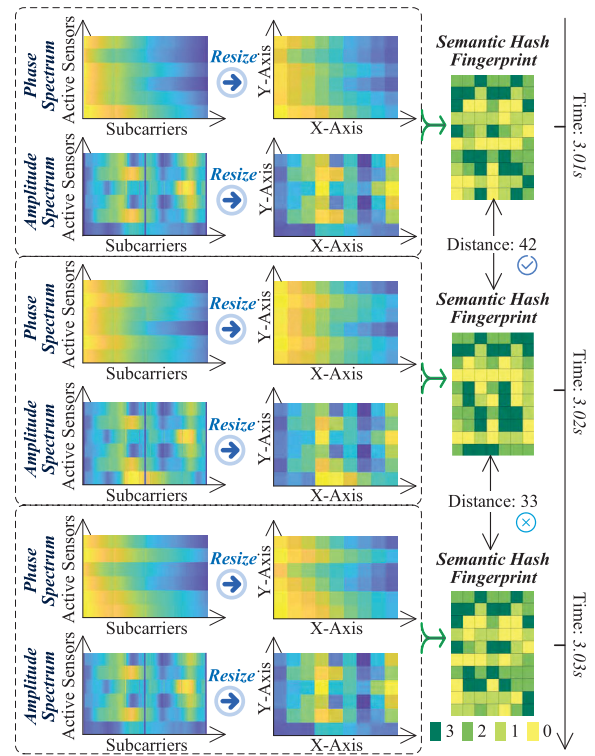

Fig. 4. The process of generating the resized matrices from the amplitude and phase spectrums, and then obtaining the semantic hash fingerprint.

is that, at each moment, we have one amplitude spectrum and one phase spectrum as shown in Fig. 4, which are both required to achieve sensing tasks. Thus, we propose a novel four-level semantic hash sampling method in **Algorithm 2** to select task-related signals spectrums for encoding, which is used before **Algorithm 1**.

As shown in **Algorithm 2**, to obtain the semantic hash matrices, the first step is to resize the T_K amplitude and phase spectrums. The purpose is to produce a small data size, which hastens the processing time [48] and preserves the features of the spectrums. Similar to the image pHash method [49], we calculate the average values of the resized amplitude and phase matrices. Different from the conventional hash method, we define four values, i.e., 0, 1, 2, and 3, as values in the hash fingerprints. Thus, we perform the operations as shown in lines $8 - 15$ of **Algorithm 2** to convert the spectrums to semantic hash fingerprints in polynomial complexity. For the kth pair of amplitude and phase spectrums, we use the Hamming distance, which measures the number of different values, between the kth hash fingerprint and the $(k-1)^{\text{th}}$ one to indicate the semantic information richness of the kth spectrums. Therefore, we can record the pair of spectrums that have the largest Hamming distance to the previous pair of spectrums. The complexity analysis for each step in our proposed algorithm is as follows:

- *Obtaining the semantic hash matrix set:* The resizing operation [49] has a complexity of $\mathcal{O}(R_x R_y)$ per matrix. Therefore, the complexity of obtaining the semantic hash matrix set for T_K spectrums is $\mathcal{O}(T_K R_x R_y)$.
- *Calculating the Hamming distance:* The complexity consists of $O(1)$ for creating the vector $\boldsymbol{\mathcal{D}}$, $O(T_K)$ for the outer loop iterating through $\boldsymbol{\mathcal{H}}_i$, and $O(R_x R_y)$ for the

inner loop iterating through $\mathcal{H}_i\{k\}$. This results in a total complexity of $\mathcal{O}(T_K R_x R_y)$.

- *Selecting the spectrum and recording the information richness:* The complexity is dominated by the linear search to find $k_{\max}$, which maximizes $\mathcal{D}(k)$. As the vector is unsorted, the complexity is $\mathcal{O}(T_K)$.

These steps are performed sequentially, leading to the total complexity of $\mathcal{O}\left(T_K\left(2R_x R_y + 1\right)\right)$. Since the constant terms do not significantly affect the growth rate, the total complexity of **Algorithm 2** simplifies to $\mathcal{O}(T_K R_x R_y)$.

V. INVERSE SEMANTIC-AWARE DECODING

In this section, we propose the inverse semantic-aware self-supervised decoding method. We also introduce how to use the recovered signal spectrums for 2D AoA and ToF estimation, which supports various sensing tasks.

A. Objective Function

We first rewrite the amplitude and phase MetaSpectrums $\boldsymbol{Z}_A$ and $\boldsymbol{Z}_P$ in the vectorized formulations, respectively. Let $\mathrm{vec}\left(\cdot\right)$ denote the matrix vectorization operation that concatenates columns into one vector, and $\mathrm{diag}\left(\mathbf{a}\right)$ denote the operation of converting the vector $\mathbf{a}$ into a diagonal matrix where the diagonal element is $\mathbf{a}$. As such, we rewrite the matrix formulations (20) and (21) as

$$\mathbf{z}_A = \boldsymbol{\Phi}\mathbf{x}_A, \tag{24}$$

and

$$\mathbf{z}_P = \boldsymbol{\Phi}\mathbf{x}_P, \tag{25}$$

respectively, where $\boldsymbol{z}_A = \mathrm{vec}\left(\boldsymbol{Z}_A\right)$, $\boldsymbol{z}_P = \mathrm{vec}\left(\boldsymbol{Z}_P\right)$, $\boldsymbol{z}_A$ and $\boldsymbol{z}_P \in \mathbb{R}^{(K+(T-1)D)L\times 1}$, $\boldsymbol{x}_A = \left[\boldsymbol{x}_{1,A}^{\mathrm{T}}\cdots\boldsymbol{x}_{i,A}^{\mathrm{T}}\cdots\boldsymbol{x}_{T,A}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{x}_P = \left[\boldsymbol{x}_{1,P}^{\mathrm{T}}\cdots\boldsymbol{x}_{i,P}^{\mathrm{T}}\cdots\boldsymbol{x}_{T,P}^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{x}_{i,A} = \mathrm{vec}\left(\boldsymbol{H}_{A'}^{(i)}\right)$, $\boldsymbol{x}_{i,P} = \mathrm{vec}\left(\boldsymbol{H}_{P'}^{(i)}\right)$, $\boldsymbol{x}_A$ and $\boldsymbol{x}_P \in \mathbb{R}^{TKL\times 1}$,

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{Q}_3\left(1\right) \cdots \boldsymbol{Q}_3\left(i\right) \cdots \boldsymbol{Q}_3\left(T\right) \\ \boldsymbol{\phi}_A^{(1)} \quad \ddots \quad \boldsymbol{\phi}_A^{(i)} \quad \ddots \quad \boldsymbol{\phi}_A^{(T)} \\ \boldsymbol{Q}_4\left(1\right) \cdots \boldsymbol{Q}_4\left(i\right) \cdots \boldsymbol{Q}_4\left(T\right) \end{bmatrix}, \tag{26}$$

$\boldsymbol{Q}_3\left(1\right) \in \mathbb{R}^{(i-1)DL\times KL}$, $\boldsymbol{Q}_4\left(1\right) \in \mathbb{R}^{(T-1)DL\times KL}$, every element in both $\boldsymbol{Q}_3$ and $\boldsymbol{Q}_4$ is zero, $\boldsymbol{\phi}_A^{(i)} = \mathrm{diag}\left(\mathrm{vec}\left(\boldsymbol{\Phi}_A^{(i)}\right)\right)$, $\boldsymbol{\phi}_A^{(i)} \in \mathbb{R}^{KL\times KL}$, and $\boldsymbol{\Phi} \in \mathbb{R}^{(K+(T-1)D)L\times TKL}$.

Note that $\boldsymbol{\Phi}$ can be obtained using (26) and the prior knowledge, i.e., the amplitude response matrices of the RIS, and $\mathbf{z}_A$ and $\mathbf{z}_P$ are the known vectorized amplitude and phase MetaSpectrums, respectively. Our goal is to decode $\mathbf{x}_A$ and $\mathbf{x}_P$ from $\mathbf{z}_A$ and $\mathbf{z}_P$, respectively. Although this problem is related to CS, most theories that are developed for CS cannot be used because that the matrix $\boldsymbol{\Phi}$ follows a very specific structure as (26). Fortunately, solid theoretical proof has shown that both $\mathbf{x}_A$ in (24) and $\mathbf{x}_P$ in (25) can be recovered even when $T > 1$ [50].

The decoding objective function can be formulated as

$$\min_{\boldsymbol{x}_A, \boldsymbol{x}_P} \alpha_1 \left\|\boldsymbol{z}_A - \boldsymbol{\Phi}\boldsymbol{x}_A\right\|^2 + \alpha_2 \left\|\boldsymbol{z}_P - \boldsymbol{\Phi}\boldsymbol{x}_P\right\|^2, \tag{27}$$

where α_1 and α_2 are the balance parameters that can be selected according to the specific wireless sensing task.

Algorithm 2 The Algorithm for Semantic Hash Sampling

Input:
- The received amplitude and phase spectrums sets in the ith time segment: $\boldsymbol{S}_{A_i}$ and $\boldsymbol{S}_{P_i}$
- The dimensions of the resized matrices: R_x and R_y

Output: The selected amplitude and phase spectrums: $\boldsymbol{Y}_A\{i\}$ and $\boldsymbol{Y}_P\{i\}$

1: *## Obtain the semantic hash matrix set*
2: Create an empty matrix set $\mathcal{H}_i \in \mathbb{R}^{R_x\times R_y\times T_K}$ to record the semantic hash values
3: **for** Every $\boldsymbol{S}_{A_i}\{k\}$ in $\boldsymbol{S}_{A_i}$ **do**
4:     Obtain amplitude and phase spectrums $\boldsymbol{H}_{A_i}^{(k)}$ and $\boldsymbol{H}_{P_i}^{(k)}$ with the prior knowledge $\boldsymbol{\Phi}_{A_i}^{(k)}$ and $\boldsymbol{\Phi}_{P_i}^{(k)}$, respectively
5:     Resize $\boldsymbol{H}_{A_i}^{(k)}$ and $\boldsymbol{H}_{P_i}^{(k)}$ into small matrices $\boldsymbol{h}_{A_i}^{(k)} \in \mathbb{R}^{R_x\times R_y}$ and $\boldsymbol{h}_{P_i}^{(k)} \in \mathbb{R}^{R_x\times R_y}$, respectively
6:     Calculate the average values of $\boldsymbol{h}_{A_i}^{(k)}$ and $\boldsymbol{h}_{P_i}^{(k)}$, denoted as $h_{A_i}^{(k)}$ and $h_{P_i}^{(k)}$, respectively
7:     **for** Every element pair in $\boldsymbol{h}_{A_i}^{(k)}$ and $\boldsymbol{h}_{P_i}^{(k)}$ **do**
8:       **if** $\boldsymbol{h}_{A_i}^{(k)}[x,y] \geqslant h_{A_i}^{(k)}$ and $\boldsymbol{h}_{P_i}^{(k)}[x,y] \geqslant h_{P_i}^{(k)}$ **then**
9:         Let $\mathcal{H}_i\{k\}[x,y] \leftarrow 3$
10:       **else if** $\boldsymbol{h}_{A_i}^{(k)}[x,y] \geqslant h_{A_i}^{(k)}$ and $\boldsymbol{h}_{P_i}^{(k)}[x,y] < h_{P_i}^{(k)}$ **then**
11:         Let $\mathcal{H}_i\{k\}[x,y] \leftarrow 2$
12:       **else if** $\boldsymbol{h}_{A_i}^{(k)}[x,y] < h_{A_i}^{(k)}$ and $\boldsymbol{h}_{P_i}^{(k)}[x,y] \geqslant h_{P_i}^{(k)}$ **then**
13:         Let $\mathcal{H}_i\{k\}[x,y] \leftarrow 1$
14:       **else**
15:         Let $\mathcal{H}_i\{k\}[x,y] \leftarrow 0$
16: *## Calculate the Hamming distance*
17: Create an empty vector $\mathcal{D} \in \mathbb{R}^{1\times T_K}$ to record the Hamming distance values
18: **for** Every $\mathcal{H}_i\{k\}$ in $\mathcal{H}_i$ **do**
19:     Create a temporary variable d
20:     **for** Every element in $\mathcal{H}_i\{k\}$ **do**
21:       **if** The element value is different from the element value in the same position in $\mathcal{H}_i\{k-1\}$ **then**
22:         $d \leftarrow d + 1$
23:     $\mathcal{D}(k) \leftarrow d$.
24: *## Select the spectrum and record the information richness*
25: Find $k_{\max}$ that maximizes $\mathcal{D}(k)$, i.e., $\mathcal{D}(k_{\max}) = \max\{\mathcal{D}\}$
26: Record the value of $\mathcal{D}(k_{\max})$
27: Let $i = k_{\max}$
28: **return** $\boldsymbol{Y}_A\{i\}$ and $\boldsymbol{Y}_P\{i\}$

For example, heartbeat and breath detection requires higher accuracy for the phase spectrum [19], and the amplitude spectrum is more significant in sensing tasks such as intrusion or fall detection [18]. We propose the algorithm for solving (27) as follows.

B. Self-Supervised Decoding Method

To solve (27), although different hand-crafted priors, e.g., total variation and sparsity, can be added as the regularization

term to improve the decoding performance, it is hard to choose a suitable prior that fits the differential encoded amplitude and phase spectrums $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$. Motivated by the success of deep convolutional neural networks (ConvNets) in inverse problems such as single-image super-resolution [51] and denoising [52], we use the implicit prior captured by the ConvNets, e.g., deep image prior [53], [54], to achieve self-supervised decoding.[3]

By considering that the unknown amplitude and phase spectrums are the outputs of neural networks, i.e., $\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e})$ and $\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e})$, respectively, the decoding problem (27) can be re-written as

$$\min_{\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_P} \alpha_1 \|\boldsymbol{z}_A - \boldsymbol{\Phi}\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e})\|^2 + \alpha_2 \|\boldsymbol{z}_P - \boldsymbol{\Phi}\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e})\|^2,$$
$$\text{s.t. } \hat{\boldsymbol{x}}_A = \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}),$$
$$\hat{\boldsymbol{x}}_P = \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e}), \qquad (28)$$

where $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$ are the parameters of networks to be learned, and $\boldsymbol{e}$ is a random vector. Since the training of $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$ is part of the decoding process, this procedure is self-supervised and no pre-training process is required.

To solve the problem (28), we introduce two auxiliary variables $\boldsymbol{t}_1$ and $\boldsymbol{t}_2 \in \mathbb{R}^{TKL}$, and corresponding weight parameters β_1 and β_2. Then, the constraints can be turned into penalty terms using the augmented Lagrangian method [55] as

$$\min_{\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_P, \boldsymbol{x}_A, \boldsymbol{x}_P} \boldsymbol{\mathcal{F}}_1(\boldsymbol{\Theta}_A, \boldsymbol{x}_A) + \boldsymbol{\mathcal{F}}_2(\boldsymbol{\Theta}_P, \boldsymbol{x}_P), \qquad (29)$$

where

$$\boldsymbol{\mathcal{F}}_1 = \alpha_1 \left(\|\boldsymbol{z}_A - \boldsymbol{\Phi}\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e})\|^2 + \beta_1 \|\boldsymbol{x}_A - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) - \boldsymbol{t}_1\|^2 \right), \qquad (30)$$

and

$$\boldsymbol{\mathcal{F}}_2 = \alpha_2 \left(\|\boldsymbol{z}_P - \boldsymbol{\Phi}\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e})\|^2 + \beta_2 \|\boldsymbol{x}_P - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e}) - \boldsymbol{t}_2\|^2 \right). \qquad (31)$$

With the help of the alternating direction method of multipliers (ADMM) [56], the problem (29) can be solved by a sequential update of the six variables, i.e., $\boldsymbol{\Theta}_A$, $\boldsymbol{\Theta}_P$, $\boldsymbol{x}_A$, $\boldsymbol{x}_P$, $\boldsymbol{t}_1$, and $\boldsymbol{t}_2$.

1) The update of $\boldsymbol{\Theta}_A$ while fixing other variables:

$$\min_{\boldsymbol{\Theta}_A} \|\boldsymbol{z}_A - \boldsymbol{\Phi}\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e})\|^2 + \beta_1 \|\boldsymbol{x}_A - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) - \boldsymbol{t}_1\|^2, \qquad (32)$$

which can be solved using the steepest descent and back-propagation optimization methods [53]. Note that $\beta_1 \|\boldsymbol{x}_A - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) - \boldsymbol{t}_1\|^2$ in (32) can be regarded as the denoising of $\boldsymbol{x}_A - \boldsymbol{t}_1$, which also serves as a proximity regularization that forces $\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e})$ to be close to $\boldsymbol{x}_A - \boldsymbol{t}_1$. This second term provides additional stabilizing and robustifying effect to the back-propagation method.

2) The update of $\boldsymbol{x}_A$ while fixing other variables:

$$\min_{\boldsymbol{x}_A} \|\boldsymbol{x}_A - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) - \boldsymbol{t}_1\|^2, \qquad (33)$$

[3]Another solution is to design a suitable explicit regularization term for decoding sensing signal spectrums and use the explicit and implicit priors jointly [54]. This is left for the future work.

which can be regarded as a denoising problem for $\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) + \boldsymbol{t}_1$. Thus, we have

$$\hat{\boldsymbol{x}}_A = \boldsymbol{\mathcal{D}}(\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) + \boldsymbol{t}_1), \qquad (34)$$

where $\boldsymbol{\mathcal{D}}(\cdot)$ represents the denoising operator that could be well-studied plug-and-play algorithms [57] or a simpler steepest-descent (SD) operator. We present the update equation for SD method as

$$\boldsymbol{x}_A^{(j+1)} = \boldsymbol{x}_A^{(j)} - s\left(\boldsymbol{x}_A^{(j)} - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) - \boldsymbol{t}_1\right), \qquad (35)$$

where s is the steepest-descent step size, and j is the inner loop iteration number.

3) The update of $\boldsymbol{t}_1$ while fixing other variables: Because $\boldsymbol{t}_1$ can be regarded as the Lagrange multipliers vector, $\boldsymbol{t}_1$ can be updated according to the augmented Lagrangian method [55] as

$$\boldsymbol{t}_1^{(k+1)} = \boldsymbol{t}_1^{(k)} + \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A^{(k)}}(\boldsymbol{e}) - \mathbf{x}_A^{(k)}, \qquad (36)$$

where k denotes the outer loop iteration number.

4) The update of $\boldsymbol{\Theta}_P$ while fixing other variables: Because the network with parameter $\boldsymbol{\Theta}_P$ is trained independently, we can update $\boldsymbol{\Theta}_P$ by solving

$$\min_{\boldsymbol{\Theta}_P} \|\mathbf{z}_P - \boldsymbol{\Phi}\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e})\|^2 + \beta_2 \|\mathbf{x}_P - \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e}) - \boldsymbol{t}_2\|^2, \qquad (37)$$

with the same method as in (32).

5) The update of $\boldsymbol{x}_P$ while fixing other variables: To minimize the difference between $\mathbf{x}_P$ and $\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_A}(\boldsymbol{e}) + \mathbf{t}_1$, we can update $\mathbf{x}_P$ as

$$\hat{\mathbf{x}}_P = \boldsymbol{\mathcal{D}}(\boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P}(\boldsymbol{e}) + \mathbf{t}_2). \qquad (38)$$

where $\boldsymbol{\mathcal{D}}$ is the same kind of denoising operator as (34).

6) The update of $\boldsymbol{t}_2$ while fixing other variables: According to the augmented Lagrangian method [55], $\boldsymbol{t}_2$ can be updated as

$$\boldsymbol{t}_2^{(k+1)} = \boldsymbol{t}_2^{(k)} + \boldsymbol{\mathcal{T}}_{\boldsymbol{\Theta}_P^{(k)}}(\boldsymbol{e}) - \mathbf{x}_P^{(k)}. \qquad (39)$$

Algorithm 3 summarizes the steps to perform the aforementioned decoding methods, and then recover the original amplitude and phase spectrums. Specifically, after decoding, we obtain the estimated $\boldsymbol{H}_{A'}\{i\}$ and $\boldsymbol{H}_{P'}\{i\}$. Let the first column in $\boldsymbol{H}_A\{i\}$ and $\boldsymbol{H}_P\{i\}$ be the same as that of $\boldsymbol{H}_{A'}\{i\}$ and $\boldsymbol{H}_{P'}\{i\}$ as

$$\boldsymbol{H}_A\{i\}[1,:] = \boldsymbol{H}_A'\{i\}[1,:], \qquad (40)$$

and

$$\boldsymbol{H}_P\{i\}[1,:] = \boldsymbol{H}_{P'}\{i\}[1,:]. \qquad (41)$$

For the second to the last columns ($j = 2, \ldots, L$), we have

$$\boldsymbol{H}_A\{i\}[j,:] = \boldsymbol{H}_{A'}\{i\}[j-1,:] + \boldsymbol{H}_{A'}\{i\}[j,:], \qquad (42)$$

and

$$\boldsymbol{H}_P\{i\}[j,:] = \boldsymbol{H}_{P'}\{i\}[j-1,:] + \boldsymbol{H}_{P'}\{i\}[j,:]. \qquad (43)$$

Note that because of the independent iterative training of the two networks and the use of the ADMM method, α_1 and α_2 have no effect on the objective function. The running time

Algorithm 3 The Algorithm for Inverse Semantic-Aware Decoding

---

Input:

- The weight parameters: β_1 and β_2
- The number of inner iterations of the denoising operator for updating $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$: N_J
- The steepest-descent parameters for updating $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$, respectively

Output: The original amplitude and phase spectrums, i.e., $\mathbf{H}_A^{(i)}$ and $\mathbf{H}_P^{(i)}$ $(i = 1, \ldots, T)$

1: *## Reconstruction of the $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$*
2: Initialize the iteration number $k = 0$
3: Set $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$ randomly
4: **while** Not converged **do**
5:    Update $\boldsymbol{\Theta}_A$ by solving (32) using steepest descent and back-propagation methods
6:    Update $\boldsymbol{x}_A$ according to (34)
7:    Update $\boldsymbol{t}_1$ according to (36)
8:    Update $\boldsymbol{\Theta}_P$ by solving (37) using steepest descent and back-propagation methods
9:    Update $\boldsymbol{x}_P$ according to (38)
10:   Update $\boldsymbol{t}_2$ according to (39)
11:   Let $k \leftarrow k + 1$
12: Record $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$ after converged
13: *## Differential decoding*
14: Recover $\mathbf{H}_{A'}\{i\}$ and $\mathbf{H}_{P'}\{i\}$ $(i = 1, \ldots, T)$ according to the definition of $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$, i.e., (24) and (25)
15: **for** Every $\boldsymbol{H}_{A'}\{i\}$ and $\boldsymbol{H}_{P'}\{i\}$ **do**
16:   Create empty $\boldsymbol{H}_A\{i\}$ and $\boldsymbol{H}_P\{i\}$ to record the decoded results
17:   Obtain $\boldsymbol{H}_A\{i\}$ and $\boldsymbol{H}_P\{i\}$ according to (40), (41), (42), and (43)
18: **return** $\mathbf{H}_A\{i\}$ and $\mathbf{H}_P\{i\}$ $(i = 1, \ldots, T)$

---

is mainly taken in updating $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$ since the inner denoising operators work efficiently. In Section VI, we set the inner iteration numbers of the denoising operators for updating $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$ to be both 600, and the outer loop maximal iteration number is 18, i.e., 18 ADMM iterations. The average running time for decoding one MetaSpectrum, which is obtained by encoding 20 original amplitude spectrums, is about 1 minute with the experiment setting in Section VI. While the self-supervised decoding approach may not be optimal for high real-time decoding of sensing signal data, our proposed method can be effectively applied to sensing tasks that require large amounts of historical data storage for analysis, such as healthcare monitoring, sleeping position detection, and historical intrusion or walking behavior analysis.

With the decoded $\mathbf{H}_A\{i\}$ and $\mathbf{H}_P\{i\}$ $(i = 1, \ldots, T)$, the original signal at each moment can be recovered to the form of complex matrices. Then, the 2D AoA and ToF can be jointly estimated [36], which can be used to complete a series of sensing tasks. Thus, the steering matrices of L-shaped array in the x and y directions, which describe how the sensor array uses each individual element to select a spatial path for the

transmission, can be expressed as

$$
\mathbf{A}_x = \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j2\pi f_k d \cos(\theta_1) \sin(\varphi_1)} & \cdots & e^{-j2\pi f_k d \cos(\theta_I) \sin(\varphi_I)} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi f_k m d \cos(\theta_1) \sin(\varphi_1)} & \cdots & e^{-j2\pi f_k m d \cos(\theta_I) \sin(\varphi_I)} \end{bmatrix},
\tag{44}
$$

and

$$
\mathbf{A}_y = \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j2\pi f_k d \sin(\theta_1) \sin(\varphi_1)} & \cdots & e^{-j2\pi f_k d \sin(\theta_I) \sin(\varphi_I)} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi f_k n d \sin(\theta_1) \sin(\varphi_1)} & \cdots & e^{-j2\pi f_k n d \sin(\theta_I) \sin(\varphi_I)} \end{bmatrix},
\tag{45}
$$

respectively. Inspired by [58], here, we take multiple subcarriers into consideration and extend the 2D AoA estimation into three dimensions, to acheive joint 2D AoA and ToF estimation via the following Proposition 2:

Proposition 2: The signal 2D AoA and ToF at time t can be estimated using

$$
P_{3D}(\theta, \varphi, \tau, t) = \frac{1}{\mathbf{A}_{0x'y'}^{\mathrm{H}} \mathbf{E}_N(t) \mathbf{E}_N^{\mathrm{H}}(t) \mathbf{A}_{0x'y'}},
\tag{46}
$$

where P_{3D} describes the signal magnitude for a given set of (θ, φ, τ), the superscript H is the conjugate transpose operator, $\mathbf{E}_N(t)$ is the noise subspace obtained by decomposing the auto-correlation matrix of the smoothed original signal at time t [58], $\mathbf{A}_{0x'y'}$ is the steering matrix that is obtained using (44) and (45) $0 < k' < K$, $0 < m' < M$, and $0 < n' < N$. Equation (47), as shown at the bottom of the next page.

Thus, we complete all the processes of the inverse semantic-aware wireless sensing framework. Specifically, we use **Algorithm 2** to sample the task-related signal spectrums. With the RIS, **Algorithm 1** can encode the sensing data, thus greatly reducing the data volume to be stored or transmitted. We use the self-supervised decoding **Algorithm 3** to recover the original sensing data. Finally, with the help of Proposition 2, various sensing tasks can be performed. For example, intrusion detection can be achieved by detecting the change of estimated 2D AoA, and the human walking trajectory can be tracked by estimating the 2D AoA and the ToF of the signals.

VI. EXPERIMENTS RESULTS

Since the key contribution of this paper is to achieve the inverse semantic-aware encoding and decoding of the sensing data with the help of RIS, we aim to answer the following research questions via experiments:

Q1) Can the proposed self-supervised decoding scheme recover the original signal spectrums and ensure the accomplishment of sensing tasks?

Q2) Can the amplitude response matrix of RIS, i.e., codebook, encrypt the sensing data?

Q3) Compared with the existing uniform sampling method, can the proposed semantic hash sampling method help to achieve more accurate completions of the sensing tasks?

We first present the experimental platform and the parameter setting of our proposed algorithms, and then answer the above questions through experimental evaluations.
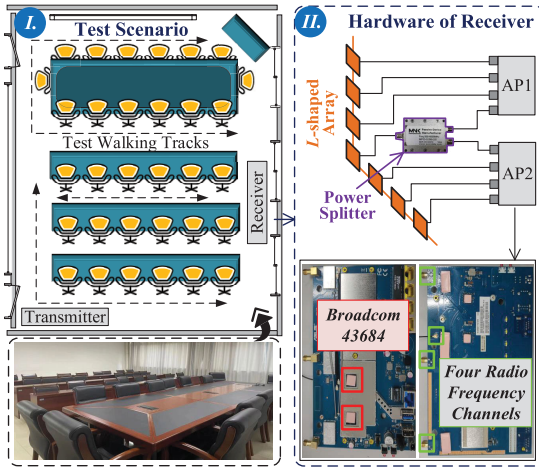
Fig. 5. Test scenario and hardware of the receiver.

A. Experiments Setting

To collect sensing data from the real-world scenario, we use three access points (APs) based on the IEEE 802.11ax protocol to build a test platform [15]. The collected sensing data is used to conduct a comprehensive evaluation of our proposed algorithms. The specific experimental scenario and hardware equipment are shown in Fig. 5. Specifically, the test scenario is a conference room with tables and chairs. Inside the room, one AP acts as a transmitter to send OFDM wireless signals with the total bandwidth of 160 MHz and 2048 sub-carriers. The center frequency of the sub-carriers is 5.805 GHz. As shown in Fig. 5 (Part II), the other two APs form an receiver with L-shaped active sensor array via a power splitter to receive signals. Since the investigation of STAR-RIS hardware is still at a very early stage, we simulate the amplitude and phase response matrices of the transmissive elements using a signal processor [28], [29], [38]. During the experiment, the data packet transmission rate, i.e., transmission frequency, is 100 Hz, which means 100 packets are transmitted per second. The human target walks along the preset trajectory to complete the data collection.

The experimental platform for running our proposed algorithms is built on a generic Ubuntu 20.04 system with an AMD Ryzen Threadripper PRO 3975WX 32-Cores CPU and an NVIDIA RTX A5000 GPU. In the self-supervised decoding

Algorithm 3, two U-net without the skip connections [53] are used as the self-supervised neural networks, i.e., $\mathcal{T}_{\boldsymbol{\Theta}_A}(\boldsymbol{e})$ and $\mathcal{T}_{\boldsymbol{\Theta}_P}(\boldsymbol{e})$. The input to the network, i.e., $\boldsymbol{e}$, is a random vector that has the same size as $\boldsymbol{x}_A$ and $\boldsymbol{x}_P$ to be recovered. During the decoding of one MetaSpectrum, $\boldsymbol{e}$ is fixed in each ADMM iteration. In addition, to avoid the local minimum that the networks stuck in the last iteration, $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$ are set to zero when each ADMM iteration is finished. In other words, both $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$ are re-trained in each iteration.

B. Experiments Performance Analysis

1) Effectiveness and Efficiency of the Proposed Inverse Semantic Decoding Method (Q1): We first set the data compression ratio as 10%. As shown in Fig. 1 (Part I), starting from 3 seconds, we select one pair of amplitude and phase spectrums in each 0.1 second time segment by using **Algorithm 2**, for RIS-aided encoding. Using the encoding **Algorithm 1** presented in Section IV-A, we can obtain one amplitude MetaSpectrum and one phase MetaSpectrum as shown in Fig. 1 (Part III) for every 10 pairs of signal spectrums. The decoded results after 15 iterations of the outer loop are shown in Fig. 1 (Part II). For both amplitude and phase spectrums, we observe that the difference between the decoded and the original spectrums is basically negligible. We present a detailed comparison of the decoded and the original amplitude spectrums in Fig. 1 (Part IV). This proves the effectiveness of our encoding and decoding methods.

In addition to the visual contrast, we show in Fig. 6 how the proposed semantic hash matrix changes with the number of outer loop decoding iterations. We set the data compression ratio as 5%. We observe that, as the number of outer loop iterations increases, both the decoded amplitude and phase spectrums at time 4.5 seconds are gradually close to the ground truth spectrums. Moreover, the Hamming distance between the semantic hash matrices of the decoded pair of amplitude and phase spectrums and that of the original signal spectrums is gradually reducing. Specifically, we can see that 12 iterations can make the Hamming distance only 2, which takes about 40 seconds on average. Furthermore, the estimated 2D AoA values using the decoded spectrum after 12 iterations are very close to the true values, which basically has no effect on

$$\mathbf{A}_{0x'y'} = [\mathbf{A}_0 \mathbf{A}_{x'} \mathbf{A}_{y'}]^{\mathrm{T}} = \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\mathbf{A}_0} \underbrace{\begin{matrix} e^{-j2\pi f_1 d\cos(\theta)\sin(\varphi)} \\ \vdots \\ e^{-j2\pi f_{k'} d\cos(\theta)\sin(\varphi)} \\ \vdots \\ e^{-j2\pi f_1 (m'-1)d\cos(\theta)\sin(\varphi)} \\ \vdots \\ e^{-j2\pi f_{k'}(m'-1)d\cos(\theta)\sin(\varphi)} \end{matrix}}_{\mathbf{A}_{x'}} \underbrace{\begin{matrix} e^{-j2\pi f_1 d\sin(\theta)\sin(\varphi)} \\ \vdots \\ e^{-j2\pi f_{k'} d\sin(\theta)\sin(\varphi)} \\ \vdots \\ e^{-j2\pi f_1 (n'-1)d\sin(\theta)\sin(\varphi)} \\ \vdots \\ e^{-j2\pi f_{k'}(n'-1)d\sin(\theta)\sin(\varphi)} \end{matrix}}_{\mathbf{A}_{y'}} \end{bmatrix}^{\mathrm{T}}$$
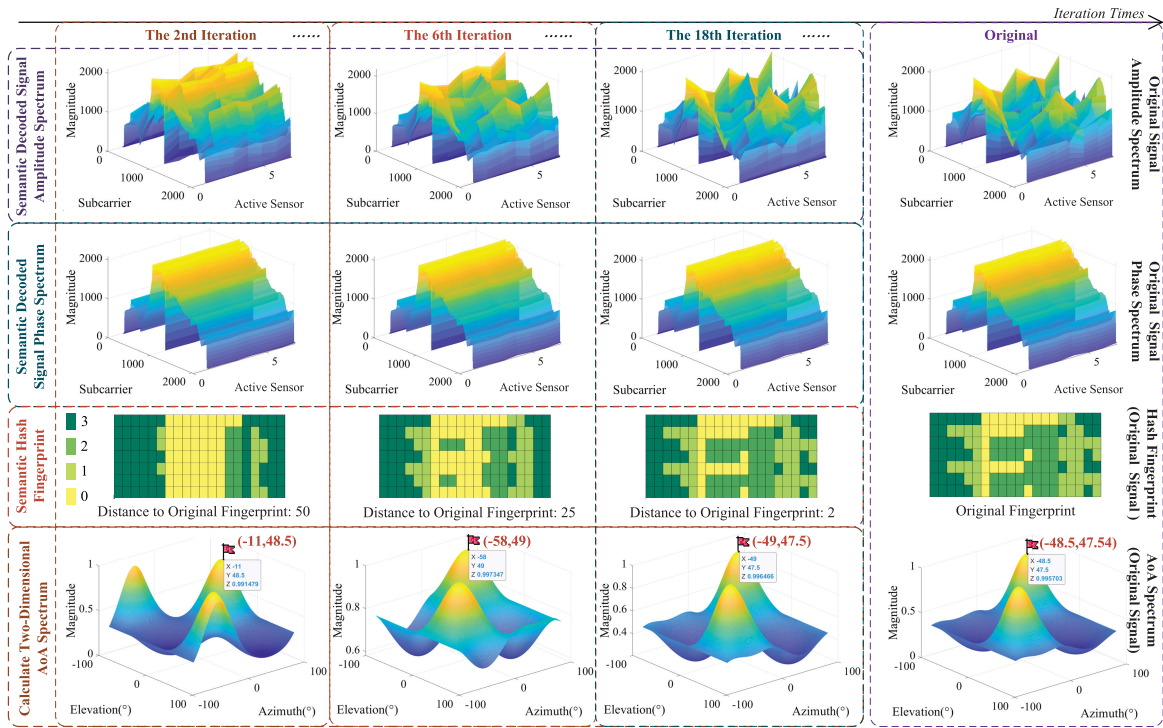
$$(47)$$

Fig. 6. The variation of the decoded amplitude and phase spectrums at time 4.5 s in the experiment, the corresponding semantic hash matrix, the estimated 2D AoA spectrum by Proposition 2 with the number of outer loop decoding iterations, where the data compression ratio is 5%.
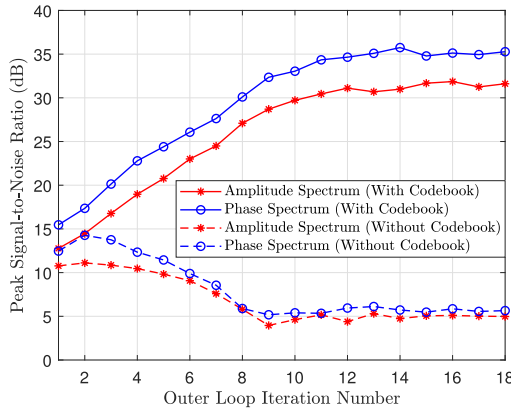

Fig. 7. The PSNR values versus the number of outer loop decoding iterations with or without codebook.
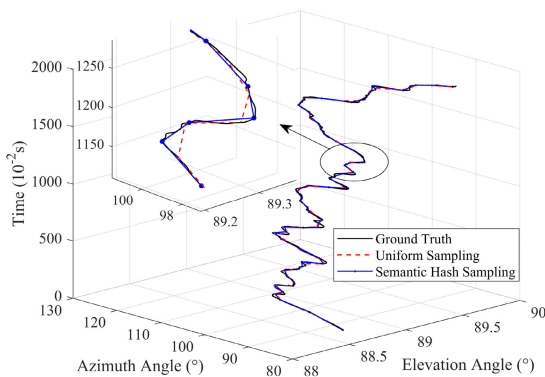

Fig. 8. Comparison between different sampling methods and ground truth in terms of 2D AoA changes with movement of human.

Fig. 9. Comparison of azimuth AoA estimation results that are obtained by using the original and decoded signals, respectively.

the practical sensing tasks. This proves the efficiency of our encoding and decoding methods.

2) Effectiveness of Using the Amplitude Response Matrix of the RIS as the Codebook (Q2): Figure 7 depicts the average peak signal-to-noise ratio (PSNR) values 10 experiments versus the number of outer loop decoding iterations, with or without the codebook $\mathbf{\Phi}_A^{(i)}$. If the codebook is available, we observe that the PSNR values of both the amplitude and phase spectrums are increasing as the number of iterations increases, and gradually reach a plateau after about 10 iterations. Although minor fluctuations occur at higher
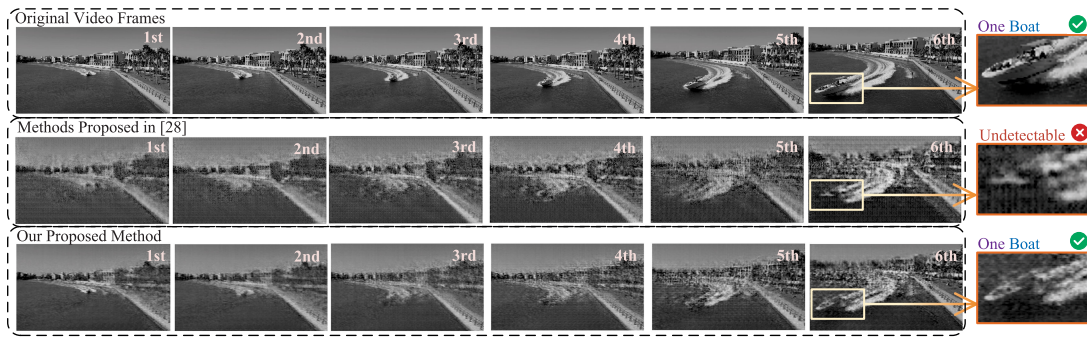
Fig. 10. Early-stop decompress results the real video frames in [45] using the method proposed in [25] and our method, respectively. Six images are compressed into one image.

iteration steps because of the dynamic nature of the decoding process, the overall trend demonstrates the effectiveness of our self-supervised decoding scheme in recovering the original signal spectrums. However, if no codebook is available or the codebook is wrong, the PSNR values decrease as the number of iterations increases. The reason is that the parameters of two decoding network, i.e., $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_P$, are learned according to a wrong objective function.

3) Effectiveness of the Proposed Semantic Hash Sampling Method (Q3): Based on the sensing data extracted via two different sampling methods, i.e., red line for the uniform sampling and black line for the semantic hash sampling methods, Fig. 8 displays estimated elevation and azimuth AoA changes over time. Note that the estimation results under two sampling schemes are obtained using the decoded amplitude and phase spectrums with the data compression ratio as 5%. First, we observe that the both elevation and azimuth AoA at every moment can be accurately estimated using the decoded data. This further validates the effectiveness of our proposed encoding and decoding algorithms (for Q1). Furthermore, by comparing the black and red lines, it can be seen that the proposed semantic hash sampling method is more efficient and effective than uniform sampling in describing the details of AoA changes, as shown in the enlarged part in Fig. 8. Because these changes are typically more informative, this shows the effectiveness of our proposed semantic hash sampling method. To compare the two schemes numerically, we consider the MSE between the ground truth and the 2D AoA estimation results after interpolation. By calculating, we obtain that the estimation error of the semantic hash sampling is 0.89, which is 67% lower than that of uniform sampling scheme whose estimation error is 2.7.

In addition to the walking human, stationary objects such as tables and chairs in the conference room also reflect wireless signals. Thus, the information that can be extracted from the signal spectrums at one certain moment is rich. Taking the azimuth AoA as an example, Fig. 9 shows the comparison of the azimuth AoA estimation results that are obtained by using the original and decoded signals, respectively, with a data compression ratio of 5%. The results demonstrate that our encoding and decoding methods effectively preserve the semantic information related to the sensing tasks, which is evident from two aspects: First, the relative magnitude characteristics among different azimuth AoA estimated from the decoded signals are consistent with the ground truth, i.e.,

the azimuth AoA estimated from the original signals. For instance, the ground truth indicates that the stronger signals' azimuth AoA are in the range of $10° - 40°$ and $110° - 150°$, as marked by the red and black boxes, respectively. In addition, the signals with AoA in $40° - 110°$ are weaker. These features are almost entirely preserved in the estimation results obtained using the decoded data. Second, we observe that the AoA estimation results of the first several strongest signals are almost unchanged before and after the inverse semantic-aware encoding and decoding, e.g., the signals marked by the red and black boxes in Fig. 9, respectively. This indicates that our proposed algorithms can effectively preserve the phase characteristics (for Q1).

VII. Conclusion and Future Directions

We have designed an inverse semantic-aware wireless sensing framework. The amplitude response matrix of the RIS can be effectively used to generate the codebook as prior knowledge for decoding. We have shown that our proposed RIS-aided encoding method can achieve effective data compression. When selecting the signal spectrums to be encoded, our proposed semantic hash sampling method is significantly better than the widely used uniform sampling method. Moreover, the self-supervised decoding method can recover signal amplitude and phase spectrums to achieve various wireless sensing tasks without affecting the performance. Since the decoding method does not require any pre-training, it can greatly save network resources. As the demand for sensing data increases, our proposed framework can contribute to building a resource-friendly next-generation Internet.

There are two potential future research directions.

- *Inverse Semantic-aware Transmission of Images.* We can explore the application of inverse semantic-aware encoding and decoding for images or audio. In the context of surveillance services, for example, a camera captures a bay to detect boats. The surveillance videos require significant storage resources. The goal is to compress multiple video frames, such as the six frames depicted in Fig. 10, into a single frame. The original frames can then be reconstructed using the proposed self-supervised decoding algorithm.
- *Cantor or Szudzik Pairing Compression.* In this paper, we encoded the amplitude and the phase spectrum separately. A possible improvement is to use the pairing

functions, e.g., cantor [59] or szudzik [60] pairing functions, to combine the two spectrum into one. As shown in Fig. 1, the pairing compression can be used as an operation after obtaining amplitude and phase spectrums to further compress the sensing data.

REFERENCES

[1] H. Du et al., "Attention-aware resource allocation and QoE analysis for metaverse xURLLC services," 2022, *arXiv:2208.05438*.

[2] M. K. Farshbafan, W. Saad, and M. Debbah, "Common language for goal-oriented semantic communications: A curriculum learning framework," 2021, *arXiv:2111.08051*.

[3] H. Seo, J. Park, M. Bennis, and M. Debbah, "Semantics-native communication with contextual reasoning," 2021, *arXiv:2108.05681*.

[4] W. Yang et al., "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, 1st Quart., 2023.

[5] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented multi-user semantic communications for VQA," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 553–557, Mar. 2022.

[6] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, "Deep learning enabled semantic communications with speech recognition and synthesis," 2022, *arXiv:2205.04603*.

[7] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, and D. I. Kim, "AI-generated incentive mechanism and full-duplex semantic communications for information sharing," 2023, *arXiv:2303.01896*.

[8] M. Zhu, C. Feng, J. Chen, C. Guo, and X. Gao, "Video semantics based resource allocation algorithm for spectrum multiplexing scenarios in vehicular networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC Workshops)*, Jul. 2021, pp. 31–36.

[9] H. Xie, Z. Qin, G. Y. Li, and B. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.

[10] L. Chen, R. Li, H. Zhang, L. Tian, and N. Chen, "Intelligent fall detection method based on accelerometer data from a wrist-worn smart watch," *Measurement*, vol. 140, pp. 215–226, Jul. 2019.

[11] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13086–13095, Aug. 2022.

[12] P. P. Ray, "A survey of IoT cloud platforms," *Future Comput. Informat. J.*, vol. 1, nos. 1–2, pp. 35–46, Dec. 2016.

[13] M. Hassanalieragh et al., "health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges," in *Proc. IEEE Int. Conf. Services Comput.*, Jun. 2015, pp. 285–292.

[14] R. R. Singh et al., "IoT embedded cloud-based intelligent power quality monitoring system for industrial drive application," *Future Gener. Comput. Syst.*, vol. 112, pp. 884–898, Nov. 2020.

[15] F. Gringoli, M. Cominelli, A. Blanco, and J. Widmer, "AX-CSI: Enabling CSI extraction on commercial 802.11ax Wi-Fi platforms," in *Proc. 15th ACM Workshop Wireless Netw. Testbeds, Experim. Eval. CHaracterization*, Jan. 2022, pp. 46–53.

[16] Z. Gao, L. Dai, S. Han, C. I, Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.

[17] K. Niu, X. Wang, F. Zhang, R. Zheng, Z. Yao, and D. Zhang, "Rethinking Doppler effect for accurate velocity estimation with commodity WiFi devices," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2164–2178, Jul. 2022.

[18] R. A. Ramadan, "Efficient intrusion detection algorithms for smart cities-based wireless sensing technologies," *J. Sensor Actuator Netw.*, vol. 9, no. 3, p. 39, Aug. 2020.

[19] J. Liu, H. Liu, Y. Chen, Y. Wang, and C. Wang, "Wireless sensing for human activity: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1629–1645, 3rd Quart., 2020.

[20] S. Yue, Y. Yang, H. Wang, H. Rahul, and D. Katabi, "BodyCompass: Monitoring sleep posture with wireless signals," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–25, Jun. 2020.

[21] K. Gao, H. Wang, H. Lv, and W. Liu, "Toward 5G NR high-precision indoor positioning via channel frequency response: A new paradigm and dataset generation method," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2233–2247, Jul. 2022.

[22] X. Gong, J. Liu, S. Yang, G. Huang, and Y. Bai, "A usability-enhanced smartphone indoor positioning solution using compressive sensing," *IEEE Sensors J.*, vol. 22, no. 3, pp. 2823–2834, Feb. 2022.

[23] R. Zhang, K. Xiong, Y. Lu, P. Fan, D. W. K. Ng, and K. B. Letaief, "Energy efficiency maximization in RIS-assisted SWIPT networks with RSMA: A PPO-based approach," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1413–1430, May 2023.

[24] X. Yuan, J. Chen, N. Zhang, J. Ni, F. R. Yu, and V. C. M. Leung, "Digital twin-driven vehicular task offloading and IRS configuration in the Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24290–24304, Dec. 2022.

[25] H. Zhang, B. Di, K. Bian, Z. Han, H. V. Poor, and L. Song, "Toward ubiquitous sensing and localization with reconfigurable intelligent surfaces," *Proc. IEEE*, vol. 110, no. 9, pp. 1401–1422, Sep. 2022.

[26] H. Zhang, H. Zhang, B. Di, K. Bian, Z. Han, and L. Song, "MetaLocalization: Reconfigurable intelligent surface aided multi-user wireless indoor localization," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7743–7757, Dec. 2021.

[27] Z. Chen, P. Chen, Z. Guo, Y. Zhang, and X. Wang, "A RIS-based vehicle DOA estimation method with integrated sensing and communication system," 2022, *arXiv:2204.11626*.

[28] J. Tang, M. Cui, S. Xu, L. Dai, F. Yang, and M. Li, "Transmissive RIS for 6G communications: Design, prototyping, and experimental demonstrations," 2022, *arXiv:2206.15133*.

[29] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, May 2022.

[30] Y. Zhang, B. Di, H. Zhang, M. Dong, L. Yang, and L. Song, "Dual codebook design for intelligent omni-surface aided communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9232–9245, Nov. 2022.

[31] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 65–88, Mar. 2021.

[32] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[33] Z. Meng, Z. Yu, K. Xu, and X. Yuan, "Self-supervised neural networks for spectral snapshot compressive imaging," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2602–2611.

[34] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng, "Adaptive non-local sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 2104–2111, Oct. 2017.

[35] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44304–44321, 2021.

[36] Y. Hua, T. K. Sarkar, and D. Weiner, "L-shaped array for estimating 2-D directions of wave arrival," in *Proc. 32nd Midwest Symp. Circuits Syst.*, 1989, pp. 390–393.

[37] H. Zheng, Z. Shi, C. Zhou, M. Haardt, and J. Chen, "Coupled coarray tensor CPD for DOA estimation with coprime L-shaped array," *IEEE Signal Process. Lett.*, vol. 28, pp. 1545–1549, 2021.

[38] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3134–3138, Sep. 2021.

[39] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[40] H. Du, J. Zhang, J. Cheng, and B. Ai, "Millimeter wave communications with reconfigurable intelligent surfaces: Performance analysis and optimization," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2752–2768, Apr. 2021.

[41] H. Du et al., "Performance and optimization of reconfigurable intelligent surface aided THz communications," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3575–3593, May 2022.

[42] M. Qiao, X. Liu, and X. Yuan, "Snapshot spatial-temporal compressive imaging," *Opt. Lett.*, vol. 45, no. 7, pp. 1659–1662, Jul. 2020.

[43] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[44] T. J. Cui et al., "Information metamaterial systems," *iScience*, vol. 23, no. 8, Aug. 2020, Art. no. 101403.

[45] A.-F. Perrin, V. Krassanakis, L. Zhang, V. Ricordel, M. P. D. Silva, and O. L. Meur, "EyeTrackUAV2: A large-scale binocular eye-tracking dataset for UAV videos," *Drones*, vol. 4, no. 1, p. 2, Jan. 2020.

[46] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *Proc. USENIX Symp. Netw. Syst. Des. Implement.*, 2016, pp. 165–178.

[47] L. Du, A. T. S. Ho, and R. Cong, "Perceptual hashing for image authentication: A survey," *Signal Process., Image Commun.*, vol. 81, Feb. 2020, Art. no. 115713.

[48] T. Tuncer, S. Dogan, M. Abdar, and P. Pławiak, "A novel facial image recognition method based on perceptual hash using quintet triple binary pattern," *Multimedia Tools Appl.*, vol. 79, nos. 39–40, pp. 29573–29593, Oct. 2020.

[49] Z. Khanam and M. N. Ahsan, "Implementation of the pHash algorithm for face recognition in a secured remote online examination system," *Int. J. Adv. Sci. Res. Eng.*, vol. 4, no. 11, pp. 1–5, 2018.

[50] S. Jalali and X. Yuan, "Snapshot compressed sensing: Performance bounds and algorithms," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8005–8024, Dec. 2019.

[51] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.

[52] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5882–5891.

[53] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[54] G. Mataev, P. Milanfar, and M. Elad, "DeepRED: Deep image prior powered by red," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 1–10.

[55] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sep. 2010.

[56] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[57] S. Sreehari et al., "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 408–423, Dec. 2016.

[58] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proc. ACM Conf. Special Interest Group Data Commun.*, Aug. 2015, pp. 269–282.

[59] M. Lisi, "Some remarks on the Cantor pairing function," *Le Matematiche*, vol. 62, no. 1, pp. 55–65, Jan. 2007.

[60] M. Szudzik, "An elegant pairing function," in *Proc. Wolfram Sci. Conf.*, 2006, pp. 1–12.

Dusit Niyato (Fellow, IEEE) received the B.Eng. degree from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests are in the areas of the Internet of Things (IoT), machine learning, and incentive mechanism design.

Jiawen Kang received the Ph.D. degree from the Guangdong University of Technology, China, in 2018. He was a Post-Doctoral Researcher with Nanyang Technological University, Singapore, from 2018 to 2021. He is currently a Full Professor with the Guangdong University of Technology, China. His research interests mainly focus on blockchain, security, and privacy protection in wireless communications and networking.

Zehui Xiong (Member, IEEE) received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He was a Visiting Scholar with Princeton University and University of Waterloo. He is currently an Assistant Professor with the Singapore University of Technology and Design and also an Honorary Adjunct Senior Research Scientist with the Alibaba-NTU Singapore Joint Research Institute, Singapore. His research interests include wireless communications, the Internet of Things, blockchain, edge intelligence, and metaverse.

Junshan Zhang (Fellow, IEEE) received the Ph.D. degree from the School of ECE, Purdue University, in August 2000. He is currently a Professor with the ECE Department, University of California at Davis. His research interests fall in the general field of information networks and data science, including edge intelligence, reinforcement learning, network optimization and control, game theory, with applications in connected and automated vehicles, and 5G and beyond.

Hongyang Du (Graduate Student Member, IEEE) received the B.Sc. degree from Beijing Jiaotong University, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Energy Research Institute @ NTU, Nanyang Technological University, Singapore, under the Interdisciplinary Graduate Program. He was recognized as an exemplary reviewer of the IEEE TRANSACTIONS ON COMMUNICATIONS in 2021. He was a recipient of IEEE Daniel E. Noble Fellowship Award in 2022.

Jiacheng Wang received the Ph.D. degree from the School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. He is currently a Research Associate in computer science and engineering with Nanyang Technological University, Singapore. His research interests include wireless sensing, semantic communications, and metaverse.

Xuemin (Sherman) Shen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on network resource management, network security, the Internet of Things, and 5G and beyond.