

Embodied Communication: Sensing-Induced Reliability Fields and Capacity Bounds

Yulin Shao

Abstract—This paper introduces embodied communication, a new wireless communication modality in which information is imprinted onto environmental states and recovered by the receiver through sensing. No dedicated communication transmitter is activated, and no additional communication spectrum is occupied; instead, the sensed environment itself becomes the carrier of information. The key insight is that sensing must be reinterpreted for communication. Rather than asking how accurately an unknown physical state can be estimated, embodied communication asks how reliably two states can be distinguished. We formalize this idea through a multi-snapshot radio frequency (RF) sensing model and derive a sensing-induced reliability field that quantifies the distinguishability between physical states. This field turns embodied symbol design into a geometric packing problem shaped by the sensing resolution of the infrastructure. For this embodied channel, we characterize the finite-snapshot ϵ -capacity through achievable designs and converses. We develop lattice-based codebooks, obtain a closed-form hexagonal design under a main-lobe approximation, and establish information-theoretic and geometric upper bounds. We further reveal an intrinsic sensing-duration tradeoff: more sensing snapshots improve reliability, but also lengthen each embodied symbol, leading to a finite optimal sensing time. These results expose a latent communication pathway in sensing-enabled infrastructure and show how the environment can be transformed from a passive backdrop into an active information carrier.

Index Terms—Embodied communication, ISAC, ambient modulation, Sensing distinguishability, lattice packing.

I. INTRODUCTION

A. Infrastructure with Sensing

The evolution toward 6G wireless networks is marked by a fundamental reorientation of the role of infrastructure [1], [2]. The base station (BS) is no longer envisioned merely as a node that transports data packets. Instead, it is increasingly becoming a multimodal sensing platform that perceives, interprets, and interacts with the physical world. Integrated sensing and communication (ISAC) [1], for instance, reuses radio-frequency (RF) spectral and hardware resources to endow the BS with radar-like sensing capability. Vision sensors such as cameras [3], [4] and LiDAR [5], [6], increasingly co-located with antenna arrays, further equip the BS with visual and geometric perception.

The sensing capability of wireless infrastructure is typically exploited in two ways.

- *Beamforming assistance.* Environmental awareness is used to enhance the wireless link itself. A representative application is predictive and site-specific beamforming [7], [8]. By tracking user mobility and reconstructing the surrounding

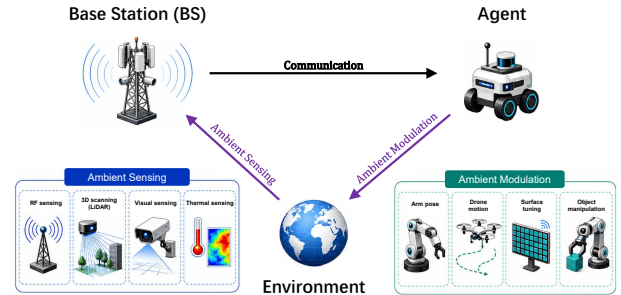


Figure 1: Conceptual illustration of embodied communication. Information is embodied in agent-induced environmental states and decoded through BS sensing observations, rather than transmitted by a dedicated uplink waveform.

geometry, the BS can preemptively steer beams around obstacles, maintain beam alignment during rapid movement, and reduce the overhead of channel state information (CSI) acquisition. In this mode, sensing forms a closed loop at the BS side: the BS observes the environment and immediately uses the observation to improve its own communication.

- *Sensing as a service.* The BS may also operate as part of a distributed sensing network, extracting physical information from the environment and delivering it to downstream applications [9]. Examples include target localization, velocity tracking, 3D scene reconstruction, micro-motion detection, autonomous vehicle coordination, industrial digital twins, and smart city analytics [10], [11]. In this mode, sensing is provided as a standalone service and is not necessarily coupled to the immediate needs of a communication link.

While both modes represent important advances, they share an implicit assumption: the physical environment is a passive object of observation. The world is seen, tracked, and measured, but it does not intentionally carry information for the BS to decode.

B. The Vision of Embodied Communication

Consider the relationships illustrated in Fig. 1. There are three entities: the BS, the communication agent, and the environment. Two links are already familiar. The first is the sensing link from the environment to the BS, through which the BS observes physical states. The second is the conventional communication link from the BS to the agent, through which information is delivered by a designed waveform.

There is, however, a third link that is easily overlooked. It points from the agent to the environment. After all, the

agent is physically situated in the environment and can deliberately alter environmental states, whether by changing posture, manipulating nearby objects, blocking or revealing a thermal source, or changing the reflective properties of surfaces. These actions are not merely physical side effects. If selected from an interpretable set of states, they can carry information.

Once this third arrow is acknowledged, a new communication pathway emerges. The agent’s deliberate manipulation of the environment, followed by the BS’s observation of that same environment, forms an information cascade from the agent to the BS through the environment. The agent does not generate a dedicated communication waveform. Rather, the agent inscribes a change onto the world, and the BS, through its sensing faculties, reads it. This constitutes a pathway through which information can flow. No additional transmitter is required, and no new spectrum is occupied. The pathway is already there, latent in the very architecture of a sensing-enabled network. It simply awaits intentional use.

We refer to this new form of communication as *embodied communication*. Formally,

Embodied communication is a communication modality in which information is encoded into deliberately induced environmental states and decoded from the receiver’s sensory observations.

The term “embodied” emphasizes that the information is physically instantiated in the state of the environment itself. It is inspired by the intellectual lineage of embodied AI and embodied cognition [12]–[15], where intelligence and meaning arise from an agent’s interaction with the physical world rather than from disembodied computation alone [13]. In a similar spirit, embodied communication treats the environment not as a passive backdrop for signal propagation, but as a medium through which expression can be enacted.

To study embodied communication, we need to identify the channel through which information flows. This channel is neither a conventional waveform channel nor a sensing channel alone. It is formed by the coupling between the agent’s physical modulation of the environment and the receiver’s sensing of the resulting state. We call this end-to-end channel the *embodied channel*. It consists of two coupled components:

- Ambient modulation: the agent-side operation that maps information to a controllable environmental state.
- Ambient sensing: the receiver-side operation that maps the induced environmental state to noisy sensory observations.

Together, they form an information-bearing channel through the environment.

This structure makes the embodied channel fundamentally different from a classical communication channel. In a conventional channel, the input alphabet consists of waveforms synthesized by a transmitter and constrained by communication resources such as power, bandwidth, and blocklength [16]. In an embodied channel, the input alphabet consists of physically realizable environmental states. It is constrained jointly by what the agent can realize and what the receiver can distinguish. The central design question is therefore: how many physical states can be carved out of a controllable environment and used as reliable embodied symbols?

Answering this question requires a different view of sensing. In classical sensing, the environment is unknown and the goal is to estimate its physical parameters as accurately as possible [17], [18]. In embodied communication, the environmental state is selected from an admissible set to represent information. The goal is therefore not only to estimate a state accurately, but to determine whether different selected states can be reliably told apart. The relevant performance object shifts from sensing accuracy to sensing distinguishability: how separable are the receiver observations induced by different environmental states?

This shift is the key bridge from sensing to communication. Traditional sensing metrics, such as estimation error or Cramer-Rao-type bounds [19], describe local accuracy around a physical state. Embodied communication requires a reliability metric between states, because symbols are confused only when the receiver cannot distinguish one induced state from another. As shown later, classical local sensing metrics can be interpreted as local approximations of this more global distinguishability geometry.

C. Contributions

This paper makes the following advances.

- We introduce embodied communication, a new communication paradigm in which information is carried by deliberately induced environmental states and recovered through receiver sensing. Unlike conventional wireless communication, the agent does not generate a dedicated waveform. Instead, information is written into the sensed world itself, enabling an information pathway that occupies no additional communication spectrum and requires no dedicated communication hardware.
- We instantiate embodied communication in an RF sensing-enabled cellular system and reveal its defining technical problem. The objective of sensing is no longer merely estimating an unknown environmental parameter, but distinguishing intentionally induced environmental states that serve as communication symbols. This shift from estimation accuracy to symbol distinguishability is the key technical departure from classical sensing.
- We develop a communication-theoretic characterization of this distinguishability. By deriving the pairwise Bhattacharyya distance between the sensing distributions induced by different environmental states, we obtain a sensing-induced reliability field over the controllable physical region. This field is the key geometric object of the embodied channel: it determines which physical states are confusable, which states are reliable, and how the BS sensing resolution shapes the embodied alphabet.
- We characterize the finite-snapshot ϵ -capacity of the embodied channel through achievable designs and converse bounds. The reliability field converts codebook design into a forbidden-region packing problem, leading to lattice-based achievable constructions and a closed-form hexagonal design under a main-lobe approximation. We also establish information-theoretic and geometric upper bounds, and show that the number of sensing snapshots creates

a fundamental tradeoff between sensing reliability and symbol duration.

II. SYSTEM MODEL AND EMBODIED CHANNEL

We consider a sensing-enabled cellular system consisting of one BS and one communication agent, as illustrated in Fig. 2. The BS is endowed with RF sensing capability, motivated by the integration of sensing functions into 5.5G and 6G infrastructures. Although embodied communication can in principle be realized over arbitrary sensing modalities, this paper focuses on RF sensing to develop a concrete communication-theoretic model.

A. RF Sensing Model at the BS

The BS uses a probing waveform to illuminate the environment and employs a colocated uniform planar array (UPA) to receive the scattered echo. As shown in Fig. 2, the UPA lies in the yz -plane and is centered at the origin. Its broadside direction is aligned with the positive x -axis. The UPA consists of $M_y \times M_z$ isotropic antenna elements with inter-element spacing $d = \lambda/2$, where λ is the carrier wavelength. The total number of antenna elements is denoted by $M = M_y M_z$.

We adopt a uniform-illumination sensing model: the probing signal is assumed to illuminate the entire agent-controllable region with constant gain. Thus, the transmit-side gain and the large-scale propagation loss are essentially constant across all admissible scatterer positions. These position-independent factors are absorbed into the average scattering power parameter introduced below. The position-dependent structure retained in the sensing observation is therefore the receive-array spatial signature, which determines the BS's ability to distinguish different induced environmental states. Small-scale scattering variations across sensing snapshots are captured by random complex scattering coefficients.

The communication agent is located in the far field of the BS array. It controls a planar physical region of size $a_y \times a_z$, where a_y and a_z denote the extents along the y - and z -axes, respectively. The region is located at distance D from the BS, with its center at $(D, 0, 0)$, and is parallel to the yz -plane. The controllable region is therefore

$$\mathcal{A} = \left[-\frac{a_y}{2}, \frac{a_y}{2} \right] \times \left[-\frac{a_z}{2}, \frac{a_z}{2} \right]. \quad (1)$$

Within this controllable region, the agent deliberately places a point scatterer at a selected location, thereby creating the simplest form of a controllable spatial pattern.¹ The resulting environmental state serves as the embodied symbol that carries the intended information. In other words, the agent communicates by writing a physical state into the environment, and the BS decodes the intended message by sensing that state.

Let $\mathbf{r} = (y, z) \in \mathcal{A}$ denote the position of the controllable point scatterer on the agent plane, where (y, z) is the offset from the center of the plane. Under the far-field and small-region assumptions, this physical coordinate is approximately mapped to the angular coordinates observed at the BS as

$$\theta \approx \frac{y}{D}, \quad \phi \approx \frac{z}{D}. \quad (2)$$

¹More generally, the agent may manipulate multiple point scatterers or extended reflective structures, leading to richer spatial patterns.

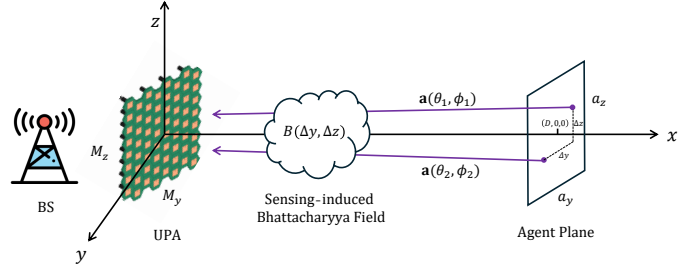


Figure 2: Embodied communication in a sensing-enabled cellular system. The agent communicates by writing a physical state into the environment, and the BS decodes the intended message by sensing the state.

Throughout the paper, the scatterer may therefore be equivalently described by its physical coordinate (y, z) on the agent plane or by its angular direction (θ, ϕ) at the BS.

One environmental sensing of the BS consists of L snapshots. During the ℓ th snapshot, where $\ell = 1, 2, \dots, L$, the BS emits a narrowband probing waveform

$$s_\ell(t) = \sqrt{E_t} \bar{s}_\ell(t), \quad (3)$$

where E_t is the transmit energy per probing snapshot and $\bar{s}_\ell(t)$ is a unit-energy baseband pulse with duration T_p , e.g., a rectangular pulse $\bar{s}(t) = \frac{1}{\sqrt{T_p}}$ for $0 \leq t \leq T_p$. The snapshots use separated time resources.

The probing illumination impinges on the controllable scatterer, and the scattered echo is received by the UPA. After down-conversion and matched filtering with $\bar{s}_\ell(t)$, the BS obtains the complex baseband sensing snapshot

$$\mathbf{y}_\ell = h_\ell \mathbf{a}(\theta, \phi) + \mathbf{n}_\ell, \quad \ell = 1, \dots, L, \quad (4)$$

where $\mathbf{y}_\ell \in \mathbb{C}^M$ and $\mathbf{a}(\theta, \phi) \in \mathbb{C}^M$ is the normalized receive steering vector corresponding to the direction of the scatterer. The receiver noise satisfies $\mathbf{n}_\ell \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$, where σ^2 denotes the post-matched-filter noise variance. The noise vectors $\{\mathbf{n}_\ell\}_{\ell=1}^L$ are independent across snapshots.

The effective scattering coefficient in the ℓ th snapshot is modeled as a circularly symmetric complex Gaussian random variable, $h_\ell \sim \mathcal{CN}(0, \rho^2)$, independently across snapshots. The parameter ρ^2 denotes the average received echo energy at the matched-filter output before receiver noise. It captures the transmit probing energy, the transmit illumination gain, the propagation and front-end gains, and a nominal radar cross section (RCS). For example, under a free-space radar model [18], one may write

$$\rho = \sqrt{E_t G_{\text{illum}}} \frac{\lambda \sqrt{\sigma_{\text{RCS}}}}{(4\pi)^{3/2} D^2}, \quad (5)$$

where G_{illum} denotes the transmit illumination gain over the agent plane, and σ_{RCS} is a nominal RCS value.

For the UPA, the antenna elements are indexed by $m_y = 0, \dots, M_y - 1$ and $m_z = 0, \dots, M_z - 1$. The normalized receive steering vector can be written as the Kronecker product of two steering vectors:

$$\mathbf{a}(\theta, \phi) = \mathbf{a}_y(\theta, \phi) \otimes \mathbf{a}_z(\phi), \quad (6)$$

where

$$\mathbf{a}_y(\theta, \phi) = \frac{1}{\sqrt{M_y}} \left[1, e^{j\pi \sin \theta \cos \phi}, \dots, e^{j\pi(M_y-1) \sin \theta \cos \phi} \right]^\top,$$

$$\mathbf{a}_z(\phi) = \frac{1}{\sqrt{M_z}} \left[1, e^{j\pi \sin \phi}, \dots, e^{j\pi(M_z-1) \sin \phi} \right]^\top.$$

The normalization ensures $\|\mathbf{a}(\theta, \phi)\| = 1$.

Using (2), we define the position-induced steering vector

$$\mathbf{a}(\mathbf{r}) \triangleq \mathbf{a}\left(\frac{y}{D}, \frac{z}{D}\right). \quad (7)$$

Then, stacking all L snapshots gives

$$\mathbf{Y} = \mathbf{a}(\mathbf{r})\mathbf{h}^\top + \mathbf{N}, \quad (8)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{C}^{M \times L}$, $\mathbf{h} = [h_1, \dots, h_L]^\top \in \mathbb{C}^L$, and $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_L] \in \mathbb{C}^{M \times L}$. The post-matched-filter per-snapshot sensing signal-to-noise ratio (SNR) of the controllable scatterer is defined as

$$\gamma_0 \triangleq \frac{\rho^2}{\sigma^2}. \quad (9)$$

Since $h_\ell \sim \mathcal{CN}(0, \rho^2)$, the conditional distribution of one sensing snapshot given the scatterer position \mathbf{r} is

$$\mathbf{y}_\ell | \mathbf{r} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}(\mathbf{r})), \quad (10)$$

where

$$\mathbf{R}(\mathbf{r}) = \rho^2 \mathbf{a}(\mathbf{r})\mathbf{a}^H(\mathbf{r}) + \sigma^2 \mathbf{I}_M = \sigma^2 (\mathbf{I}_M + \gamma_0 \mathbf{a}(\mathbf{r})\mathbf{a}^H(\mathbf{r})).$$

Thus, the scatterer position is encoded in the covariance structure of the received sensing snapshots.

In practice, the BS observes the sample covariance

$$\widehat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{L} \sum_{\ell=1}^L \mathbf{y}_\ell \mathbf{y}_\ell^H = \frac{1}{L} \mathbf{Y} \mathbf{Y}^H. \quad (11)$$

The number of snapshots L is the sensing resource that can be leveraged in one embodied channel use.

B. Ambient Modulation

The agent communicates by selecting one admissible scatterer position from a predefined position codebook. Let

$$\mathfrak{C} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_J\} \subseteq \mathcal{A}$$

denote the set of admissible positions, where $\mathbf{r}_j = (y_j, z_j)$, $j = 1, \dots, J$. The cardinality $J = |\mathfrak{C}|$ determines the size of the embodied symbol alphabet.

Let $W \in \{1, 2, \dots, J\}$ denote the message to be conveyed in one embodied channel use. Ambient modulation is defined as the mapping

$$W = j \quad \mapsto \quad \mathbf{r}_j \in \mathfrak{C}. \quad (12)$$

We assume ideal ambient modulation in this paper. That is, for any selected codeword $\mathbf{r}_j \in \mathfrak{C}$, the agent can realize the corresponding scatterer position without placement error, unintended scatterers, or temporal drift within the embodied channel use. Accordingly, the ambient modulation channel is deterministic: $\Pr(\mathbf{r} = \mathbf{r}_j | W = j) = 1$.

This assumption allows us to isolate the fundamental limitation imposed by the ambient sensing process, namely the

ability of the BS to distinguish different physical positions through noisy multi-snapshot RF observations. The impact of imperfect ambient modulation, including finite placement accuracy, actuation noise, and parasitic scattering, is left for future investigation.

C. Embodied Channel Law

Combining the position modulation model and the multi-snapshot RF sensing model yields the embodied communication channel. For a transmitted message $W = j$, the agent places the scatterer at position \mathbf{r}_j , and the BS observes $\mathbf{Y} = \mathbf{a}(\mathbf{r}_j)\mathbf{h}_j^\top + \mathbf{N}$. Unlike a coherent communication model, the instantaneous scattering vector \mathbf{h}_j is not known to the BS and is not treated as receiver-side CSI. It is a random nuisance component induced by the scattering process. Therefore, the channel transition law is obtained by marginalizing over the random scattering coefficients.

Since the columns of \mathbf{Y} are independent conditioned on $W = j$, the embodied channel transition law is

$$p(\mathbf{Y} | W = j) = \prod_{\ell=1}^L \frac{1}{\pi^M \det(\mathbf{R}_j)} \exp(-\mathbf{y}_\ell^H \mathbf{R}_j^{-1} \mathbf{y}_\ell), \quad (13)$$

where $\mathbf{R}_j \triangleq \mathbf{R}(\mathbf{r}_j) = \sigma^2 (\mathbf{I}_M + \gamma_0 \mathbf{a}(\mathbf{r}_j)\mathbf{a}^H(\mathbf{r}_j))$.

The information flow can be summarized as

$$W \rightarrow \mathbf{r}_W \rightarrow \mathbf{R}_W \rightarrow \mathbf{Y} \rightarrow \widehat{W}. \quad (14)$$

The first mapping represents position modulation by the agent, while the second represents RF sensing by the BS. Together, they define a discrete-input continuous-output noncoherent channel whose input symbols are physical positions rather than conventional transmitter-generated waveforms.

Remark 1 (Distinctive structure of the embodied channel). *At first sight, (13) resembles a standard discrete-input continuous-output channel. The distinction lies in the origin and constraint of the input alphabet. In a conventional communication channel, the input symbols are waveform codewords generated by a transmitter and constrained by resources such as power, bandwidth, and blocklength. In the embodied channel, the input symbols are physical states of the environment, namely scatterer positions on the agent plane: $\mathfrak{C} \subseteq \mathcal{A}$. Thus, the alphabet is not freely synthesized in signal space. It must be carved out of a finite physical region and can only contain positions that the BS sensing system can reliably distinguish.*

Mathematically, each physical position induces the cascade

$$\mathbf{r} \mapsto \mathbf{a}(\mathbf{r}) \mapsto \mathbf{R}(\mathbf{r}) \mapsto p(\mathbf{Y} | \mathbf{r}). \quad (15)$$

Choosing a codeword \mathbf{r}_j therefore simultaneously chooses a physical state, an array steering vector, and a covariance matrix of the sensing observation. Consequently, the input alphabet and the channel law are coupled through the sensing geometry. The central capacity question is not merely how much information can be transmitted over a given alphabet, but how many physical states can be selected from \mathcal{A} such that the induced distributions $\{p(\mathbf{Y} | \mathbf{r}_j)\}$ are reliably distinguishable. This coupling between physical alphabet design and

sensing-induced statistical distinguishability is the defining mathematical feature of the embodied channel.

In the next section, we define the ϵ -capacity of this embodied channel and characterize how the BS sensing process constrains the number of physical positions that can be reliably distinguished from L noisy snapshots.

III. THE ϵ -CAPACITY OF EMBODIED CHANNEL

In this section, we define the ϵ -capacity of the embodied channel introduced. The definition follows a finite-blocklength reliability [20] viewpoint: for a fixed number of sensing snapshots L and a target error probability ϵ , we ask *how many physical scatterer positions can be reliably used as embodied symbols*. This differs from the Shannon capacity of a fixed input alphabet. Here, the input alphabet itself is a design object, since the codewords are physical positions selected from the agent-controllable region \mathcal{A} .

A. The ϵ -Capacity

Consider an embodied codebook $\mathcal{C} = \{\mathbf{r}_1, \dots, \mathbf{r}_J\} \subseteq \mathcal{A}$, where message $W = j$ is mapped to position \mathbf{r}_j . Given the observation \mathbf{Y} , the BS applies a decoder

$$g: \mathbb{C}^{M \times L} \rightarrow \{1, \dots, J\},$$

and outputs $\widehat{W} = g(\mathbf{Y})$.

For a fixed codebook \mathcal{C} and decoder g , we define the maximum decoding error probability as

$$P_{\max}(\mathcal{C}, g) = \max_{j \in \{1, \dots, J\}} \Pr(g(\mathbf{Y}) \neq j \mid W = j).$$

The optimal maximum error probability of the codebook is

$$P_{\max}^*(\mathcal{C}) = \inf_g P_{\max}(\mathcal{C}, g), \quad (16)$$

where the infimum is over all measurable decoders.

For a target error probability $\epsilon \in (0, 1)$ and a fixed number of snapshots L , define the maximum ϵ -reliable embodied alphabet size as

$$\begin{aligned} J_\epsilon^*(L) &= \max_{\mathcal{C} \subseteq \mathcal{A}} |\mathcal{C}| \\ \text{s.t. } P_{\max}^*(\mathcal{C}) &\leq \epsilon. \end{aligned}$$

The ϵ -capacity of the embodied channel is then defined as

$$C_\epsilon(L) = \frac{1}{LT_p} \log_2 J_\epsilon^*(L). \quad (17)$$

The ϵ -capacity in (17) is the central object of this paper. It measures the largest number of physical states that can be carved out of the agent plane and reliably distinguished by the BS sensing system. To characterize this quantity, we first study the most elementary problem: distinguishing two candidate scatterer positions.

B. Pairwise Distinguishability

Consider two candidate positions $\mathbf{r}_i = (y_i, z_i)$, $\mathbf{r}_j = (y_j, z_j)$ with corresponding steering vectors $\mathbf{a}_i = \mathbf{a}(\mathbf{r}_i)$ and $\mathbf{a}_j = \mathbf{a}(\mathbf{r}_j)$. The BS must distinguish between the two hypotheses:

$$\mathcal{H}_i: \mathbf{y}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_i),$$

$$\mathcal{H}_j: \mathbf{y}_\ell \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_j).$$

For a single snapshot, the optimal likelihood-ratio test compares the likelihoods induced by \mathbf{R}_i and \mathbf{R}_j . For L independent snapshots, the log-likelihood ratio is the sum of the corresponding single-snapshot log-likelihood ratios. Although the exact pairwise error probability of this covariance discrimination problem generally does not admit a simple closed form, it can be tightly characterized by classical error exponents. In particular, we use the Bhattacharyya bound [21], which gives an analytically tractable upper bound on the pairwise error probability.

For two zero-mean complex Gaussian distributions with covariance matrices \mathbf{R}_i and \mathbf{R}_j , the single-snapshot Bhattacharyya distance is

$$B_{ij} = \log \frac{\det\left(\frac{\mathbf{R}_i + \mathbf{R}_j}{2}\right)}{\sqrt{\det(\mathbf{R}_i) \det(\mathbf{R}_j)}}. \quad (18)$$

Since the L snapshots are conditionally independent, the corresponding pairwise error probability satisfies

$$P_{i \rightarrow j} \leq \exp(-LB_{ij}). \quad (19)$$

We now evaluate B_{ij} for the embodied sensing model.

Theorem 1 (Pairwise embodied sensing exponent). *For two scatterer positions \mathbf{r}_i and \mathbf{r}_j , define the squared steering-vector correlation $\eta_{ij} \triangleq |\mathbf{a}_i^H \mathbf{a}_j|^2$. Then the Bhattacharyya distance between the two induced sensing distributions is*

$$B_{ij} = \log \frac{(1 + \frac{\gamma_0}{2})^2 - \frac{\gamma_0^2}{4} \eta_{ij}}{1 + \gamma_0}. \quad (20)$$

Consequently, the pairwise error probability from L snapshots satisfies

$$P_{i \rightarrow j} \leq \left[\frac{1 + \gamma_0}{(1 + \frac{\gamma_0}{2})^2 - \frac{\gamma_0^2}{4} \eta_{ij}} \right]^L. \quad (21)$$

Proof. Since $\|\mathbf{a}_i\| = \|\mathbf{a}_j\| = 1$, the matrix determinant lemma gives $\det(\mathbf{R}_i) = \det(\mathbf{R}_j) = \sigma^{2M}(1 + \gamma_0)$. Thus, the position-dependent part of (18) is contained in $\det\left(\frac{\mathbf{R}_i + \mathbf{R}_j}{2}\right)$, where

$$\frac{\mathbf{R}_i + \mathbf{R}_j}{2} = \sigma^2 \left[\mathbf{I}_M + \frac{\gamma_0}{2} (\mathbf{a}_i \mathbf{a}_i^H + \mathbf{a}_j \mathbf{a}_j^H) \right].$$

Defining $\mathbf{U}_{ij} = [\mathbf{a}_i, \mathbf{a}_j] \in \mathbb{C}^{M \times 2}$, we have

$$\det\left(\frac{\mathbf{R}_i + \mathbf{R}_j}{2}\right) = \sigma^{2M} \det\left[\mathbf{I}_M + \frac{\gamma_0}{2} \mathbf{U}_{ij} \mathbf{U}_{ij}^H\right].$$

By Sylvester's determinant identity,

$$\det\left[\mathbf{I}_M + \frac{\gamma_0}{2} \mathbf{U}_{ij} \mathbf{U}_{ij}^H\right] = \det\left[\mathbf{I}_2 + \frac{\gamma_0}{2} \mathbf{U}_{ij}^H \mathbf{U}_{ij}\right].$$

Moreover,

$$\mathbf{U}_{ij}^H \mathbf{U}_{ij} = \begin{bmatrix} 1 & \mathbf{a}_i^H \mathbf{a}_j \\ \mathbf{a}_j^H \mathbf{a}_i & 1 \end{bmatrix}.$$

Hence,

$$\det\left[\mathbf{I}_2 + \frac{\gamma_0}{2} \mathbf{U}_{ij}^H \mathbf{U}_{ij}\right] = \det\left[\begin{array}{cc} 1 + \frac{\gamma_0}{2} & \frac{\gamma_0}{2} \mathbf{a}_i^H \mathbf{a}_j \\ \frac{\gamma_0}{2} \mathbf{a}_j^H \mathbf{a}_i & 1 + \frac{\gamma_0}{2} \end{array}\right]$$

$$= \left(1 + \frac{\gamma_0}{2}\right)^2 - \frac{\gamma_0^2}{4}\eta_{ij}.$$

Therefore,

$$\det\left(\frac{\mathbf{R}_i + \mathbf{R}_j}{2}\right) = \sigma^{2M} \left[\left(1 + \frac{\gamma_0}{2}\right)^2 - \frac{\gamma_0^2}{4}\eta_{ij} \right].$$

Substituting the above determinant expressions into (18) yields (20). The pairwise error bound (21) follows from (19). \square

Theorem 1 shows that the two positions enter the pairwise error exponent only through the squared steering-vector correlation η_{ij} . A larger η_{ij} means that the two positions induce more similar BS array responses, which reduces B_{ij} and increases the pairwise error probability. A smaller η_{ij} means that the two positions are more separated in the BS sensing space, which increases B_{ij} and improves discrimination reliability.

Remark 2 (From sensing accuracy to sensing distinguishability). *The role of sensing in embodied communication is different from that in conventional localization. Classical sensing typically asks how accurately an unknown scatterer position can be estimated. Here, the scatterer position is intentionally selected from a finite physical codebook and therefore acts as a communication symbol. The relevant question is thus how reliably two physical states can be distinguished.*

This is why we characterize the BS sensing process through the binary hypothesis testing distance B_{ij} rather than through an estimation-accuracy metric. The resulting pairwise distinguishability directly controls the probability of confusing one embodied symbol with another, and later becomes the geometry used to design reliable position codebooks.

C. Sensing-Induced Reliability Field

Theorem 1 expresses pairwise reliability in terms of η_{ij} . We now connect η_{ij} to physical displacement on the agent plane.

Let $\Delta_{ij} = (\Delta y_{ij}, \Delta z_{ij}) \triangleq \mathbf{r}_i - \mathbf{r}_j$. Under the far-field and small-region approximation in (2), the steering-vector correlation depends only on Δ_{ij} .

Lemma 2 (Steering correlation induced by physical displacement). *Under the UPA model, the squared correlation between two steering vectors associated with positions separated by $\Delta = (\Delta y, \Delta z)$ is*

$$\eta(\Delta) = \eta_y(\Delta y)\eta_z(\Delta z), \quad (22)$$

where

$$\eta_y(\Delta y) = \frac{1}{M_y^2} \left| \frac{\sin\left(\frac{\pi M_y \Delta y}{2D}\right)}{\sin\left(\frac{\pi \Delta y}{2D}\right)} \right|^2, \quad \eta_z(\Delta z) = \frac{1}{M_z^2} \left| \frac{\sin\left(\frac{\pi M_z \Delta z}{2D}\right)}{\sin\left(\frac{\pi \Delta z}{2D}\right)} \right|^2.$$

Proof. From the Kronecker structure in (6), we can write $\mathbf{a}(\mathbf{r}) = \mathbf{a}_y(\mathbf{r}) \otimes \mathbf{a}_z(\mathbf{r})$. Therefore, for two positions \mathbf{r}_i and \mathbf{r}_j , we have

$$\begin{aligned} |\mathbf{a}_i^H \mathbf{a}_j|^2 &= |(\mathbf{a}_{y,i}^H \otimes \mathbf{a}_{z,i}^H)(\mathbf{a}_{y,j} \otimes \mathbf{a}_{z,j})|^2 \\ &= |\mathbf{a}_{y,i}^H \mathbf{a}_{y,j}|^2 |\mathbf{a}_{z,i}^H \mathbf{a}_{z,j}|^2. \end{aligned}$$

Using the small-angle approximation, we have $\sin \theta \cos \phi \simeq \frac{y}{D}$ and $\sin \phi \simeq \frac{z}{D}$. The y -axis inner product can be written as

$$\mathbf{a}_{y,i}^H \mathbf{a}_{y,j} = \frac{1}{M_y} \sum_{m=0}^{M_y-1} \exp\left(j\pi m \frac{y_j - y_i}{D}\right).$$

Taking the magnitude squared and using the finite geometric-sum identity gives

$$|\mathbf{a}_{y,i}^H \mathbf{a}_{y,j}|^2 = \frac{1}{M_y^2} \left| \frac{\sin\left(\frac{\pi M_y (y_i - y_j)}{2D}\right)}{\sin\left(\frac{\pi (y_i - y_j)}{2D}\right)} \right|^2.$$

The derivation of $\eta_z(\Delta z)$ follows identically from

$$\mathbf{a}_{z,i}^H \mathbf{a}_{z,j} = \frac{1}{M_z} \sum_{n=0}^{M_z-1} \exp\left(j\pi n \frac{z_j - z_i}{D}\right).$$

Multiplying the two squared magnitudes yields (22). \square

Combining Theorem 1 and Lemma 2, we can define the sensing-induced Bhattacharyya field:

$$B(\Delta) = \log \frac{\left(1 + \frac{\gamma_0}{2}\right)^2 - \frac{\gamma_0^2}{4}\eta(\Delta)}{1 + \gamma_0}. \quad (23)$$

Then, the pairwise error probability satisfies

$$P_{i \rightarrow j} \leq \exp[-LB(\Delta_{ij})]. \quad (24)$$

The field $B(\Delta)$ is the fundamental reliability object induced by the BS sensing system on the agent plane. It maps a physical displacement between two possible embodied symbols to a sensing error exponent. It is generally anisotropic and not solely a function of the Euclidean distance $\|\Delta\|_2$.

For a target pairwise error probability ϵ_p , define the required single-snapshot exponent

$$B_{\text{req}}(\epsilon_p, L) = \frac{1}{L} \log \frac{1}{\epsilon_p}. \quad (25)$$

The pairwise confusability region is

$$\mathcal{S}_{\text{conf}}(\epsilon_p, L) = \{\Delta \in \mathbb{R}^2 : B(\Delta) < B_{\text{req}}(\epsilon_p, L)\}.$$

If the displacement between two codewords lies outside $\mathcal{S}_{\text{conf}}(\epsilon_p, L)$, then the Bhattacharyya bound guarantees pairwise error probability no larger than ϵ_p .

Remark 3 (Relation to classical sensing metrics). *Classical studies often characterize localization performance through estimation-theoretic metrics such as the Fisher information matrix (FIM) and the Cramer-Rao bound (CRB).*

The Bhattacharyya field $B(\Delta)$ provides a pairwise error exponent between two candidate embodied symbols and leads directly to codebook-level reliability conditions. In contrast, the CRB provides a local lower bound on estimation variance and does not directly yield a maximum-error guarantee for a finite position codebook.

The two viewpoints are nevertheless connected. Let $\theta = [y, z]^T$ denote the scatterer position, and let $\mathbf{J}(\theta)$ be the single-snapshot FIM associated with the Gaussian model $\mathcal{CN}(\mathbf{0}, \mathbf{R}(\theta))$. For two nearby positions θ and $\theta + \delta$, the

Bhattacharyya distance admits the local expansion

$$B(\boldsymbol{\theta}, \boldsymbol{\theta} + \boldsymbol{\delta}) = \frac{1}{8} \boldsymbol{\delta}^T \mathbf{J}(\boldsymbol{\theta}) \boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|^2). \quad (26)$$

Thus, the FIM underlying the CRB describes the local curvature of the pairwise distinguishability field. However, embodied communication requires a global codebook design over finite separations, for which the full function $B(\boldsymbol{\Delta})$ is needed. In particular, $B(\boldsymbol{\Delta})$ captures anisotropy, sidelobes, and nonlocal ambiguities of the BS array response that are not represented by a purely local CRB.

D. Reliability of a Codebook

We now lift the pairwise reliability law to a codebook-level guarantee under maximum error.

For a codebook \mathfrak{C} , the maximum-likelihood decoder is

$$\widehat{W} = \arg \max_{j \in \{1, \dots, J\}} p(\mathbf{Y} | W = j). \quad (27)$$

Using (13), the negative log-likelihood under hypothesis $W = j$ is, up to constants independent of j , $L \log \det(\mathbf{R}_j) + \text{tr}(\mathbf{R}_j^{-1} \mathbf{Y} \mathbf{Y}^H)$. Therefore,

$$\widehat{W} = \arg \min_j [L \log \det(\mathbf{R}_j) + \text{tr}(\mathbf{R}_j^{-1} \mathbf{Y} \mathbf{Y}^H)],$$

where $\det(\mathbf{R}_j) = \sigma^{2M}(1 + \gamma_0)$, $\forall j$, which is common to all hypotheses. Substituting

$$\mathbf{R}_j^{-1} = \frac{1}{\sigma^2} \left(\mathbf{I}_M - \frac{\gamma_0}{1 + \gamma_0} \mathbf{a}_j \mathbf{a}_j^H \right)$$

and removing terms independent of j , we obtain

$$\widehat{W} = \arg \max_j \mathbf{a}_j^H \mathbf{Y} \mathbf{Y}^H \mathbf{a}_j. \quad (28)$$

Thus, in the embodied channel, the optimal decoder selects the candidate position whose steering vector captures the largest accumulated sensing energy over the L snapshots.

The reliability of the entire codebook is controlled by the weakest pair of codewords. Define the minimum pairwise Bhattacharyya distance of \mathfrak{C} as

$$B_{\min}(\mathfrak{C}) = \min_{i \neq j} B(\mathbf{r}_i - \mathbf{r}_j).$$

Theorem 3 (Sufficient condition for ϵ -reliability). *For a codebook \mathfrak{C} with J codewords, the maximum error probability under ML decoding satisfies*

$$P_{\max}(\mathfrak{C}, g_{\text{ML}}) \leq (J - 1) \exp[-LB_{\min}(\mathfrak{C})]. \quad (29)$$

Consequently, a sufficient condition for ϵ -reliability is

$$B_{\min}(\mathfrak{C}) \geq \frac{1}{L} \log \frac{J - 1}{\epsilon}. \quad (30)$$

Proof. The proof follows the union bound over the $J - 1$ competing codewords, together with the pairwise bound in (24), and is omitted for brevity. \square

Theorem 3 gives a direct codebook design rule. A J -position codebook is guaranteed to be ϵ -reliable if every pairwise displacement lies outside the confusability region associated with the pairwise budget $\epsilon/(J - 1)$. This also

indicates that a larger codebook requires a larger minimum pairwise error exponent to maintain the same overall reliability.

IV. ACHIEVABLE BOUNDS AND CODEBOOK DESIGNS

Section III defines the ϵ -capacity $C_\epsilon(L)$ and establishes a sufficient reliability condition for a codebook. Specifically, a J -position codebook $\mathfrak{C} \subseteq \mathcal{A}$ is guaranteed to be ϵ -reliable if (30) is satisfied. In this section, we develop achievable lower bounds on $C_\epsilon(L)$ by constructing codebooks that satisfy this condition.

A. Forbidden-Region Packing Lower Bound

The key observation is that the reliability condition defines a forbidden region in displacement space. Two physical positions can be used as reliable codewords only if their displacement does not fall into this region. Therefore, embodied codebook design becomes a packing problem on the agent plane under a sensing-induced forbidden displacement geometry.

For a candidate codebook size J , define the codebook-level Bhattacharyya threshold

$$B_J \triangleq \frac{1}{L} \log \frac{J - 1}{\epsilon}.$$

The corresponding forbidden displacement region is

$$\mathcal{F}_J = \{\boldsymbol{\Delta} \in \mathbb{R}^2 : B(\boldsymbol{\Delta}) < B_J\}. \quad (31)$$

Therefore, a sufficient condition for ϵ -reliability is $\mathbf{r}_i - \mathbf{r}_j \notin \mathcal{F}_J$, $\forall i \neq j$. This leads to the following alphabet size:

$$J_{\text{F}}^*(L, \epsilon) = \max\{J : \exists \mathfrak{C} \subseteq \mathcal{A}, |\mathfrak{C}| = J, \mathbf{r}_i - \mathbf{r}_j \notin \mathcal{F}_J, \forall i \neq j\}. \quad (32)$$

Every codebook counted by (32) satisfies the sufficient reliability condition in Theorem 3. Hence, the forbidden-region packing rate

$$R_{\text{F}}(L, \epsilon) = \frac{1}{LT_p} \log_2 J_{\text{F}}^*(L, \epsilon) \quad (33)$$

is achievable, and $C_\epsilon(L) \geq R_{\text{F}}(L, \epsilon)$.

The optimization in (32), however, is difficult because it allows arbitrary finite point sets in a continuous region. We next introduce lattice codebooks as a structured and analytically tractable class of embodied codebooks.

1) *Lattice codebooks:* A natural structured construction is to select codewords from a two-dimensional lattice on the agent plane. Let

$$\Lambda(\mathbf{G}) = \{\mathbf{G}\mathbf{k} : \mathbf{k} \in \mathbb{Z}^2\},$$

where $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2] \in \mathbb{R}^{2 \times 2}$ is a full-rank lattice generator [22]. The corresponding lattice codebook is

$$\mathfrak{C}_\Lambda(\mathbf{G}) = \Lambda(\mathbf{G}) \cap \mathcal{A},$$

and we denote by $J_\Lambda(\mathbf{G}) \triangleq |\mathfrak{C}_\Lambda(\mathbf{G})|$ the number of lattice points retained inside the agent plane.

Proposition 4 (Lattice-codebook achievable rate). *For any full-rank lattice generator \mathbf{G} , if*

$$\mathbf{G}\mathbf{k} \notin \mathcal{F}_{J_\Lambda(\mathbf{G})}, \quad \forall \mathbf{k} \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}, \quad (34)$$

the lattice codebook $\mathfrak{C}_\Lambda(\mathbf{G})$ is ϵ -reliable under ML decoding. The corresponding achievable embodied rate is

$$R_\Lambda(\mathbf{G}) = \frac{1}{LT_p} \log_2 J_\Lambda(\mathbf{G}). \quad (35)$$

Proof. For a lattice codebook, all pairwise displacements belong to the difference lattice: $\mathbf{r}_i - \mathbf{r}_j \in \Lambda(\mathbf{G})$, $i \neq j$.

If (34) holds, then for all distinct codewords $\mathbf{r}_i, \mathbf{r}_j \in \mathfrak{C}_\Lambda(\mathbf{G})$, $\mathbf{r}_i - \mathbf{r}_j \notin \mathcal{F}_{J_\Lambda(\mathbf{G})}$. By the definition of \mathcal{F}_J in (31), this implies

$$B(\mathbf{r}_i - \mathbf{r}_j) \geq \frac{1}{L} \log \frac{J_\Lambda(\mathbf{G}) - 1}{\epsilon}, \quad \forall i \neq j. \quad (36)$$

Therefore, Theorem 3 gives $P_{\max}(\mathfrak{C}_\Lambda(\mathbf{G}), g_{\text{ML}}) \leq \epsilon$. The lattice codebook is feasible for the ϵ -capacity problem and achieves (35). \square

The best lattice achievable rate is

$$R_\Lambda^*(L, \epsilon) = \max_{\mathbf{G} \in \mathbb{R}^{2 \times 2}} \frac{1}{LT_p} \log_2 J_\Lambda(\mathbf{G}), \quad \text{s.t. (34)}.$$

Then, $R_\Lambda^*(L, \epsilon) \leq C_\epsilon(L)$ is a lower bound.

Remark 4. The lattice formulation contains rectangular grids as a special case by choosing $\mathbf{G} = [[s_y, 0]^\top, [0, s_z]^\top]$. However, a general lattice is more flexible. It can rotate, shear, and adapt its fundamental cell to the anisotropic shape of the forbidden region. This additional geometric freedom is essential when $B(\Delta)$ is not isotropic.

For large agent planes, boundary effects are small and the number of retained lattice points is approximately

$$J_\Lambda(\mathbf{G}) \approx \frac{a_y a_z}{|\det \mathbf{G}|}. \quad (37)$$

Thus, lattice codebook design can be interpreted as minimizing the fundamental cell area $|\det \mathbf{G}|$ while ensuring that no nonzero lattice point falls into the forbidden region. The next subsection shows that, under a main-lobe approximation, this lattice packing problem has a closed-form geometric solution.

2) *Main-lobe closed-form design:* The exact forbidden region \mathcal{F}_J is induced by the UPA Dirichlet kernels in (22). It can be anisotropic and may contain sidelobe-induced nonconvex structure. However, in many practical regimes, the dominant part of the forbidden region is governed by the main lobe around the origin. We now use a second-order approximation to obtain a closed-form lattice design.

Lemma 5 (Quadratic approximation of the reliability field). *In the main-lobe regime, the sensing-induced Bhattacharyya field admits the approximation*

$$B(\Delta y, \Delta z) \approx \kappa (\alpha_y \Delta y^2 + \alpha_z \Delta z^2), \quad (38)$$

where $\kappa = \frac{\gamma_0^2}{4(1+\gamma_0)}$, and

$$\alpha_y = \frac{\pi^2(M_y^2 - 1)}{12D^2}, \quad \alpha_z = \frac{\pi^2(M_z^2 - 1)}{12D^2}.$$

Equivalently, we can write

$$B(\Delta) \approx \Delta^T \mathbf{G}_B \Delta, \quad (39)$$

where

$$\mathbf{G}_B = \kappa \begin{bmatrix} \alpha_y & 0 \\ 0 & \alpha_z \end{bmatrix}.$$

Proof. In the main-lobe regime, the relevant pairwise displacements are small enough for a second-order approximation of the steering correlation.

For small x ,

$$\frac{1}{M^2} \left| \frac{\sin(Mx/2)}{\sin(x/2)} \right|^2 \approx 1 - \frac{M^2 - 1}{12} x^2.$$

Using $x_y = \pi \Delta y / D$ and $x_z = \pi \Delta z / D$, we obtain

$$\eta_y(\Delta y) \approx 1 - \frac{\pi^2(M_y^2 - 1)}{12D^2} \Delta y^2 = 1 - \alpha_y \Delta y^2,$$

$$\eta_z(\Delta z) \approx 1 - \frac{\pi^2(M_z^2 - 1)}{12D^2} \Delta z^2 = 1 - \alpha_z \Delta z^2.$$

Therefore,

$$\begin{aligned} \eta(\Delta y, \Delta z) &= \eta_y(\Delta y) \eta_z(\Delta z) \\ &\approx (1 - \alpha_y \Delta y^2)(1 - \alpha_z \Delta z^2) \\ &\approx 1 - \alpha_y \Delta y^2 - \alpha_z \Delta z^2, \end{aligned}$$

where the fourth-order term has been neglected.

Substituting this into (23), we get

$$\begin{aligned} B(\Delta y, \Delta z) &= \log \left[1 + \frac{\gamma_0^2}{4(1+\gamma_0)} (1 - \eta(\Delta y, \Delta z)) \right] \\ &\approx \log \left[1 + \kappa (\alpha_y \Delta y^2 + \alpha_z \Delta z^2) \right]. \end{aligned}$$

In the main-lobe regime, the argument of the logarithm is close to one, so $\log(1+x) \approx x$, yielding (38). \square

Lemma 5 implies that the level sets of $B(\Delta)$ are approximately ellipses. Accordingly, the forbidden region in (31) is approximated by

$$\mathcal{F}_J^{\text{ml}} = \{ \Delta \in \mathbb{R}^2 : \Delta^T \mathbf{G}_B \Delta < B_J \}. \quad (40)$$

Its semi-axes are

$$r_y(J) = \sqrt{\frac{B_J}{\kappa \alpha_y}}, \quad r_z(J) = \sqrt{\frac{B_J}{\kappa \alpha_z}}.$$

The ellipse can be converted into a disk through the linear transformation

$$\tilde{\mathbf{r}} = \mathbf{T} \mathbf{r}, \quad \mathbf{T} = \mathbf{G}_B^{1/2} = \begin{bmatrix} \sqrt{\kappa \alpha_y} & 0 \\ 0 & \sqrt{\kappa \alpha_z} \end{bmatrix}. \quad (41)$$

For a displacement vector $\Delta = \mathbf{r}_i - \mathbf{r}_j$, its transformed displacement is $\tilde{\Delta} \triangleq \mathbf{T} \Delta = \tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_j$. Then $\Delta^T \mathbf{G}_B \Delta = \|\tilde{\Delta}\|_2^2$, and the forbidden ellipse becomes the disk

$$\tilde{\mathcal{F}}_J^{\text{ml}} = \{ \tilde{\Delta} : \|\tilde{\Delta}\|_2^2 < B_J \}. \quad (42)$$

Theorem 6 (Main-lobe hexagonal lattice design). *Under the main-lobe approximation in Lemma 5, the densest lattice codebook is obtained by placing a hexagonal lattice with minimum distance $\sqrt{B_J}$ in the transformed plane $\tilde{\mathbf{r}} = \mathbf{T} \mathbf{r}$ and mapping it back to the original agent plane by \mathbf{T}^{-1} . Ignoring boundary effects, the resulting achievable alphabet*

size is approximately

$$J_{\text{hex}} \approx \frac{2\Xi L}{\sqrt{3} \log \frac{J-1}{\epsilon}}, \quad (43)$$

where

$$\Xi = \frac{\pi^2 a_y a_z \gamma_0^2 \sqrt{(M_y^2 - 1)(M_z^2 - 1)}}{48D^2(1 + \gamma_0)}.$$

For large J and small ϵ ,

$$J_{\text{hex}} \approx \left\lceil \frac{\Xi_h L}{W_0\left(\frac{\Xi_h L}{\epsilon}\right)} \right\rceil, \quad (44)$$

where $\Xi_h = \frac{2\Xi}{\sqrt{3}}$ and W_0 is the principal branch of the Lambert- W function.

Proof. In the transformed plane, the forbidden region is a disk of radius $\sqrt{B_J}$. Thus, a feasible lattice must have minimum Euclidean distance at least $\sqrt{B_J}$. Among two-dimensional lattices with a given minimum distance, the hexagonal lattice has the smallest fundamental cell area. For minimum distance $\sqrt{B_J}$, this cell area is $A_{\text{hex}}(B_J) = \frac{\sqrt{3}}{2} B_J$.

The transformed agent plane has area

$$|\tilde{\mathcal{A}}| = |\det \mathbf{T}| |\mathcal{A}| = a_y a_z \sqrt{\det \mathbf{G}_B} \triangleq \Xi.$$

Ignoring boundary effects, the number of retained lattice points is therefore

$$J \approx \frac{|\tilde{\mathcal{A}}|}{A_{\text{hex}}(B_J)} = \frac{2\Xi}{\sqrt{3} B_J},$$

where $B_J = \frac{1}{L} \log \frac{J-1}{\epsilon}$.

This expression is informative but implicit because J appears on both sides. To obtain a conservative explicit expression, we use

$$\log \frac{J-1}{\epsilon} \approx \log \frac{J}{\epsilon},$$

for large J and small ϵ . Then

$$J \approx \frac{2\Xi L}{\sqrt{3} \log \frac{J}{\epsilon}}.$$

Solving the equality $J \log \frac{J}{\epsilon} = \frac{2\Xi L}{\sqrt{3}}$ gives a closed-form expression through the Lambert- W function. Indeed, let $u = \log \frac{J}{\epsilon}$. Then $J = \epsilon e^u$, and the equality becomes $\epsilon u e^u = \frac{2\Xi L}{\sqrt{3}}$. Hence, we have

$$u = W_0\left(\frac{2\Xi L}{\sqrt{3}\epsilon}\right), \quad J = \frac{2\Xi L}{\sqrt{3} W_0\left(\frac{2\Xi L}{\sqrt{3}\epsilon}\right)}.$$

Eq. (44) follows by defining $\Xi_h = \frac{2\Xi}{\sqrt{3}}$. \square

Consequently, the achievable rate is

$$R_{\text{hex}}(L, \epsilon) = \frac{1}{LT_p} \log_2 J_{\text{hex}}.$$

Theorem 6 gives a closed-form geometric design rule under the main-lobe approximation. The BS sensing system induces a reliability metric \mathbf{G}_B on the agent plane. After whitening this metric, the approximately optimal lattice is hexagonal. Mapping it back to the original agent plane yields an inverse-

scaled hexagonal lattice that is denser along directions where the BS has finer sensing resolution.

Remark 5. It is worth noting that the result is approximate as it ignores sidelobes and boundary effects. For finite systems, the resulting lattice should be truncated to \mathcal{A} and its exact reliability should be verified using the true field $B(\Delta)$.

B. Snapshot-Number Optimization

The number of sensing snapshots L is a distinctive resource in embodied communication. Increasing L improves reliability because pairwise errors decay as $\exp[-LB(\Delta)]$. At the same time, one embodied channel use occupies LT_p seconds, reducing the rate normalization factor. Thus, L creates a sensing-duration tradeoff.

The explicit lower bound in (44) makes this tradeoff transparent. Ignoring the floor operation, the rate is

$$R_{\text{hex}}(L, \epsilon) \approx \frac{1}{LT_p} \log_2 \left[\frac{\Xi_h L}{W_0\left(\frac{\Xi_h L}{\epsilon}\right)} \right]. \quad (45)$$

The term inside the logarithm increases with L , because more snapshots allow more reliable position discrimination and hence a denser physical alphabet. However, the prefactor $1/(LT_p)$ decreases with L . The balance between these two effects yields a finite optimal sensing duration.

Theorem 7 (Optimal snapshot number). *Under the main-lobe hexagonal lattice approximation, the continuous relaxation of $\max_{L>0} R_{\text{hex}}(L, \epsilon)$ has the stationary point*

$$L_{\text{cont}}^* \approx \frac{\epsilon}{\Xi_h} y^* e^{y^*}, \quad (46)$$

where $y^* = \frac{q + \sqrt{q^2 + 4q}}{2}$ and $q = -\log \epsilon$. For integer snapshot numbers, one may evaluate $L = \lfloor L_{\text{cont}}^* \rfloor$ and $L = \lceil L_{\text{cont}}^* \rceil$ using the exact reliability condition based on the true field $B(\Delta)$.

Proof. Let $x = \frac{\Xi_h L}{\epsilon}$ and $y = W_0(x)$. Then $x = ye^y$, and hence $L = \frac{\epsilon}{\Xi_h} ye^y$. Therefore,

$$J_{\text{hex}} \approx \frac{\Xi_h L}{W_0(\Xi_h L/\epsilon)} = \epsilon e^y.$$

Up to the constant factor $1/(T_p \log 2)$, maximizing (45) is equivalent to maximizing

$$g(y) = \frac{\log J_{\text{hex}}}{L} = \frac{\log \epsilon + y}{(\epsilon/\Xi_h) ye^y}.$$

The constant ϵ/Ξ_h does not affect the optimizer. Let $q = -\log \epsilon$. Then the objective is proportional to

$$\tilde{g}(y) = \frac{y - q}{ye^y}.$$

The feasible rate requires $y > q$. Taking the derivative gives

$$\frac{d}{dy} \left[\frac{y - q}{ye^y} \right] = e^{-y} \left[\frac{q}{y^2} - \frac{y - q}{y} \right].$$

Setting the derivative to zero yields $\frac{q}{y^2} = \frac{y - q}{y}$. The positive root is

$$y^* = \frac{q + \sqrt{q^2 + 4q}}{2}.$$

Substituting $L = (\epsilon/\Xi_h)ye^y$ yields (46). \square

Theorem 7 shows that increasing the sensing duration is not always beneficial. The result also reveals the role of system parameters. A larger transformed-area parameter Ξ_h reduces the required optimal snapshot number because the BS-agent geometry already supports a dense physical alphabet. In contrast, a smaller Ξ_h , caused for example by larger distance D , smaller sensing SNR γ_0 , or smaller array aperture, requires more snapshots to achieve reliable discrimination.

V. UPPER BOUNDS ON ϵ -CAPACITY

Section IV constructs reliable embodied codebooks and thereby establishes achievable lower bounds on $C_\epsilon(L)$. We now derive upper bounds that limit how large the embodied alphabet can be. We present two complementary converses. The first is an information-theoretic converse based on Fano's inequality. It upper-bounds the number of reliably distinguishable physical positions by the amount of statistical information that the BS sensing observations can carry about the message. The second is a geometric packing converse. It translates binary testing impossibility into a necessary physical separation between codewords, and then upper-bounds how many such separated positions can fit inside the agent plane.

A. Information-Theoretic Converse

$C_\epsilon(L)$ is defined through the largest number of messages that can be decoded with maximum error probability no larger than ϵ . Fano's inequality applies to any such reliable decoding problem with a discrete message, an observation, and a decoder [23]. Therefore, it can be used to upper-bound the maximum feasible embodied alphabet size.

We first define the sensing covariance set induced by the agent plane:

$$\mathcal{Q}_A = \text{conv} \{ \mathbf{a}(\mathbf{r})\mathbf{a}^H(\mathbf{r}) : \mathbf{r} \in \mathcal{A} \}.$$

Each element of \mathcal{Q}_A is a convex mixture of rank-one spatial covariance matrices generated by feasible physical positions. This set captures the fact that the embodied alphabet cannot excite arbitrary array covariance directions; it is constrained by the finite physical support \mathcal{A} .

Define the support-constrained single-snapshot information upper bound

$$C_{\text{snap}}^{\text{ub}}(\mathcal{A}) = \sup_{\mathbf{Q} \in \mathcal{Q}_A} \log_2 \frac{\det(\mathbf{I}_M + \gamma_0 \mathbf{Q})}{1 + \gamma_0}. \quad (47)$$

Theorem 8 (Support-constrained information converse). *The ϵ -capacity of the embodied channel satisfies*

$$C_\epsilon(L) \leq C_{\text{info}}^{\text{ub}}(L, \epsilon) \triangleq \frac{C_{\text{snap}}^{\text{ub}}(\mathcal{A}) + \frac{1}{L}h_2(\epsilon)}{(1 - \epsilon)T_p}, \quad (48)$$

where $h_2(\epsilon) = -\epsilon \log_2 \epsilon - (1 - \epsilon) \log_2 (1 - \epsilon)$ is the binary entropy function.

Proof. Consider any J -position codebook $\mathcal{C} \subseteq \mathcal{A}$ for which there exists a decoder g satisfying $P_{\text{max}}(\mathcal{C}, g) \leq \epsilon$. The maximum-error guarantee holds for every message individually. Hence, it also holds under any prior distribution on the

message set. For the converse, we choose the auxiliary prior $W \sim \text{Unif}\{1, \dots, J\}$.

Under this uniform prior, the average error probability is no larger than ϵ . By Fano's inequality,

$$H(W|\mathbf{Y}) \leq h_2(\epsilon) + \epsilon \log_2(J - 1) \leq h_2(\epsilon) + \epsilon \log_2 J.$$

Therefore,

$$\begin{aligned} I(W; \mathbf{Y}) &= H(W) - H(W|\mathbf{Y}) \\ &\geq \log_2 J - h_2(\epsilon) - \epsilon \log_2 J \\ &= (1 - \epsilon) \log_2 J - h_2(\epsilon), \end{aligned}$$

and hence

$$\log_2 J \leq \frac{I(W; \mathbf{Y}) + h_2(\epsilon)}{1 - \epsilon}. \quad (49)$$

It remains to upper-bound $I(W; \mathbf{Y})$ for an arbitrary feasible codebook. For one snapshot, $\mathbf{y}_\ell|W = j \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_j)$, where $\mathbf{R}_j = \sigma^2(\mathbf{I}_M + \gamma_0 \mathbf{a}(\mathbf{r}_j)\mathbf{a}^H(\mathbf{r}_j))$ and $\det(\mathbf{R}_j) = \sigma^{2M}(1 + \gamma_0)$, $\forall j$. Conditioned on $W = j$, the L snapshots are independent. Thus,

$$h(\mathbf{Y}|W = j) = \sum_{\ell=1}^L h(\mathbf{y}_\ell|W = j) = L \log_2 [(\pi e)^M \sigma^{2M}(1 + \gamma_0)].$$

Since this expression is common to all j , we can write

$$h(\mathbf{Y}|W) = L \log_2 [(\pi e)^M \sigma^{2M}(1 + \gamma_0)]. \quad (50)$$

We next upper-bound the unconditional entropy $h(\mathbf{Y})$. Although the snapshots are conditionally independent given W , they are not generally independent after marginalizing over W , because they share the same selected physical position. Therefore, we work with the stacked observation

$$\mathbf{y}_{\text{vec}} = \text{vec}(\mathbf{Y}) = [\mathbf{y}_1^T, \dots, \mathbf{y}_L^T]^T \in \mathbb{C}^{ML}.$$

Let

$$\mathbf{Q}_\mathcal{C} = \frac{1}{J} \sum_{j=1}^J \mathbf{a}(\mathbf{r}_j)\mathbf{a}^H(\mathbf{r}_j).$$

Since each $\mathbf{r}_j \in \mathcal{A}$, we have $\mathbf{Q}_\mathcal{C} \in \mathcal{Q}_A$. The per-snapshot unconditional covariance is

$$\bar{\mathbf{R}} = \mathbb{E}[\mathbf{R}_W] = \sigma^2(\mathbf{I}_M + \gamma_0 \mathbf{Q}_\mathcal{C}).$$

For two distinct snapshots $\ell \neq m$, conditioned on W they are independent and zero mean, so $\mathbb{E}[\mathbf{y}_\ell \mathbf{y}_m^H | W] = \mathbf{0}$. Averaging over W gives $\mathbb{E}[\mathbf{y}_\ell \mathbf{y}_m^H] = \mathbf{0}$, $\ell \neq m$. Therefore,

$$\text{Cov}(\mathbf{y}_{\text{vec}}) = \mathbf{I}_L \otimes \bar{\mathbf{R}}.$$

For a fixed covariance matrix, the proper complex Gaussian distribution maximizes differential entropy. Hence,

$$\begin{aligned} h(\mathbf{Y}) &= h(\mathbf{y}_{\text{vec}}) \leq \log_2 [(\pi e)^{ML} \det(\mathbf{I}_L \otimes \bar{\mathbf{R}})] \\ &= L \log_2 [(\pi e)^M \det(\bar{\mathbf{R}})]. \end{aligned} \quad (51)$$

Combining (50) and (51),

$$\begin{aligned} I(W; \mathbf{Y}) &= h(\mathbf{Y}) - h(\mathbf{Y}|W) \leq L \log_2 \frac{\det(\bar{\mathbf{R}})}{\sigma^{2M}(1 + \gamma_0)} \\ &= L \log_2 \frac{\det(\mathbf{I}_M + \gamma_0 \mathbf{Q}_\mathcal{C})}{1 + \gamma_0}. \end{aligned}$$

Since $\mathbf{Q}_{\mathcal{C}} \in \mathcal{Q}_{\mathcal{A}}$, (47) gives $I(W; \mathbf{Y}) \leq LC_{\text{snap}}^{\text{ub}}(\mathcal{A})$. Substituting it into (49), we obtain

$$\log_2 J \leq \frac{LC_{\text{snap}}^{\text{ub}}(\mathcal{A}) + h_2(\epsilon)}{1 - \epsilon}. \quad (52)$$

Dividing by LT_p and maximizing over all ϵ -reliable codebooks yields (48). \square

A closed-form universal relaxation follows by enlarging $\mathcal{Q}_{\mathcal{A}}$. Every $\mathbf{Q} \in \mathcal{Q}_{\mathcal{A}}$ satisfies

$$\mathbf{Q} \succeq \mathbf{0}, \quad \text{tr}(\mathbf{Q}) = 1. \quad (53)$$

Therefore,

$$\begin{aligned} C_{\text{snap}}^{\text{ub}}(\mathcal{A}) &\leq \sup_{\mathbf{Q} \succeq \mathbf{0}, \text{tr}(\mathbf{Q})=1} \log_2 \frac{\det(\mathbf{I}_M + \gamma_0 \mathbf{Q})}{1 + \gamma_0} \\ &= M \log_2 \left(1 + \frac{\gamma_0}{M}\right) - \log_2(1 + \gamma_0), \end{aligned}$$

where the last equality follows from the arithmetic-geometric mean inequality and is attained by $\mathbf{Q} = \frac{1}{M} \mathbf{I}_M$ in the relaxed set. This yields the following closed-form corollary.

Corollary 9 (Universal information upper bound). *The ϵ -capacity of embodied communication satisfies*

$$C_{\epsilon}(L) \leq \frac{C_{\text{snap}}^{\text{univ}} + \frac{1}{L} h_2(\epsilon)}{(1 - \epsilon) T_p}, \quad (54)$$

where $C_{\text{snap}}^{\text{univ}} = M \log_2 \left(1 + \frac{\gamma_0}{M}\right) - \log_2(1 + \gamma_0)$.

The universal bound in Corollary 9 is easy to evaluate but may be loose when the angular support of \mathcal{A} occupies only a small portion of the BS field of view. The support-constrained bound in Theorem 8 is generally tighter because it only allows covariance mixtures generated by physically feasible scatterer positions.

B. Geometric Packing Converse

The information-theoretic converse limits the amount of information that can be extracted from the sensing observations. We next derive a complementary converse based on the physical geometry of the agent plane. The intuition is: if two physical positions are too close in the sensing geometry, then no binary detector can distinguish them with sufficiently small error. Thus, any maximum-error reliable codebook must have a minimum physical separation, which limits the number of codewords that can be placed in \mathcal{A} .

We first establish a necessary pairwise condition.

Lemma 10 (Necessary pairwise Bhattacharyya separation). *Consider two codeword positions \mathbf{r}_i and \mathbf{r}_j with displacement Δ_{ij} . Suppose that the BS can distinguish these two positions from L snapshots with error probability at most ϵ under both hypotheses. Their single-snapshot Bhattacharyya distance must satisfy*

$$B(\Delta_{ij}) \geq B_{\text{nec}}(\epsilon, L), \quad (55)$$

where $B_{\text{nec}}(\epsilon, L) = \frac{1}{2L} \log \frac{1}{4\epsilon(1-\epsilon)}$.

Proof. Let $P_i^{(L)}$ and $P_j^{(L)}$ denote the L -snapshot distributions induced by \mathbf{r}_i and \mathbf{r}_j , respectively. Their Bhattacharyya co-

efficient is $\beta_{ij}^{(L)} = \exp[-LB(\Delta_{ij})]$. For equal priors, the minimum binary Bayes error probability satisfies

$$P_b^* = \frac{1}{2} \left(1 - \|P_i^{(L)} - P_j^{(L)}\|_{\text{TV}}\right),$$

where $\|\cdot\|_{\text{TV}}$ denotes total variation distance. Using the standard relation between total variation distance and Bhattacharyya coefficient,

$$\|P_i^{(L)} - P_j^{(L)}\|_{\text{TV}} \leq \sqrt{1 - \left(\beta_{ij}^{(L)}\right)^2},$$

we obtain

$$P_b^* \geq \frac{1}{2} \left(1 - \sqrt{1 - \exp[-2LB(\Delta_{ij})]}\right).$$

If a binary test has error probability no larger than ϵ under both hypotheses, then its equal-prior average error probability is also no larger than ϵ . Therefore, it is necessary that

$$\frac{1}{2} \left(1 - \sqrt{1 - \exp[-2LB(\Delta_{ij})]}\right) \leq \epsilon.$$

Rearranging gives $\exp[-2LB(\Delta_{ij})] \leq 4\epsilon(1 - \epsilon)$, which is equivalent to (55). \square

Lemma 10 states that reliable pairwise discrimination requires a minimum amount of sensing distinguishability. Accordingly, define the necessary forbidden region

$$\mathcal{F}_{\text{nec}}(\epsilon, L) = \{\Delta \in \mathbb{R}^2 : B(\Delta) < B_{\text{nec}}(\epsilon, L)\}. \quad (56)$$

Any pair of codewords in an ϵ -reliable maximum-error codebook must have a displacement outside $\mathcal{F}_{\text{nec}}(\epsilon, L)$.

The exact shape of $\mathcal{F}_{\text{nec}}(\epsilon, L)$ is generally anisotropic and may be nonconvex because of array sidelobes. For an explicit geometric upper bound, we extract from it a Euclidean necessary separation. Define

$$d_{\text{nec}}(\epsilon, L) = \sup \{d \geq 0 : \{\Delta : \|\Delta\|_2 \leq d\} \subseteq \mathcal{F}_{\text{nec}}(\epsilon, L)\}.$$

Thus, $d_{\text{nec}}(\epsilon, L)$ is the radius of the largest Euclidean ball around the origin that is entirely contained in the necessary forbidden region.

Theorem 11 (Geometric packing upper bound). *Assume $\epsilon < 1/2$. The embodied ϵ -capacity satisfies*

$$\begin{aligned} C_{\epsilon}(L) &\leq \frac{1}{LT_p} \log_2 \left[\frac{4}{\pi d_{\text{nec}}^2(\epsilon, L)} (a_y a_z \right. \\ &\quad \left. + (a_y + a_z) d_{\text{nec}}(\epsilon, L) + \frac{\pi d_{\text{nec}}^2(\epsilon, L)}{4} \right). \end{aligned}$$

Proof. Consider any ϵ -reliable embodied codebook \mathcal{C} with a decoder g such that $\Pr(g(\mathbf{Y}) \neq i | W = i) \leq \epsilon, \forall i$. The multi-hypothesis decoder g can be converted into a binary test between hypotheses i and j by declaring hypothesis i if $g(\mathbf{Y}) = i$, and declaring hypothesis j otherwise. Under hypothesis i , this binary test makes an error only when the original decoder fails to output i , so its error probability is at most ϵ . Under hypothesis j , the binary test declares i only when the original decoder outputs i , which is contained in the

event that the original decoder fails to output j . Thus, its error probability under hypothesis j is also at most ϵ .

By Lemma 10, every pair of distinct codewords must satisfy

$$B(\mathbf{r}_i - \mathbf{r}_j) \geq B_{\text{nec}}(\epsilon, L).$$

That is, $\mathbf{r}_i - \mathbf{r}_j \notin \mathcal{F}_{\text{nec}}(\epsilon, L)$, $\forall i \neq j$.

By the definition of $d_{\text{nec}}(\epsilon, L)$, any displacement with Euclidean norm no larger than $d_{\text{nec}}(\epsilon, L)$ lies inside $\mathcal{F}_{\text{nec}}(\epsilon, L)$. Therefore,

$$\|\mathbf{r}_i - \mathbf{r}_j\|_2 > d_{\text{nec}}(\epsilon, L), \quad \forall i \neq j.$$

Now place a disk of radius $d_{\text{nec}}(\epsilon, L)/2$ centered at each codeword. These disks are non-overlapping. Since the codeword centers lie in the rectangle \mathcal{A} , all disks are contained in the expanded region

$$\mathcal{A} \oplus \mathcal{B} \left(0, \frac{d_{\text{nec}}(\epsilon, L)}{2} \right),$$

where \oplus denotes the Minkowski sum [24]. The area of this expanded region is

$$a_y a_z + (a_y + a_z) d_{\text{nec}}(\epsilon, L) + \frac{\pi d_{\text{nec}}^2(\epsilon, L)}{4}.$$

Each disk has area $\frac{\pi d_{\text{nec}}^2(\epsilon, L)}{4}$. Because the disks are non-overlapping, their total area cannot exceed the area of the expanded region. Therefore,

$$J \leq \frac{a_y a_z + (a_y + a_z) d_{\text{nec}}(\epsilon, L) + \frac{\pi d_{\text{nec}}^2(\epsilon, L)}{4}}{\frac{\pi d_{\text{nec}}^2(\epsilon, L)}{4}}.$$

Taking $\log_2(\cdot)/(LT_p)$ and maximizing over all ϵ -reliable codebooks gives the upper bound. \square

The bound in Theorem 11 is explicit once $d_{\text{nec}}(\epsilon, L)$ is known. It is conservative because it uses the largest Euclidean ball contained in the necessary forbidden region rather than the full anisotropic shape of $\mathcal{F}_{\text{nec}}(\epsilon, L)$. We now give a closed-form approximation for $d_{\text{nec}}(\epsilon, L)$ under the main-lobe model of Lemma 5.

Corollary 12 (Main-lobe geometric upper bound). *Let $\alpha_{\text{max}} = \max\{\alpha_y, \alpha_z\}$. Under the main-lobe quadratic approximation, the necessary Euclidean separation is approximated by*

$$d_{\text{nec}}^{\text{ml}}(\epsilon, L) \approx \sqrt{\frac{1}{2\kappa L \alpha_{\text{max}}} \log \frac{1}{4\epsilon(1-\epsilon)}}. \quad (57)$$

Substituting $d_{\text{nec}}^{\text{ml}}(\epsilon, L)$ into the geometric packing upper bound gives the closed-form approximation

$$C_\epsilon(L) \lesssim \frac{1}{LT_p} \log_2 \left[1 + \frac{4(a_y + a_z)}{\pi} \sqrt{\frac{\Omega_{\text{max}} L}{\log \frac{1}{4\epsilon(1-\epsilon)}}} + \frac{4a_y a_z}{\pi} \frac{\Omega_{\text{max}} L}{\log \frac{1}{4\epsilon(1-\epsilon)}} \right],$$

where $M_{\text{max}} \triangleq \max\{M_y^2 - 1, M_z^2 - 1\}$ and $\Omega_{\text{max}} \triangleq \frac{\pi^2 \gamma_0^2 M_{\text{max}}}{24 D^2 (1 + \gamma_0)}$.

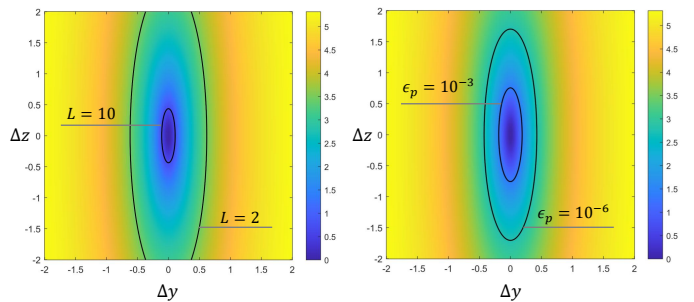


Figure 3: Sensing-induced reliability field $B(\Delta)$ determines the pairwise distinguishability of embodied symbols.

Proof. Under the main-lobe approximation,

$$B(\Delta y, \Delta z) \approx \kappa (\alpha_y \Delta y^2 + \alpha_z \Delta z^2).$$

For a fixed Euclidean radius d , the maximum value of this quadratic form over all directions is

$$\max_{\|\Delta\|_2=d} B(\Delta) \approx \kappa \alpha_{\text{max}} d^2.$$

If $\kappa \alpha_{\text{max}} d^2 < B_{\text{nec}}(\epsilon, L)$, then even the best-resolved direction at radius d cannot meet the necessary binary distinguishability threshold. Hence all displacements of norm no larger than $\sqrt{\frac{B_{\text{nec}}(\epsilon, L)}{\kappa \alpha_{\text{max}}}}$ belong to the necessary forbidden region under the quadratic approximation. This gives (57). \square

For achievability, a conservative Euclidean design must protect against the worst-resolved direction of the BS sensing field. For the converse, however, a pair of positions can be ruled out only if even the best-resolved direction cannot provide enough distinguishability. This is why the closed-form necessary separation in (57) is governed by α_{max} .

VI. NUMERICAL RESULTS

This section numerically evaluates the sensing-induced geometry and the achievable performance of embodied communication. Unless otherwise specified, the carrier frequency is $f_c = 7$ GHz, the BS is equipped with a 64×16 UPA, the BS-agent distance is $D = 100$ m, the number of snapshots $L = 5$, and the target maximum error probability is $\epsilon = 10^{-3}$. The controllable agent region is a square with $a_y = a_z = 2$ m. To remove the dependence on the particular probing-pulse duration, we report the rate in normalized units of bits per probing duration, i.e., $R = \frac{1}{L} \log_2 J$.

A. Sensing-Induced Reliability Fields

We first examine the reliability field $B(\Delta)$ in (23) induced by the BS sensing system on the agent plane. This field maps a physical displacement between two possible scatterer positions to the single-snapshot Bhattacharyya distance of the corresponding sensing distributions. It is therefore the basic geometric object that determines how physical positions can be used as embodied symbols.

Fig. 3 shows the two-dimensional field $B(\Delta)$ together with the contours corresponding to different snapshot numbers L and target pairwise error probabilities ϵ_p .

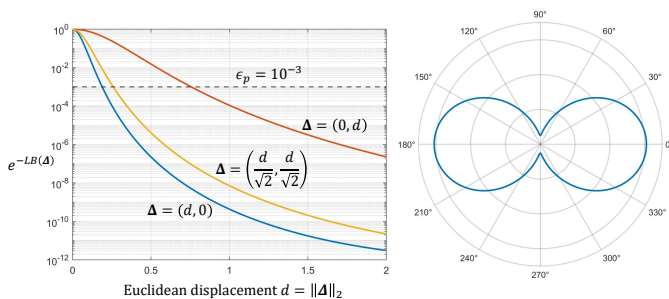


Figure 4: Directional pairwise error behavior. The same Euclidean displacement may lead to different error exponents.

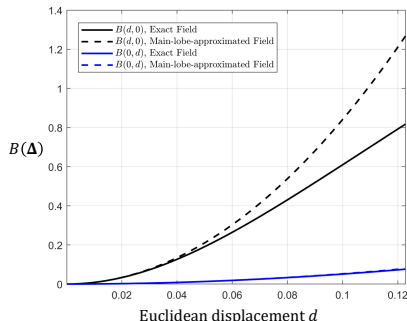


Figure 5: Exact Bhattacharyya distance versus the main-lobe quadratic approximation.

- For a fixed ϵ_p , increasing L reduces the required single-snapshot exponent $B_{\text{req}} = \frac{1}{L} \log \frac{1}{\epsilon_p}$, and therefore shrinks the forbidden displacement region. More snapshots allow the BS to distinguish closer physical positions and hence support denser embodied codebooks.
- For a fixed L , decreasing ϵ_p has the opposite effect: it requires a larger Bhattacharyya exponent, enlarges the forbidden region, and reduces the number of admissible codewords. This directly illustrates the operational meaning of the ϵ -capacity in Section III.

Fig. 4 further plots the pairwise error bound $\exp[-LB(\Delta)]$ as a function of Euclidean displacement along representative directions. The decay rates are direction-dependent, which means that two codeword pairs with the same Euclidean separation can have different sensing distinguishability. This behavior is caused by the anisotropic UPA aperture and confirms that the relevant codebook geometry is not ordinary Euclidean distance, but the sensing-induced field $B(\Delta)$.

The directional polar plot in Fig. 4 provides another view of the same phenomenon by plotting $B(d \cos \psi, d \sin \psi)$ against the direction angle ψ for a fixed displacement radius d . The anisotropic polar profile reflects the unequal resolution of the UPA along different directions. This is precisely why the lattice codebook in Section IV is designed in the transformed reliability domain rather than directly as an isotropic grid on the physical plane.

We next validate the main-lobe approximation used to derive the closed-form lattice design in Section IV. Recall that the exact reliability field is induced by the UPA Dirichlet kernels, while the main-lobe approximation replaces it locally by

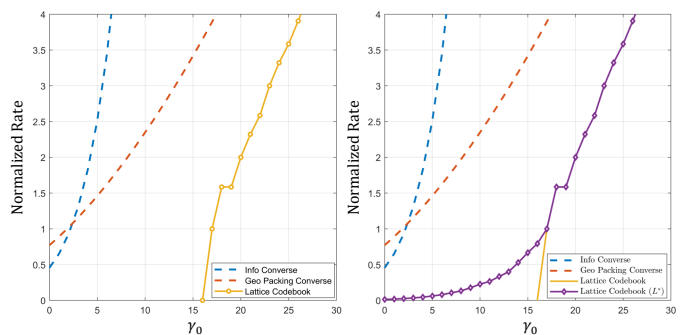


Figure 6: Achievable normalized rate versus sensing SNR, where $M_y = M_z = 64$ and $L = 1$. Optimizing over L improves the rate by balancing sensing reliability and symbol duration.

the quadratic form $B(\Delta) \approx \Delta^T \mathbf{G}_B \Delta$. This approximation underlies the transformed hexagonal lattice construction in Theorem 6.

Fig. 5 compares the exact Bhattacharyya distance with its quadratic approximation along representative displacement directions. The approximation is accurate near the origin, which is the dominant regime for dense codebook packing. For larger displacements, the exact field deviates from the quadratic model because of sidelobes and nonlocal array-response effects.

B. Achievable Rate and Converses

We now evaluate the achievable rate of the proposed hexagonal lattice codebook and compare it with the converses established in Section V.

Fig. 6 contains two subplots. We first focus on the left subplot, where the number of snapshots is fixed as $L = 1$. As the sensing SNR increases, the achievable rate increases. This is consistent with the analytical dependence of the reliability field on γ_0 : a larger SNR increases the Bhattacharyya distance between nearby positions, reduces the effective forbidden region, and allows more physical positions to be reliably used as codewords.

However, the fixed- L result only captures one operating point. As discussed in Section IV-B, the number of snapshots itself is a key design variable in embodied communication. Increasing L improves sensing reliability because pairwise error probabilities decay exponentially with L . At the same time, one embodied symbol occupies LT_p seconds, so the normalized rate is penalized by the factor $1/L$. This creates a sensing-duration tradeoff: a larger L can support a larger alphabet, but too large an L reduces the rate per unit probing duration.

To make this tradeoff explicit, Fig. 7 plots the normalized rate as a function of L for several SNR values. As shown,

- The variation of L creates a multimodal rate curve and there exists an optimal sensing duration.
- Higher SNR shifts the optimum toward smaller L . This is because each snapshot is more informative at high SNR, hence fewer snapshots are sufficient to resolve a dense physical alphabet. At lower SNR, more snapshots are needed to accumulate enough sensing evidence.

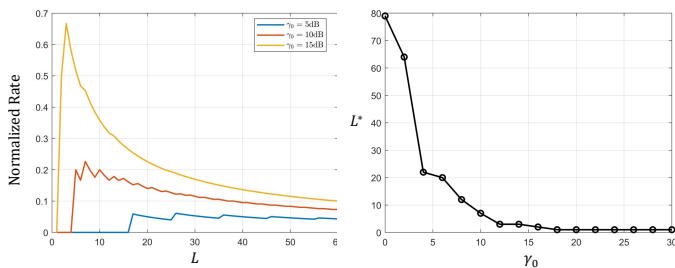


Figure 7: Normalized rate as a function of L , and the optimal snapshot number versus sensing SNR.

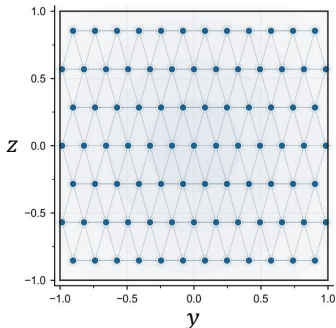


Figure 8: Example hexagonal lattice codebook on the agent plane, where $M_y = 64$ and $M_z = 32$.

- Since the rate is normalized by $1/L$, excessive sensing duration eventually reduces the rate.

Fig. 7 further plots the optimized snapshot number L^* versus SNR. As the SNR increases, fewer snapshots are needed to achieve the same level of position distinguishability. This agrees with the closed-form expression in Theorem 7, where a larger transformed-area parameter Ξ_h reduces the optimal sensing duration. We point out that small fluctuations of the optimized integer-valued L^* can occur because both L and the supported alphabet size are discrete, and because finite-region truncation affects the realized lattice codebook.

With the role of L clarified, we return to the right subplot of Fig. 6, where L is optimized for each SNR. Compared with the fixed- L case, optimizing the sensing duration improves the achievable rate, especially in the low-SNR regimes where a single snapshot is insufficient to resolve a dense alphabet.

Finally, Fig. 8 visualizes a generated hexagonal lattice codebook on the agent plane. Because the BS-induced reliability metric is anisotropic, the resulting codebook is not an ordinary isotropic grid in the physical domain. Instead, it is stretched and compressed according to the sensing resolution of the UPA. In directions where the BS provides finer discrimination, the codewords can be placed more densely; in weaker directions, the spacing is larger.

This layout gives a direct physical interpretation of embodied communication. The agent plane becomes a physical alphabet, and the BS sensing system determines how finely that alphabet can be carved.

VII. CONCLUSION

This paper has positioned embodied communication as a new way to think about information exchange in sensing-rich environments. Its central implication is that communication need not be confined to transmitter-generated waveforms. When agents can intentionally shape the physical states around them, and when infrastructure can perceive those states, the environment itself becomes an information interface. This viewpoint expands the role of future infrastructure from a data pipe or sensing node to an interpreter of purposeful physical change.

The broader impact of embodied communication is that it creates a communication pathway for entities that may not possess, activate, or continuously operate a conventional radio transmitter. Information can instead be conveyed through controllable presence, motion, configuration, reflection, blockage, or other observable physical effects. This may enable low-power, low-complexity, passive, or mechanically constrained agents to participate in networked systems in ways that are difficult to support through conventional wireless links alone.

The vision also goes beyond the cellular and RF setting studied in this paper. Any sensing infrastructure may serve as the receiver, including cameras, LiDAR, acoustic arrays, thermal sensors, magnetic sensors, or multimodal perception systems. Likewise, the communicating agent may be a robot, vehicle, human, wearable object, industrial actuator, reconfigurable surface, or embedded device. Embodied communication should therefore be viewed not as a special case of RF sensing, but as a general architectural principle for sensor-rich intelligent environments.

REFERENCES

- [1] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1728–1767, 2022.
- [2] A. Liu, Z. Huang, M. Li, Y. Wan, W. Li, T. X. Han, C. Liu, R. Du, D. K. P. Tan, J. Lu *et al.*, "A survey on fundamental limits of integrated sensing and communication," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 994–1034, 2022.
- [3] W. Xu, F. Gao, X. Tao, J. Zhang, and A. Alkhateeb, "Computer vision aided mmWave beam alignment in V2X communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 4, pp. 2699–2714, 2022.
- [4] R. Zhang, Y. Shao, M. Li, L. Lu, and Y. C. Eldar, "Optical integrated sensing and communication with light-emitting diode," *IEEE Internet of Things Journal*, vol. 12, no. 9, pp. 12896–12911, 2024.
- [5] Y. Shao, C. Bian, L. Yang, Q. Yang, Z. Zhang, and D. Gunduz, "Point cloud in the air," *IEEE Communications Magazine*, 2025.
- [6] C. Bian, Y. Shao, and D. Gündüz, "Over-the-air learning-based geometry point cloud transmission," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 12, pp. 4042–4058, 2025.
- [7] F. Liu, W. Yuan, C. Masouros, and J. Yuan, "Radar-assisted predictive beamforming for vehicular links: Communication served by sensing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7704–7719, 2020.
- [8] Y. Heng, J. Mo, and J. G. Andrews, "Learning site-specific probing beams for fast mmWave beam alignment," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 5785–5800, 2022.
- [9] A. Boukerche, H. A. Oliveira, E. F. Nakamura, and A. A. Loureiro, "Vehicular ad hoc networks: A new challenge for localization-based systems," *Computer Commun.*, vol. 31, no. 12, pp. 2838–2849, 2008.
- [10] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using wifi," in *ACM SIGCOMM*, 2015, pp. 269–282.
- [11] Y. Shao, D. Gündüz, and S. C. Liew, "Federated edge learning with misaligned over-the-air computation," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 3951–3964, 2021.

- [12] M. Savva, A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik *et al.*, “Habitat: A platform for embodied AI research,” in *IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9339–9347.
- [13] R. Chrisley, “Embodied Artificial Intelligence,” *Artificial intelligence*, vol. 149, no. 1, pp. 131–150, 2003.
- [14] Y. Shao, Q. Cao, and D. Gündüz, “A theory of semantic communication,” *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, pp. 12 211–12 228, 2024.
- [15] L. Shapiro, *Embodied cognition*. Routledge, 2019.
- [16] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [17] M. A. Herman and T. Strohmer, “High-resolution radar via compressed sensing,” *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.
- [18] S. M. Kay, “Statistical signal processing: estimation theory,” *Prentice Hall*, vol. 1, pp. Chapter–3, 1993.
- [19] S. T. Smith, “Statistical resolution limits and the complexified Cramer-Rao bound,” *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1597–1609, 2005.
- [20] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [21] T. Kailath, “The divergence and Bhattacharyya distance measures in signal selection,” *IEEE Transactions on Communication Technology*, vol. 15, no. 1, pp. 52–60, 1967.
- [22] R. Zamir, *Lattice coding for signals and networks: A structured coding approach to quantization, modulation, and multiuser information theory*. Cambridge University Press, 2014.
- [23] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.
- [24] G. Varadhan and D. Manocha, “Accurate Minkowski sum approximation of polyhedral models,” in *IEEE Pacific Conference on Computer Graphics and Applications*, 2004, pp. 392–401.