SCIENCE CHINA
Information Sciences

• Supplementary File •

Measuring discrete sensing capability for ISAC via task mutual information

Fei SHANG[1], Haohua DU[2*], Panlong YANG[3], Xin HE[4], Jingjing WANG[2] & Xiang-Yang LI[1*]

[1]*School of Computer Science and Technology, University of Science and Technology of China, Hefei* 230026*, China*
[2]*School of Cyber Science and Technology, Beihang University, Beijing* 100191*, China*
[3]*School of Computer Science, Nanjing University of Information Science and Technology, Nanjing* 210044, China
[4]*School of Computer and Information, Anhui Normal University, Wuhu,* 241008, China

Appendix A Sensing channel encoder model

Appendix A.1 Lower bound on expected error

The current evaluation of sensing systems' performance predominantly relies on experimental assessments. While experimental evaluations are highly effective in gauging system performance, conducting rigorous controlled experiments in real-world scenarios is exceedingly challenging. Consequently, in many instances, it is difficult to ascertain whether the failure to achieve the desired accuracy is due to inadequately designed sensing features or simply unforeseen interference during the data acquisition process. In this section, we give a lower bound on the expected error value based on DTMI, which helps us analyze the ultimate performance of the sensing system.

Theorem A1. For a sensing task W with m statuses, we use n independent features to describe the status of the target. The expected value of the error P_E^n satisfies the following lower bound:

$$P_E^n \geqslant \frac{H(W) - I(X^n;Y^n) - H(P_E^n)}{\log m},$$

where $H(P_E^n) = -P_E^n \log P_E^n - (1-P_E^n)\log(1-P_E^n)$.

Proof. We first prove that the sensing model we defined forms a Markov chain. Then we combine Fano's inequality [1] and some properties of Markov chains to give a lower bound for P_E^n.

According to the definition and properties of Markov chains [2], for the sensing model we described, the target status W, the feature X^n, the received channel embedding Y^n, and the sensing result $\hat{W}$ form two Markov chains, i.e., $W \to X^n \to Y^n \to \hat{W}$ and $\hat{W} \to Y^n \to X^n \to W$. According to the Fano's inequality [2], if three random variables X, Y, Z form a Markov chain, i.e., $X \to Y \to Z$, we have:

$$\Pr(X \neq Z) \geqslant \frac{H(X|Z) - H(\Pr(X \neq Z))}{\log(|\mathcal{X}|)}, \tag{A1}$$

where $H(X|Y)$ is the conditional entropy of X given Y. For the Markov chain $W \to X^n \to Y^n \to \hat{W}$, according to the total probability formula and Ferno's inequality, we have:

$$P_E^n = \Pr(\hat{W} \neq W) \geqslant \frac{H(W|\hat{W}) - H(P_E^n)}{\log(|\mathcal{W}|)} = \frac{H(W) - I(W;\hat{W}) - H(P_E^n)}{\log m}. \tag{A2}$$

According to the Data-processing inequality [2], if three random variables X, Y, and Z form a Markov chain, $X \to Y \to Z$, then we have $I(X;Z) \leqslant I(X;Y)$, where $I(X;Y)$ is the mutual information between X and Y. For the Markov chain $W \to X^n \to Y^n \to \hat{W}$, we have $I(W;\hat{W}) \leqslant I(W;Y^n)$. And for the Markov chain $\hat{W} \to Y^n \to X^n \to W$, we have $I(Y^n;W) \leqslant I(Y^n;X^n)$. As a result, we have:

$$I(W;\hat{W}) \leqslant I(X^n;Y^n). \tag{A3}$$

Substituting Equ. (A3) into Equ. (A2), we have:

$$P_E^n \geqslant \frac{H(W) - I(X^n;Y^n) - H(P_E^n)}{\log m}.$$

* Corresponding author (email: duhaohua@buaa.edu.cn, xiangyangli@ustc.edu.cn)

Appendix A.2 Upper bound on expected error

In communication, Shannon's second theorem [3] posits that for a given signal, error-free transmission can always be achieved as long as we employ code words that are sufficiently long to encode the message. This issue is equally pertinent in sensing: when the dimensionality n of the feature is sufficiently large, what is the upper bound on the expected error? In this section, we derive an upper bound based on DTMI (Theorem A2) and provide a sufficient condition under which error-free sensing can be attained (Theorem A3).

Theorem A2. For a sensing task with m statuss, we use n independent features to describe the status of the target. For sufficiently large n, the expected value of the error P_E^n satisfies the following upper bound:

$$P_E^n \leqslant \varepsilon + \sum_{k=1}^{m} p(w_k) \sum_{j \neq k}^{m} 2^{3n\varepsilon - \sum_{i=1}^{n} I(X_i(w_j);Y_i(w_k))}.$$

Proof. The expected error P_E^n is influenced by the decision rule g, with the maximum likelihood criterion being a commonly employed rule in practical scenarios. However, for the sake of facilitating analysis, we introduce a novel decision rule defined in conjunction with the matching set $B_\varepsilon^{(n)}$ (Definition 3), where in the result $\hat{W}$ is determined as w_i whenever the channel embedding Y^n and the feature $X^n(w_i)$ corresponding to the message w_i form a jointly matching set. Under this rule, we first estimate the probability of X^n, Y^n constituting a jointly matching set (Lemma A1 to A3) and subsequently present a suboptimal upper bound on the expected error (it is noted that employing alternative decision criteria might yield tighter upper bounds).

Definition 1. If a sequence $X^n = [X_1, \cdots, X_n]$ of length n, where each dimension is statistically independent of one another, we refer to sequence X^n as an n-dimensional **independent sequence**. Their joint probability density function is given by:

$$p(x^n) = \Pi_{i=1}^{n} p(x_i). \tag{A4}$$

Definition 2. For two n-dimensional independent sequences X^n and Y^n, if the joint distribution of (X^n, Y^n) is given by

$$p(x^n, y^n) = \Pi_{i=1}^{n} p(x_i, y_i), \tag{A5}$$

we refer to (X^n, Y^n) as a n-dimensional **jointly independent sequence**.

Definition 3. The **jointly matching set** $B_\varepsilon^{(n)}$ of jointly independent sequence is defined as:

$$\begin{aligned} B_\varepsilon^{(n)} = \{&(X^n, Y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \\ &\left| -\frac{1}{n}\log p(x^n) - \frac{1}{n}\sum_{i=1}^{n} H(X_i) \right| < \varepsilon \\ &\left| -\frac{1}{n}\log p(y^n) - \frac{1}{n}\sum_{i=1}^{n} H(Y_i) \right| < \varepsilon \\ &\left| -\frac{1}{n}\log p(x^n, y^n) - \frac{1}{n}\sum_{i=1}^{n} H(X_i, Y_i) \right| < \varepsilon \Bigg\}, \end{aligned} \tag{A6}$$

where (X^n, Y^n) is the n-dimensional jointly independent sequence. $H(X_i)$, $H(Y_i)$, and $H(X_i, Y_i)$ are the entropy of X_i, Y_i, and (X_i, Y_i), respectively.

Definition 4. **The decoding rule** g. To obtain sensing outcomes from Y^n, we employ the following rule g:

- We declare that the target statue is w_i if $(X^n(w_i), Y^n) \in B_\varepsilon^{(n)}$ and there is no other status w_j such that $(X^n(w_j), Y^n) \in B_\varepsilon^{(n)}$.
- If there are multiple statuss w_j such that $(X^n(w_j), Y^n) \in B_\varepsilon^{(n)}$ or there is no status w_i such that $(X^n(w_i), Y^n) \in B_\varepsilon^{(n)}$, an error is declared.

To estimate the probability of an event occurring, we first prove the following lemma about matching sets.

Lemma A1. For a n-dimensional jointly independent sequence (X^n, Y^n) and a matching set $B_\varepsilon^{(n)}$, when $n \to \infty$, the probability that (X^n, Y^n) is in the matching set $B_\varepsilon^{(n)}$ is close to 1, which is

$$\Pr((X^n, Y^n) \in B_\varepsilon^{(n)}) \to 1. \tag{A7}$$

Proof. According to the Chebyshev's Law of Large Numbers, when the number of observations n is sufficiently large, the sample mean of n independent and identically distributed random variables converges in probability to their common expected value. Observing that the entropy is essentially the expectation of the logarithm of the reciprocal of probabilities, we leverage these two premises to underpin our proof.

According to Chebyshev's Law of Large Numbers, given $\varepsilon > 0$, there exists n_1, so that for all $n > n_1$, the following holds:

$$\begin{aligned} P_1 &= \Pr\left(\left| -\frac{1}{n}\log p\left(X^n\right) - \frac{1}{n}\sum_{i=1}^{n} H(X_i) \right| \geqslant \varepsilon \right) \\ &= \Pr\left(\left| \frac{1}{n}\sum_{i=1}^{n} \log p\left(X_i\right) - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\log p(X_i)\right) \right| \geqslant \varepsilon \right) < \frac{\varepsilon}{3}. \end{aligned} \tag{A8}$$

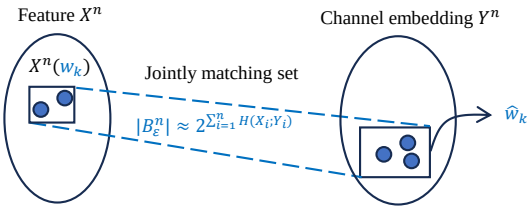

Figure A1 For independent sequences X^n and Y^n, the number of elements in their jointly matching set is approximately $2^{\sum_{i=1}^n H(X_i,Y_i)}$. We decode the channel embedding as $\hat{w}_k$ when Y^n forms a joint matching sequence with only one feature $X^n(w_k)$.

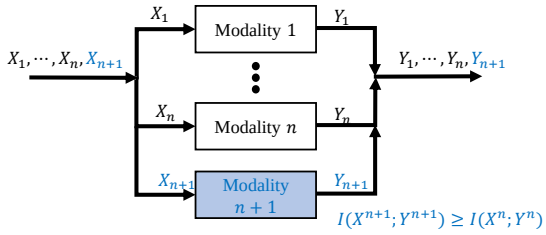

Figure A2 The sensing channel of multi-modal sensing system.

Similarly, there exists n_2 and n_3, so that for all $n > n_2$, the following holds:

$$P_2 = \Pr\left(\left|-\frac{1}{n}\log p\left(Y^n\right) - \frac{1}{n}\sum_{i=1}^n H(Y_i)\right| \geqslant \varepsilon\right) < \frac{\varepsilon}{3}, \tag{A9}$$

and for all $n > n_3$, the following holds:

$$P_3 = \Pr\left(\left|-\frac{1}{n}\log p\left(X^n,Y^n\right) - \frac{1}{n}\sum_{i=1}^n H(X_i,Y_i)\right| \geqslant \varepsilon\right) < \frac{\varepsilon}{3}. \tag{A10}$$

Let $n_0 = \max\{n_1, n_2, n_3\}$, then for all $n > n_0$, the following holds:

$$\Pr((X^n,Y^n) \in B_\varepsilon^{(n)}) > 1 - (P_1 + P_2 + P_3) = 1 - \varepsilon. \tag{A11}$$

Going further, we consider the scenario where (X^n, Y^n) forms a jointly independent sequence (Definition 2), and we examine the probability of them constituting a joint matching set. Initially, drawing upon Definition 3, we estimate the counts of elements in both the matching set and the jointly matching set, which are related to the entropy. Specifically, the number of elements in the matching set for X^n and Y^n are approximately $2^{\sum_{i=1}^n H(X_i)}$ and $2^{\sum_{i=1}^n H(Y_i)}$, respectively, while the count of their joint matching sequences is roughly $2^{\sum_{i=1}^n H(X_i,Y_i)}$. Building on this foundation, Lemma A3 furnishes an estimate for the probability that (X^n, Y^n) forms a joint matching set.

Lemma A2. The upper bound of the number of elements in the matching set of jointly independent sequence $B_\varepsilon^{(n)}$ is given by:

$$\|B_\varepsilon^{(n)}\| \leqslant 2^{n\varepsilon + \sum_{i=1}^n H(X_i,Y_i)}, \tag{A12}$$

where $H(X_i, Y_i)$ is the entropy of (X_i, Y_i), and $\|.\|$ denotes the number of elements in the set.

Proof. According to the Definition 3, if $(X^n, Y^n) \in B_\varepsilon^{(n)}$, we have:

$$p(x^n, y^n) \geqslant 2^{-n\varepsilon - \sum_{i=1}^n H(X_i,Y_i)}. \tag{A13}$$

As a result,

$$1 = \sum_{(x^n,y^n)\in\mathcal{X}^n\times(Y)^n} p(x^n,y^n) \geqslant \sum_{(x^n,y^n)\in B_\varepsilon^{(n)}} p(x^n,y^n) \geqslant 2^{-n\varepsilon-\sum_{i=1}^n H(X_i,Y_i)}|B_\varepsilon^{(n)}|. \tag{A14}$$

Therefore, we have

$$\|B_\varepsilon^{(n)}\| \leqslant 2^{n\varepsilon+\sum_{i=1}^n H(X_i,Y_i)}.$$

Lemma A3. For a n-dimensional jointly independent sequence $(\hat{X}^n, \hat{Y}^n)$ and a matching set $B_\varepsilon^{(n)}$, if $(\hat{X}^n, \hat{Y}^n) \sim p(x^n)p(y^n)$, i.e., $\hat{X}^n$ and $\hat{Y}^n$ are independent with the same marginals as $p(x^n, y^n)$, then

$$\Pr((\hat{X}^n,\hat{Y}^n) \in B_\varepsilon^{(n)}) \leqslant 2^{3n\varepsilon - \sum_{i=1}^n I(X_i;Y_i)}, \tag{A15}$$

where $I(X_i; Y_i)$ is the mutual information between X_i and Y_i.

Proof. According to the definition of the jointly matching set, we have:

$$\log(p(x^n)) \leqslant n\varepsilon - \sum_{i=1}^n H(X_i), \text{ and } \log(p(y^n)) \leqslant n\varepsilon - \sum_{i=1}^n H(Y_i). \tag{A16}$$

The probability of a joint independent sequence $(\hat{X}^n, \hat{Y}^n)$ in B_ε^n is given by:

$$\begin{aligned}\Pr((\hat{X}^n,\hat{Y}^n) \in B_\varepsilon^{(n)}) = \sum_{(x^n,y^n)\in B_\varepsilon^{(n)}} p(x^n)p(y^n) &\leqslant \left|B_\varepsilon^{(n)}\right| 2^{n\varepsilon-\sum_{i=1}^n H(X_i)} 2^{n\varepsilon-\sum_{i=1}^n H(Y_i)} \\ \leqslant 2^{3n\varepsilon+\sum_{i=1}^n (H(X_i,Y_i)-H(X_i)-H(Y_i))} &= 2^{3n\varepsilon-\sum_{i=1}^n I(X_i;Y_i)}.\end{aligned} \tag{A17}$$

We first estimate the probability that the sensing result $\hat{W}$ is wrong when the target status is $W = w_i$. We can assume without loss of generality that the target status is w_1. We consider the following events:

$$C_i = \left\{ (X^n(w_i), Y^n(w_1)) \in B_\varepsilon^{(n)} \right\}, \quad i \in \{1, \cdots, m\}, \tag{A18}$$

where $Y^n(w_1)$ is the received channel embedding when the target status is w_1. Based on the decision rule, the conditional error probability at this point is given by:

$$\xi_1 = Pr\left(\bar{C_1} \bigcup_{i=2}^{m} C_i \right) \leqslant Pr\left(\bar{C_1}\right) + \sum_{i=2}^{m} Pr\left(C_i\right), \tag{A19}$$

where $\bar{C_1}$ is the complement of C_1. According to Lemma A1, we have:

$$\Pr\left(\bar{C_1}\right) \leqslant \varepsilon. \tag{A20}$$

Besides, for $j \in \{2, \cdots, m\}$, the feature $X^n(w_j)$ is independent of $X^n(w_1)$, so is $X^n(j)$ and $Y^n(w_1)$. Hence, according to Lemma A3, we have:

$$\Pr\left(C_j\right) \leqslant 2^{3n\varepsilon - \sum_{i=1}^{n} I(X_i(w_j); Y_i(w_1))}. \tag{A21}$$

Substituting the above results into Eq. (A19), we have:

$$\xi_1 \leqslant \varepsilon + \sum_{j=2}^{m} 2^{3n\varepsilon - \sum_{i=1}^{n} I(X_i(w_j); Y_i(w_1))}. \tag{A22}$$

According to Definition of P_E^n, we have:

$$P_E^n = \sum_{k=1}^{m} p(w_k)\xi_k \leqslant \varepsilon + \sum_{k=1}^{m} p(w_k) \sum_{j \neq k}^{m} 2^{3n\varepsilon - \sum_{i=1}^{n} I(X_i(w_j); Y_i(w_k))}. \tag{A23}$$

Finally, Theorem A3 provides a sufficient condition for error-free sensing, indicating that for achieving error-free sensing, a sufficient number of features with high DTMI must be identified [1)].

Theorem A3. For a sensing task with $m = 2^{nR}$ statuss, we use n independent features to describe the status of the target. For a sufficiently large n, if R satisfies the following equation,

$$R < \min_{k \neq j} I(\bar{X}^n(w_k); \bar{Y}^n(w_j)) - 3\varepsilon, \tag{A24}$$

where $\bar{X}(w_j)$ and $\bar{Y}(w_j)$ is the mean $X^n(w_j)$ and $Y^n(w_j)$, we have $\xi_j \to 0$.

Proof. In Theorem A2, we derive an upper bound estimate for the expected error P_E^n. Capitalizing on the convexity property of mutual information, we leverage Jensen's inequality to provide a sufficient condition for a tight error estimation. This approach ensures that our estimate effectively captures the inherent relationship between the variables, harnessing the convexity to yield a more robust and accurate analysis of the error's expected magnitude without loss of generality.

According to the Jensen's inequality, if f is a convex function and X is a random variable, we have:

$$f(\mathbb{E}(X)) \leqslant \mathbb{E}(f(X)). \tag{A25}$$

Since the mutual information is a convex function [2], we have:

$$nI(\bar{X}^n; \bar{Y}^n) \leqslant n \sum_{i=1}^{n} \frac{1}{n} I(X_i; Y_i), \tag{A26}$$

where $\bar{X}^n$ and $\bar{Y}^n$ is the mean of X^n and Y^n. As a result, for a $j \in \{1, \cdots, m$, the Equ. (A19) can be rewritten as:

$$\xi_j \leqslant \varepsilon + \sum_{k \neq j}^{m} 2^{3n\varepsilon - nI(\bar{X}^n(w_k); \bar{Y}^n(w_j))}. \tag{A27}$$

As a result, for $m = 2^{nR}$ and sufficiently large n, if R satisfies the Equ. (A24), we have:

$$\xi_j \leqslant \varepsilon + 2^{3n\varepsilon + nR - n \min_{k \neq j} I(\bar{X}^n(w_k); \bar{Y}^n(w_j))} \to 2\varepsilon. \tag{A28}$$

Appendix B Corollary

Previous excellent sensing systems have summarized many valuable experiences, such as multi-modal systems tend to achieve better sensing performance. However, these experiences currently lack theoretical explainability. In this section, we employ sensing channel encoder model and DTMI as tools to attempt to explain some classic phenomena.

1) This requirement diverges from the conclusion in communications, where merely having a sufficient number of codewords is typically sufficient.

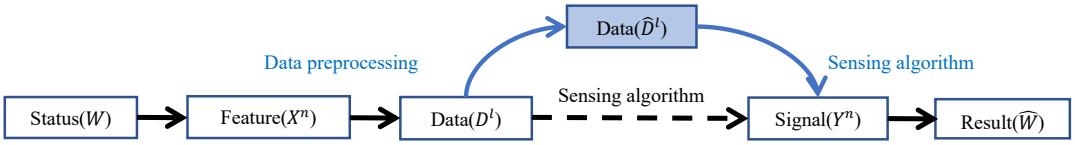

Figure B1 The sensing system with data pre-processing.

Appendix B.1 Why do multimodal systems tend to exhibit superior performance?

In a communication system, Shannon's second theorem stipulates that the error rate can be reduced to an arbitrary low level, provided that the codewords are sufficiently lengthy. Similarly, many previous research works have shown that using multi-modality for sensing helps achieve better performance, which can be explained by the theorem we proved previously. In this subsection, we will theoretically explain why multi-modal sensing systems are more capable of achieving superior sensing performance based on the DTMI.

Fig. A2 shows a schematic diagram of a multi-modal system. For the target state W, we use n modalities to sense it. The channels of different modalities are directly independent of each other. For example, in order to identify the material of the target, we use three modalities: vision, sound wave, and radio frequency signal for sensing. The transmission of visual signal, sound wave signal, and radio frequency signal is independent of each other. According to the Theorem A1, when the number of states m remains unchanged, the lower bound of the expected value of the error P_E^n is related to $I(X^n;Y^n)$. Note that both mutual information and conditional mutual information are non-negative. When we add a new mode, we have

$$\begin{aligned} I(X^{n+1};Y^{n+1}) &= I(X^n, X_{n+1}; Y^n, Y_{n+1}) \\ &= I(X^n;Y^n) + I(X^n;Y_{n+1}|Y^n) + I(X_{n+1};Y^{n+1}|X^n) \\ &\geqslant I(X^n;Y^n), \end{aligned} \tag{B1}$$

where $X^{n+1} = [X_1, X_2, \ldots, X_n, X_{n+1}]$ and $Y^{n+1} = [Y_1, Y_2, \ldots, Y_n, Y_{n+1}]$. Therefore, the more modalities we use, the larger the mutual information $I(X^n;Y^n)$, the lower the theoretical lower bound of the expected value of the error.

Appendix B.2 How do we compare which of two sensing features is better?

In the process of designing a sensing system, it is crucial to carefully craft the sensing features. To show that feature X is better than feature X', we usually need to run many micro-benchmarks. While experimental validation is a compelling method of verification, it frequently involves intricate setup procedures and can be time-consuming. Moreover, due to the challenge of deploying tests across a wide range of scenarios, it is often difficult to ascertain whether feature X is truly superior to feature X' or if this conclusion holds only in specific contexts.

In this paper, we propose DTMI which can reflect the performance of sensing features to a certain extent. Specifically, we consider two features X and X'. After passing through the sensing channel, their corresponding channel embeddings are Y and Y', respectively. According to Theorem A1 and Theorem A2, both the upper and lower bounds of the expected error are related to the DTMI. If the DTMI $I(X;Y) > I(X';Y')$, the upper and lower bounds of the expected value of the error P_E will be reduced, which means that it is easier to achieve good performance using X as sensing features. This necessitates alternative approaches, beyond experimental validation, to assess the performance of designed sensing features.

Appendix B.3 Is data pre-processing a "cure-all" solution?

Since data contains a lot of noise and interference, sensing systems usually include a data preprocessing module when they are designed, which is used to improve data quality for subsequent processing. Previous studies have shown that preprocessing can often improve sensing performance. Now our questions are: can we accomplish any sensing task with arbitrary accuracy through sufficiently sophisticatedly designed data preprocessing algorithms?

We refine the sensing channel encoder model, and the result is illustrated in Fig. B1. Specifically, for the n-dimensional independent features X^n, after transmission through an actual physical channel, we obtain an l-dimensional data D^l at the receiver. For instance, to localize a target using radio frequency (RF) signals, we employ angle of arrival (AoA) as a feature. At the receiver, what we receive is the amplitude and phase of the RF signals, which are D^l. Subsequently, we subject the received data D^l to data preprocessing, yielding a processed data $\hat{D}^l$. Then we utilize the sensing algorithm to process the data $\hat{D}^l$ to obtain the channel embedding Y^n, and finally use the judgment algorithm to obtain the result $\hat{W}$. In particular, when no data preprocessing is used, it is equivalent to $\hat{D}^l = D^l$.

Corollary 1. If the following equation holds,

$$H(W) - I(X^n;D^l) > 1, \tag{B2}$$

lossless sensing cannot be achieved simply by improving the effect of data preprocessing.

Proof. According to the definition Markov chain, the channel shown in Fig. B1 constitutes a Markov chain $W \to X^n \to D^l \to \hat{D}^l \to Y^n \to \hat{W}$. Note that "whether the sensing result is correct" is a binary event, so we have $H(P_E^n) \leqslant 1$. According to the Theorem A1 and the Data-Processing Inequality, we have

$$P_E^n \geqslant \frac{H(W) - I(X^n;Y^n) - H(P_E^n)}{\log m} \geqslant \frac{H(W) - I(X^n;D^l) - 1}{\log m} \geqslant 0. \tag{B3}$$

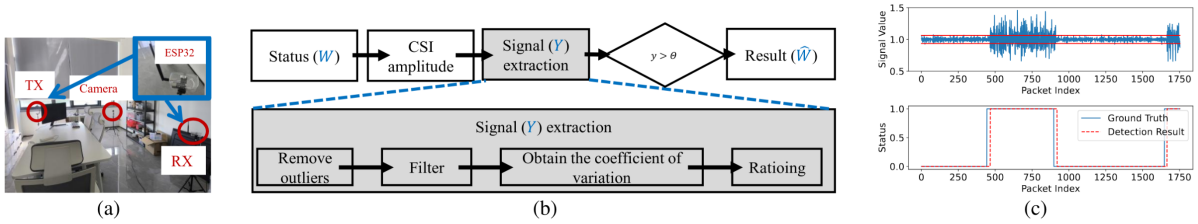

Figure C1 Human detection in home environments based on WiFi. (a) Experimental environment and device deployment. (b) Human detection algorithm based on thresholding method. For state W, we use the coefficient of variation as a feature (encoding). Then, we obtain the channel embedding Y from the received signal and subsequently use the threshold method for state discrimination to obtain the sensing result $\hat{W}$. (c) Channel embedding extraction.

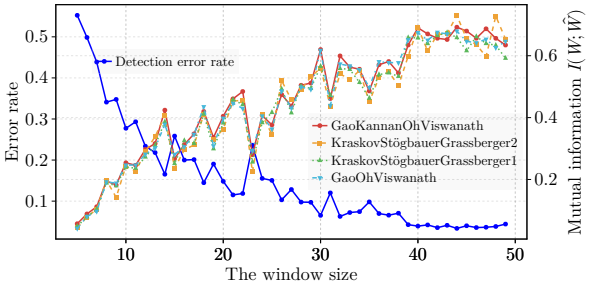

Figure C2 The accuracy exhibits a similar trend to the mutual information estimated by numerical algorithms.

Therefore, lossless sensing cannot be achieved simply by improving the effect of data preprocessing.

Appendix C Case study

We illustrate the role of system performance evaluation based on sensing channel encoder model and DTMI through several case studies. We begin by examining the application of DTMI in binary classification tasks, using examples of human detection in home settings via WiFi and appliance cabinet door displacement detection in industrial scenarios via RFID. For multi-class classification, we consider two instances: the classic sensing problem in ISAC systems – direction estimation, and device identification based on an open-source traffic dataset. Furthermore, DTMI can provide estimates of upper and lower bounds for sensing system errors, which is beneficial for optimizing and balancing ISAC systems.

Appendix C.1 Binary classification task.

(1) **Human detection based on Wi-Fi devices.**

Indoor human detection plays a pivotal role in services such as elderly monitoring. In particular, device-free passive human detection has garnered significant attention in recent years. While methods based on infrared, pressure sensors, and the like have been applied to human detection, they either rely on specialized hardware or come at a higher cost. Moreover, vision-based and infrared-based methods are only effective within line-of-sight (LOS) coverage. Wi-Fi devices, being one of the most widely deployed radio frequency devices, have led to the implementation of numerous radio frequency sensing systems around them. In recent years, with the advancement of wireless sensing technology, Wi-Fi-based approaches have proven to be a promising method for indoor human detection. We deployed an experiment based on Wi-Fi devices in a residential setting and estimated mutual information using numerical methods. The experimental results indicate that DTMI exhibits a similar trend to accuracy. We used the cor function from Statistics.jl library in Julia language to calculate Pearson coefficient, where one input is mutual information and another input is sensing accuracy. In this case study, their Pearson correlation coefficient exceeds 0.9.

The experimental setup is depicted in Fig. C1(a), where we conducted experiments in a 4 m × 6 m office using an ESP32 device as both transmitter and receiver, each equipped with a single antenna. Additionally, a camera was placed within the environment to capture video footage for recording ground truth. The sampling rate of the ESP32 is set to 100 Hz. Ten volunteers are invited to participate in the tests. Each data acquisition session lasted 10 minutes: the first 5 minutes ensured the room is empty, followed by 5 minutes with human activity (walking) inside the room. The demo are shown in https://github.com/zaoanhh/DTMI/blob/main/human%20dection/demo.

State W has two possible values: "personnel present" and "personnel absent". After obtaining CSI data, we initially sliced the data, then performed data preprocessing to eliminate outliers and apply filtering. Finally, channel embedding Y is extracted from this processed data and compared against empirical thresholds to ascertain the presence or absence of individuals, which is the result $\hat{W}$. The entire data processing procedure is illustrated in Fig. C1(b). The corfficient of variation of k-th subcarrier is $\delta^k_{\Delta T} = \sigma^k_{\Delta T}/\mu^k_{\Delta T}$, where ΔT is the width of the time window, $\mu^k_{\Delta T}$ and $\sigma^k_{\Delta T}$ are the mean and standard deviation of the k-th subcarrier, respectively. And the channel embedding y is given by $y = \frac{1}{n}\sum_{i=1}^{1}\left|\delta^i\Delta_T/\delta^i\Delta_{T-1}\right|$, where n is the number of subcarriers. If y falls within the experiential threshold range, we

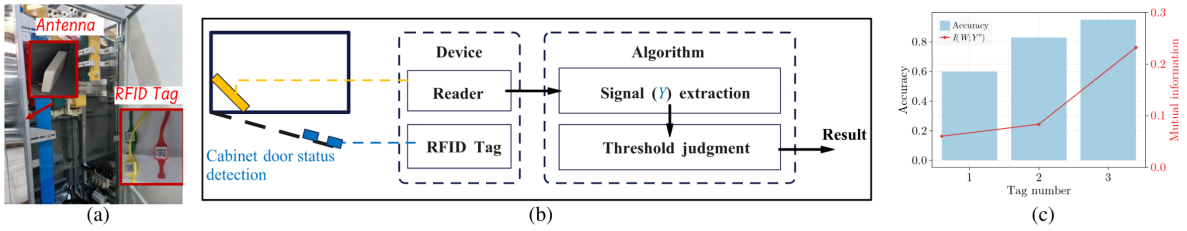

Figure C3 RFID-based electrical cabinet door state monitoring. (a) Schematic diagram of device deployment. (b) The cabinet door status monitoring algorithm. (c) The identification accuracy (represented by the bar chart on the left y-axis) and the mutual information (indicated by the red line on the right y-axis) exhibit a consistent trend of variation.

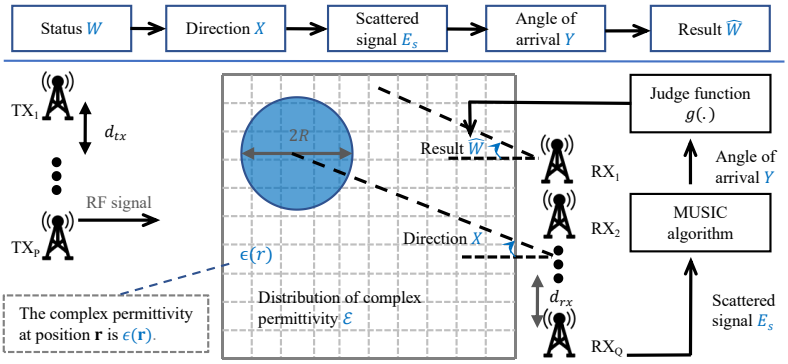

Figure C4 The sensing channel of direction estimation based on Music algorithm and electromagnetic signal. For state W, we use the azimuth angle as a feature (encoding). Then, the channel embedding Y is obtained from the received signal using the MUSIC algorithm, and subsequently, the state is discriminated using the maximum likelihood rule to obtain the sensing result $\hat{W}$.

consider the environment to be "person absent"; otherwise, it is determined to be person present. The entire data processing workflow is illustrated in Fig. C1(b). Here, the threshold range is $[0.935, 1.065]$. Figure C1(c) shows a example of the channel embedding extraction process.

In Fig. C2, the blue solid line illustrates the error rate of human detection as the width of the time window varies ΔT. The dashed lines of other colors represent the mutual information $I(W;\hat{W})$ under different numerical estimation algorithms, namely KraskovStogbauerGrassberger1 [4], KraskovStogbauerGrassberger2 [4], GaoKannanOhViswanath [5], and GaoOhViswanath [6]. The results demonstrate that the trend of accuracy change is highly consistent with the trend of mutual information change, indicating that in such tasks, DTMI can serve as an additional performance metric, complementing accuracy, to evaluate system performance.

(2) **RFID-based electrical cabinet door state monitoring.**

Ensuring electrical safety is crucial during the manufacturing process. Take the electrical cabinet as an example; if its door is inadvertently opened without timely detection, there are potential safety hazards, including the risk of electrical fires and electric shock. In the field of terminal sensing in power systems, electromagnetic transformer-type sensors have traditionally dominated. In recent years, non-electric quantity sensing technologies such as vibration, stroke, arc light, and spectral sensing have gained widespread application in digital electrical equipment and power systems. However, these sensing technologies frequently depend on specialized sensors that boast high sensitivity and accuracy. These sensors are typically burdened with several drawbacks, including complexities in power supply, large size and weight, high energy consumption, vulnerability to electromagnetic interference, difficult installation processes, and exorbitant costs. Consequently, they fall short of meeting the requirements for the development of modern smart power equipment. Given the cost-effectiveness and ease of deployment of RFID tags, we have developed an algorithm for monitoring cabinet door status using multiple tags. Furthermore, we employ the mutual information of tasks, as proposed in this paper, to assess the system's performance.

We conduct relevant tests in a factory setting. For an industrial metal electrical cabinet (measuring approximately $1\,\mathrm{m} \times 1\,\mathrm{m} \times 2\,\mathrm{m}$) used in production, our objective is to monitor the status of the cabinet door. The RFID reader is ImpinJ Speedway R420 reader. The RFID system operates in the $920\,\mathrm{MHz} \sim 926\,\mathrm{MHz}$. Two states W are defined: when the door opening angle is less than $5°$, it is considered "closed"; otherwise, it is deemed "open". We affix several (1 to 3) anti-metal RFID tags onto the cabinet door and positioned the antenna within the cabinet body. The deployment configuration of the equipment is illustrated in Fig. C3(a). After collecting the RSSI (Received Signal Strength Indicator) from each tag, we perform differential processing against an initial value, followed by calculating the average of these differential values across multiple tags. If the average differential exceeds an empirically determined threshold (set here as 2.5), we conclude that the sensing result is "open"; otherwise, it is concluded as "closed". The detailed steps of data processing are depicted in Fig. C3(b).

The results of the state monitoring are shown in Fig. C3(c). Due to the cabinet being made of metal, the electromagnetic waves suffer from severe multipath interference. Consequently, when only one tag is used, the stability of the data is poor, and the empirical threshold becomes almost unusable after the tag position shifts by just a few centimeters. This issue

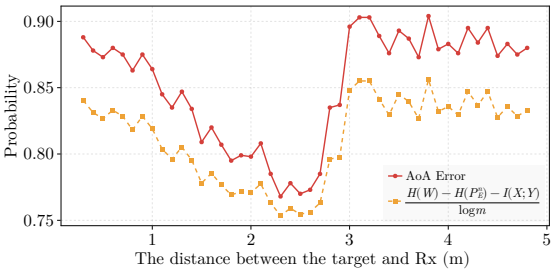

Figure C5 The sensing channel of AoA estimation based on Music algorithm and electromagnetic signal.

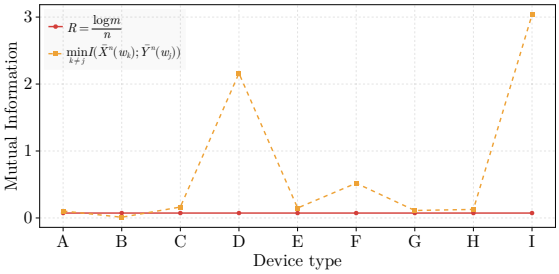

Figure C6 Device type identification based on traffic characteristics.

leads to an identification accuracy of less than 60%. This is well reflected by the mutual information $I(W;Y^n)$ $(n=1)$, which has a small value in this case. Since the spacing of the tags exceeds half a wavelength, their mutual influence is minimal, and thus we can approximately consider the reflection signals from different tags as independent of each other. Consequently, following corollary introduced in Sec. Appendix B.1, as the number of tags increases, so does the mutual information. We employ GaoOhViswanath [6] method to estimate the mutual information, and the red line in Fig. C3(c) illustrates its trend, which increases with the number of tags. As the mutual information increases, so does the accuracy of state identification.

Appendix C.2 Multiple classification tasks.

(1) **Direction estimation based on MUSIC algorithm and electromagnetic signal.**

Location sensing represents one of the most prevalent and fundamental tasks in the sensing field. A plethora of superior systems have been developed utilizing location sensing. Nevertheless, for an extended period, there has been a dearth of methods other than experimental evaluations to assess the influence of numerous factors, including the distance between the target and both the transmitter and receiver, on localization accuracy. In this case study, we use direction estimation based on the Music algorithm (one of the most popular localization algorithms) [7] and electromagnetic signal to show the application of the proposed framework.

We consider a two-dimensional direction estimation problem. The basic model setup is shown in the Fig. C4. There are P transmitting antennas and the position of the p-th transmitting antenna is denoted as $\mathbf{r}_{tx_p}$. The receiver has Q receiving antennas and the position of the q-th receiving antenna is denoted as $\mathbf{r}_{rx_q}$. The distance between two adjacent antennas is d_{rx} and d_{tx} for the receiver and transmitter, respectively. The distribution of complex permittivity in space is $\mathcal{E}$, and the permittivity at position $\mathbf{r}$ is $\mathcal{E}=\epsilon(\mathbf{r})$. For ease of calculation, we set the shape of the target to be a circle with a radius of R. We set m states, each state corresponds to a direction interval. The direction is defined as the angle (the X in Fig. C4) between the line connecting the center of the target circle and the center of the receiving antenna array and the vertical line of the antenna array. The direction interval is $[-\pi,\pi]$, which is evenly divided into m sub-intervals. The scattered signals E_s are calculated using Maxwell's equations and the method of moments [8]. After adding Gaussian random noise to E_s, we estimate channel embedding Y using the MUSIC algorithm. Finally, we use the maximum likelihood algorithm to determine the direction X corresponding to channel embedding Y, and then output the category to which X belongs as the result $\hat{W}$.

We first simulated the effect of the distance between the target and the receiver on the direction estimation accuracy. During the simulation, we set the parameters as follows. We set the number of states $m=9$. The frequency of the electromagnetic signal is 5.0 GHz. The distance between the transmitter and the receiver is 8.0 m. There are $P=1$ transmitting antennas and $Q=3$ receiving antennas. The distance between two adjacent receiving antennas is 0.03 m, i.e., $d_{rx}=0.03$ m. The diameter of the target is $2R=0.2$ m. The distance between the target and the receiver changes from 0.3 m to 5 m. The material of the target is water, and the permittivity is given by empirical formula [9]. In order to solve the scattered waves E_s using the moment method, we discretized the space so that each subunit is a square with a side length of 0.01 m.

We estimate the mutual information using a numerical algorithm [10]. The results are shown in Fig. C5. The results show that when the target is too close to the receiver, the accuracy of the direction estimation is very poor. We believe this is because the existence of phenomena such as diffraction makes it difficult to use the ray tracing model (the basic assumption of the MUSIC algorithm) to equivalent signal transmission [11]. When the distance is too large, the accuracy will also decrease. We believe this is because the scattered wave signal becomes weaker, resulting in a decrease in angular resolution. In addition, the changing trend of accuracy is basically consistent with the changing trend of the error lower bound given by our DTMI. And we used the cor function from Statistics.jl library in Julia language to calculate Pearson coefficient, where one input is mutual information and another input is sensing accuracy. Their Pearson correlation coefficient exceeds 0.95.

(2) **Device type identification based on traffic characteristics.**

Security and privacy issues have always been a hot topic among researchers [12]. In recent years, with the development of the Internet of Things (IoT) and WiFi technology, attackers have devised more diverse means to steal private information. For instance, many attackers place concealed cameras and other IoT devices designed to pilfer private information in public environments such as hotels. After acquiring this private information, these devices continuously transmit the data through gateways. To detect illegal devices, Yan et al. [13] leveraged the characteristic that different devices generate distinct

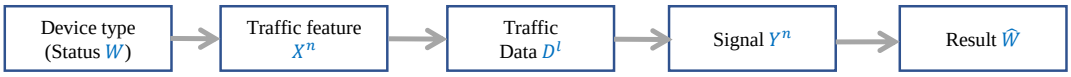

Figure C7 Device type identification based on traffic characteristics. For device type W, we construct feature X^n (encoding) using network traffic. Then, the channel embedding Y^n is obtained from the received signal, and subsequently, the state is discriminated using the maximum likelihood rule to obtain the sensing result $\hat{W}$.

Table C1 The code names for device types and their actual names.

Device name	**Device type**
XIAOMI Bedside Lamp	A
HUAWEI TC5206	B
XIAOMI Induction Cooker	C1, C2
HUAWEI Matebook	D
XIAOMI Microwave Oven	E1, E2
Oneplus6T	F
XIAOMI Rice Cooker	G1, G2
XIAOMI EPS	H1, H2, H3
XIAOMI Table Lamp	I1, I2

traffic patterns, using the traffic at the gateway for device type identification. Their research findings indicated a minimum accuracy rate of 99.17% for identifying common devices like various models of Xiaomi phones, routers, etc. In this paper, based on their open-source code and data, our analysis shows that lossless detection can be achieved when the bit rate satisfies the sufficient condition given in Theorem A3.

At this moment, the schematic diagram illustrating the sensing channel encoder model is depicted in Fig. C7. Post-processing of the traffic data, we employ the methodology put forth by Yan. [13] and colleagues to derive a 30-dimensional signal intended for appliance classification. Our dataset encompasses traffic information from ten distinct device categories, whose precise nomenclature and coding are presented in Table C1. Notably, instances where identical device names are associated with multiple codes signify the existence of several units of the same device category. As an illustration, Type "C" comprises two devices, labeled "C1" and "C2", which denote two separate models of Xiaomi induction stoves. The evaluation procedure incorporates a five-fold cross-validation strategy, alongside adopting the KNN classifier as the analytical tool for discrimination. Throughout every iteration of cross-validation, the signals hailing from the subset earmarked for training are denoted as X^n, whereas those belonging to the testing subset are marked as Y^n, precedented by applying algorithm "GaoOhViswanath" [6] to gauge mutual information. Fig. C6 illustrates the results of our calculations. Here, the possible state number $m = 10$, and 30-dimensional features are used for device type recognition. In this case, the corresponding sensing bitrate is $R = \log m/n$. We find that the data at this time satisfies the sufficient conditions given by Theorem A3, and the goal of non-destructive sensing can be achieved at this time. the DTMI values are mostly above the lower bound given by Theorem A3, and the overall recognition accuracy of KNN has exceeded 99%.

References

1 Te Sun Han, Verdu S. Generalizing the Fano inequality. IEEE Transactions on Information Theory, 1994, 40(4): 1247-1251.
2 Cover T M. Elements of information theory. John Wiley & Sons, 1999.
3 Shannon C E. A mathematical theory of communication.
4 Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Physical Review E, 2004, 69(6): 066138.
5 Gao W H, Kannan S, Oh S, et al. Estimating mutual information for discrete-continuous mixtures.
6 Gao W, Oh S, Viswanath P. Demystifying Fixed k -Nearest Neighbor Information Estimators. IEEE Transactions on Information Theory, 2018, 64(8): 5629-5661.
7 Kotaru M, Joshi K, Bharadia D, et al. SpotFi: Decimeter Level Localization Using WiFi. In: Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, 2015. 269-282.
8 Shang F, Yang P, Yan D, et al. LiquImager: Fine-grained Liquid Identification and Container Imaging System with COTS WiFi Devices. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2024, 8(1): 1-29.
9 Kaatze U. Complex permittivity of water as a function of frequency and temperature. Journal of Chemical & Engineering Data, 1989, 34(4): 371-374.
10 Haaga K A, RomeoV, Datseris G, et al. JuliaDynamics/CausalityTools.jl: V2.10.1. 2023.
11 Shang F, Yang P, Yan Y, et al. LiqRay: Non-invasive and fine-grained liquid recognition system. In: Proceedings of the 28th Annual International Conference on Mobile Computing And Networking, 2022. 296-309.
12 Han F, Yang P, Du H, et al. Accuth: Anti-Spoofing Voice Authentication via Accelerometer. In: Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, 2022. 637-650.
13 Yan D, Yan Y, Yang P, et al. Real-time identification of rogue WiFi connections in the wild. IEEE Internet of Things Journal, 2022, 10(7): 6042-6058.