This letter has been submitted for possible publication at IEEE Wireless Communication Letter (WCL) 2025. This is the authors' version.

PULSE: Physics-Aware Temporal Embedding Learning for Domain Adaptive Wireless Sensing

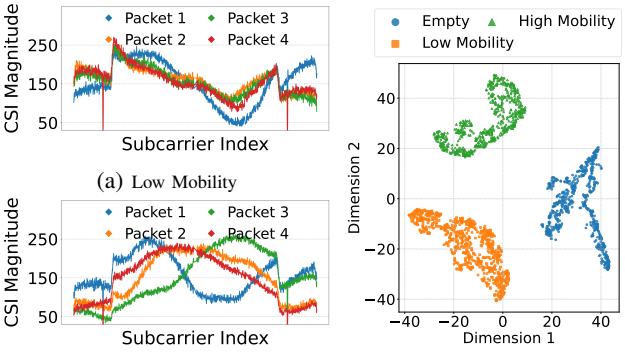
Rifat Zabin, Md. Golam Rabiul Alam

***Abstract*—We present PULSE, a lightweight domain-adaptive sensing framework that learns the physics-aware temporal embeddings, extracted from Wi-Fi channel frequency response (CFR). Unlike the existing approaches where the CFR is directly used as the input the tensor for Learning model, PULSE extracts temporal descriptors and embeds them through a light-weight 1D convolutional network that jointly learns discriminative representations and sensing semantics. The framework achieves over 99% accuracy while reducing 85% of the input tensor dimensionality and maintaining low inference latency, demonstrating suitability for real-time edge inference. Furthermore, through supervised contrastive pretraining and few-shot adaptation, PULSE generalizes effectively to unseen domains using only 5 s worth of labeled data, outperforming the state-of-the-art frameworks. We have extensively evaluated PULSE sesning framework with publicly available dataset of 20 different activities collected over multiple, subjects and propagation environment. For reproducibility, we pledge to share our code base at: https://github.com/rifatzabin/PULSE.**

I. Introduction

Wireless sensing is perceived from the variations in multipath propagation due to the changes in physical world [1]. It is emerging as a non-intrusive and privacy-preserving technology with applications spanning healthcare, smart homes, and entertainment systems. However, real-time inference on resource-constrained edge devices remains challenging due to the high dimensionality of channel frequency response (CFR) data and the computational burden of processing it at fine temporal scales [2]. Moreover, existing lightweight Wi-Fi sensing models—although designed to reduce overhead—often suffer from degraded reliability, especially in dynamic environments where subtle activity variations are easily masked by lossy feature representations [1], [3]. State-of-the-art (SOTA) approach, such as dynamic entropy-based subchannel selection, still rely on full CFR and high-resolution spatial features extracted via CNN and CBAM modules [3], making it compute intensive for edge devices. Meneghello *et al.* proposed a phase sanitization technique leveraging Doppler shifts to build an environment-independent human activity recognition system [4]. However, it is limited to high motion activity and involves only 7 classes. Haque *et al.* developed ATEN, a domain generalization framework integrating TCN to extract long-range temporal dependencies for gesture recognition from mmWave MIMO CFR [5]. However, it is compute intensive because of its long-range temporal dependencies. Similarly, Yang *et al.* transformed CFR into time–frequency spectrograms and achieved domain adaptability through spectrogram augmentation, self-contrastive learning, and adversarial training [6]. Despite recent progress, most wireless sensing models remain environment-dependent, computationally heavy, and poorly interpretable, limiting their real-time adaptability and deployment on edge devices.

R. Zabin, M. G. R. Alam, are with the Department of Computer Science and Engineering, BRAC University, Bangladesh. Corresponding author email: rifat.zabin@g.bracu.ac.bd.



(a) Low Mobility (b) High Mobility (c) Class Embeddings

Fig. 1: Temporal evolution of CFR for different classes of activities.

To address those issues, we investigate the temporal discriminability of wireless signals in performing the senisng tasks. Our intuition is that physics-aware temporal features of the CFR can reveal distinct class-dependent patterns, enabling accurate sensing with lightweight models while reducing input dimensionality and computational cost. We validate this insight through a preliminary analysis of Wi-Fi CFR data for three simple activity classes– empty, high-mobility, and low-mobility through extracting physics-aware temporal descriptors (as detailed in Section]II-A). As presented in Fig. 1a and 1b, The CFR magnitude traces exhibited distinct temporal patterns across activities— low-mobility gestures produced smooth, stable CFR variations, whereas high-mobility activities caused sharp, transient fluctuations. Moreover, when the temporal descriptors are projected into latent space, the resulting t-SNE embedding as presented in Fig. 1c revealed clearly separable clusters corresponding to different activities. This observation highlighted a key insight: *the temporal evolution of CFR alone encodes sufficient discriminative structure to perform sensing tasks including activity recognition, gesture recognition, and gait recognition even without full CFR matrices.*

This motivated the design of PULSE, which leverages the intrinsic temporal structure of wireless channels for sensing. Instead of relying on raw, high-dimensional CFR, PULSE distills physically interpretable temporal features and embeds them into a latent representation space, achieving substantial dimensionality reduction and robustness—ideal for resource-constrained edge devices. A lightweight 1D CNN then learns discriminative temporal embeddings directly from these descriptors, jointly performing feature embedding and prediction for diverse sensing tasks. To ensure adaptability across unseen domains i.e., unseen environment and subjects, PULSE employs a few-shot pretraining strategy combining cross-entropy and supervised contrastive learning to produce transferable embeddings.

Summary of Novel Contributions:

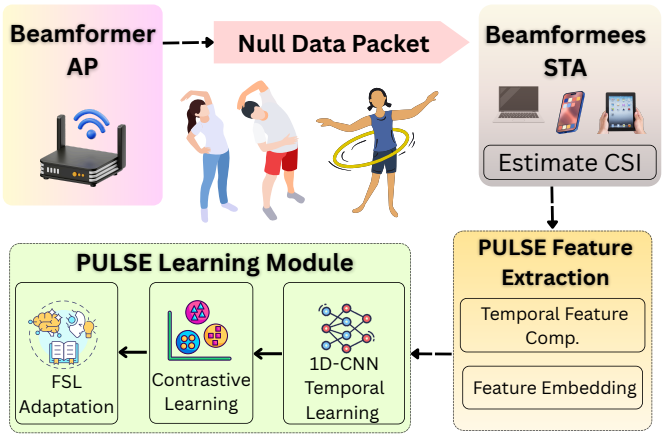


Fig. 2: System Overview of the Proposed PULSE Wireless Sensing Framework

- PULSE introduces a *physics-aware temporal embedding framework* that extracts temporal descriptors pertaining to channel dynamics from CFR, which preserves the discriminative structure while reducing input size by up to 75% for efficient edge sensing.
- PULSE proposes a unified *temporal embedding based learning pipeline*, where a lightweight 1D CNN jointly performs feature embedding and activity inference. Unlike conventional sensing approaches, this joint learning enhances discriminability and minimizes redundancy in temporal representations.
- PULSE integrates a *supervised contrastive pretraining and prototype-based few-shot adaptation* strategy, achieving strong cross-domain generalization from only a few seconds of labeled data.
- PULSE achieves **99.8% accuracy** in recognizing 20 activities within known environments and **95–96%** when tested in unseen domains, improving cross-domain performance by up to **9%** over SOTA approach– *BeamSense* [7].

II. PULSE System Model

In IEEE 802.11ac/ax systems, the multi-antenna transmitter (beamformer) obtains channel knowledge through a *channel sounding* procedure. It periodically transmits *Null Data Packet* (NDP) frames containing known *Long Training Fields* (LTFs), which each receiver (beamformee) uses to estimate the *Channel Frequency Response* (CFR) $\mathbf{H} \in \mathbb{C}^{K\times M\times N}$, where K, M, and N represent the numbers of sub-channels, transmit, and receive antennas, respectively. The CFR captures the frequency-selective propagation between each antenna pair, and after quantization, is fed back to the beamformer for precoding. These estimated CFR values, which encode the temporal and spatial dynamics of the wireless channel, serve as the raw input to subsequent learning task.

Building upon this foundation, PULSE transforms the acquired CFR into compact, interpretable temporal embeddings for efficient and adaptive sensing as illustrated in Fig 2. PULSE operates through three sequential stages: (i) temporal feature extraction from CFR streams, (ii) feature embedding and discriminative learning using a lightweight 1D CNN, and (iii) few-shot domain adaptation for robust and generalizable inference.

A. Temporal Feature Extraction with PULSE

The PULSE framework begins by converting raw *Channel Frequency Response* (CFR) into temporally structured features that describe how the wireless channel evolves over time. Each CFR trace is stored as a complex matrix $\mathbf{H} \in \mathbb{C}^{T\times F}$, where T and F denote the temporal and frequency dimensions, respectively. This matrix captures the instantaneous amplitude and phase response of the wireless channel between transmitter and receiver antennas. However, CFR at this granularity is both high-dimensional and noisy. To make it more suitable for learning, PULSE extracts low-dimensional, physically meaningful features that retain temporal variations linked to human motion or environmental dynamics.

Sub-channel Grouping and Feature Computation Since CFR often exhibits strong frequency correlation, PULSE first divides all the sub-channels, F into k contiguous groups, each having F/k number of component. This band-based grouping provides a trade-off between spectral resolution and robustness, ensuring that small-scale fading effects do not dominate the temporal behavior. Within each group, both amplitude and phase components of the CFR are analyzed through magnitude and phase decomposition. For every time instant t and sub-channel group k, PULSE computes nine temporal features that summarize statistical and temporal changes within that sub-channel group:

$$
\begin{aligned}
&f_1(t) = \text{mean}_{F_k}(|H|), && f_2(t) = \text{std}_{F_k}(|H|),\\
&f_3(t) = \text{median}_{F_k}(|H|), && f_4(t) = p_{75}(|H|) - p_{25}(|H|),\\
&f_5(t) = \Delta_t f_1(t), && f_6(t) = \Delta_t \text{mean}_{F_k}(\angle H),\\
&f_7(t) = \text{std}_{F_k}(\angle H), && f_8(t) = \Delta_t \text{mean}_{F_k}(\angle H),\\
&f_9(t) = \log\big(\varepsilon + \text{mean}_{F_k}(|H|^2)\big)
\end{aligned}
\tag{1}
$$

where $\Delta_t(\cdot)$ denotes the first-order difference between consecutive time samples and $f_9(t)$ is log-energy emphasizing instantaneous signal strength variation, where ε avoids numerical underflow. Collectively, these statistics capture both the slow-varying amplitude envelope and the fine phase fluctuations that occur when human motion or object displacement alters the propagation path. When concatenated across all k sub-channel groups, the resulting feature dimension becomes $C = 9k$, producing a compact yet expressive temporal feature matrix $\mathbf{F} \in \mathbb{R}^{C\times T}$.

Temporal Window Segmentation To align with learning frameworks, the continuous feature stream is partitioned into smaller temporal windows of fixed length W and stride S. Each segment $\mathbf{X}_i \in \mathbb{R}^{C\times W}$ corresponds to a short-duration snapshot of channel behavior, effectively converting long CFR recordings into multiple training samples that capture localized motion patterns. Using non-overlapping or partially overlapping windows allows PULSE to control redundancy and temporal resolution depending on the dynamics of the activity. This transformation also leads to a significant reduction in input dimensionality compared to the raw CFR representation. While the original CFR tensor $\mathbf{H} \in \mathbb{C}^{T\times F}$ contains high-resolution complex responses over all F sub-channels, the extracted feature tensor $\mathbf{F} \in \mathbb{R}^{C\times T}$ condenses this information

into a compact set of physically meaningful descriptors of dimension $C = k \times d$, with k denoting the number of sub-channel groups and d the number of temporal–statistical features per group. Since $(k \times d) \ll 2F$, this conversion typically can achieve a dimensionality reduction of approximately 65–85%, depending on k and the feature configuration. Such compression retains the most informative temporal variations while discarding redundant spectral correlations, allowing PULSE to remain computationally lightweight and memory-efficient for real-time inference.

Normalization and Output After segmentation, PULSE computes per-channel normalization parameters

$$\mu_c = \frac{1}{NW}\sum_{i,t} X_{i,c,t}, \qquad \sigma_c = \sqrt{\frac{1}{NW}\sum_{i,t}(X_{i,c,t}-\mu_c)^2}, \tag{2}$$

ensuring that all feature channels contribute equally during model training and preventing amplitude-dominant bands from biasing the learning process. Finally, the processed data are saved as tensors $\mathbf{X} \in \mathbb{R}^{N\times C\times W}$ with corresponding class labels $\mathbf{y} \in \mathbb{N}^N$, normalization parameters $\mu, \sigma \in \mathbb{R}^C$, and a class-to-ID mapping file for reproducibility.

B. Feature Embedding and Activity Learning with 1D CNN

Following temporal feature extraction, PULSE derives embeddings from the extracted features and utilizes a lightweight 1D CNN to perform end-to-end supervised learning for classification task within each environment. The model simultaneously learns discriminative feature embeddings and the activity decision boundaries from normalized temporal windows $\mathbf{X} \in \mathbb{R}^{N\times C\times W}$ representing N samples, C no. of temporal features, and W temporal windows. The dataset is divided into training, validation, and test subsets using stratified partitioning to preserve class balance. Normalization parameters are estimated only from the training set and applied to all splits to ensure statistical consistency: $\boldsymbol{\mu} = \mathrm{mean}_{n,t}(\mathbf{X}_{\text{train}})$, and $\boldsymbol{\sigma} = \mathrm{std}_{n,t}(\mathbf{X}_{\text{train}})$. To enhance robustness against temporal jitter and small channel fluctuations, each input window undergoes mild augmentations, including circular time shifts and Gaussian perturbations $\mathcal{N}(0, \sigma^2_{\text{noise}})$. These preserve the temporal structure of CFR while improving invariance to phase drift and environmental variations.

The CNN backbone $f_\theta(\cdot)$ consists of multiple temporal convolutional blocks with residual connections and dropout regularization. It produces both a compact temporal embedding $\mathbf{h}$ and a class prediction $\mathbf{z}$: $\mathbf{z}, \mathbf{h} = f_\theta(\widetilde{\mathbf{x}}), \qquad \mathbf{z} \in \mathbb{R}^M$. Here, $\mathbf{h}$ captures high-level temporal dynamics, while $\mathbf{z}$ encodes class-level logits for M classes. The model is trained using a supervised classification loss. By default, cross-entropy with label smoothing ε is used: $\mathcal{L}_{\text{CE}} = -\sum_{k=1}^{K} q_k \log p_k, p_k = \frac{e^{z_k}}{\sum_j e^{z_j}}$ where q_k denotes the smoothed target distribution. Alternatively, focal loss can be applied to emphasize hard samples: $\mathcal{L}_{\text{Focal}} = \left(1 - e^{-\mathrm{CE}(\mathbf{z},y)}\right)^{\gamma}\mathrm{CE}(\mathbf{z}, y)$ with focusing parameter $\gamma > 0$.

Training proceeds until the validation loss ceases to improve for a predefined patience interval, and the best-performing weights are restored for final evaluation.

C. Few-Shot Adaptation for Domain Generalization

To generalize effectively from a trained source environment to a new target environment with limited supervision, PULSE employs a two-stage adaptation strategy that leverages reusable embeddings and non-parametric classification for low-data transfer.

Source Pretraining: A temporal 1D CNN encoder with attention pooling is pretrained on the source-domain data to learn transferable temporal embeddings. Each normalized input window $\widetilde{\mathbf{x}} \in \mathbb{R}^{C\times W}$ is transformed into a compact D-dimensional embedding $\mathbf{e} \in \mathbb{R}^D$, where D denotes the latent embedding dimension learned by the encoder. The encoder is optimized using a joint objective that integrates cross-entropy (CE) and supervised contrastive (SupCon) losses, designed to enhance class discrimination while preserving temporal coherence. The total loss function is expressed as $\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{CE}} + \lambda_{\text{sc}}\,\mathcal{L}_{\text{SupCon}}$, where:

$$\mathcal{L}_{\text{SupCon}} = -\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{1}{|\mathcal{P}(i)|}\sum_{p\in\mathcal{P}(i)}\log\frac{\exp\big(\langle \mathbf{z}_i, \mathbf{z}_p\rangle/\tau\big)}{\sum_{a\in\mathcal{A}(i)}\exp\big(\langle \mathbf{z}_i, \mathbf{z}_a\rangle/\tau\big)}. \tag{3}$$

where $\mathbf{z}$ denotes the normalized projection of $\mathbf{e}$, $\mathcal{P}(i)$ and $\mathcal{A}(i)$ represent the positive and all anchor samples for instance i, and sim is the cosine similarity. This pretraining process aligns temporally correlated embeddings and reinforces inter-class boundaries, producing a robust feature space suitable for downstream few-shot adaptation.

PULSE Few-Shot Adaptation: When applied to a new environment, only a few labeled samples per class are required to perform domain alignment. The target data $\mathbf{X}$ are split into a small K-shot support set $\mathcal{S}$ and a disjoint query set $\mathcal{Q}$. To compensate for statistical shift between domains, PULSE employs a normalization blending strategy:

$$\boldsymbol{\mu} = (1-\beta)\boldsymbol{\mu}_{\text{src}} + \beta\boldsymbol{\mu}_{\text{sup}}, \qquad \boldsymbol{\sigma} = (1-\beta)\boldsymbol{\sigma}_{\text{src}} + \beta\boldsymbol{\sigma}_{\text{sup}}, \tag{4}$$

where $\beta \in [0, 1]$ controls the contribution of source and target statistics. All support and query samples are passed through the pretrained encoder to obtain ℓ_2-normalized embeddings. For each class c, a prototype $\mathbf{P}_c$ is computed as the mean of its support embeddings:

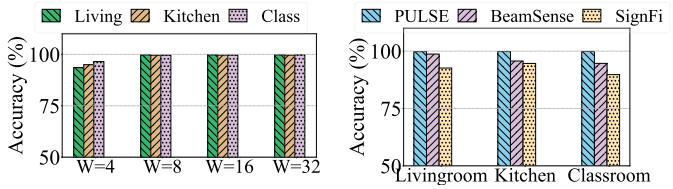
$$\mathbf{P}_c = \frac{\left(\frac{1}{|\mathcal{S}_c|}\sum_{(\mathbf{x},y)\in\mathcal{S}_c}\mathbf{e}(\mathbf{x})\right)}{\left\|\frac{1}{|\mathcal{S}_c|}\sum_{(\mathbf{x},y)\in\mathcal{S}_c}\mathbf{e}(\mathbf{x})\right\|_2}, \tag{5}$$

and query samples are classified by cosine similarity:

$$\hat{y}(\mathbf{x}_q) = \arg\max_c \left\langle \frac{\mathbf{e}(\mathbf{x}_q)}{\|\mathbf{e}(\mathbf{x}_q)\|_2},\ \mathbf{P}_c\right\rangle. \tag{6}$$

This nearest-prototype classification in embedding space allows non-parametric adaptation with minimal overhead and high stability under small K.

Through this adaptation mechanism, PULSE effectively aligns temporal embeddings between source and target domains, achieving robust performance with as little as 5 s worth of labeled data per class. The same CNN encoder serves both pretraining and adaptation, requiring no architectural



(a) Impact of temporal window length on PULSE accuracy (b) Performance comparison of PULSE with baseline models

Fig. 3: PULSE Accuracy analysis under varying temporal window W and baseline comparison using the selected configuration $W = 8$

modification while maintaining full interpretability and low computational complexity.

III. Experimental Dataset

The experiments in this study are conducted using the publicly available *BeamSense* dataset [7], which provides a large-scale, real-world collection of Wi-Fi Channel State Information (CFR) data for human activity sensing. The dataset was recorded in three distinct indoor environments—*kitchen*, *living room*, and *classroom*—to capture diverse propagation conditions with varying furniture, obstacles, and layouts. Three spatially separated Wi-Fi stations (STAs), denoted as *STA1*, *STA2*, and *STA3*, were deployed at different positions within each environment to emulate spatial diversity in sensing. Each STA captured CFR data corresponding to twenty human activities, encompassing both stationary and dynamic motions- *jogging, clapping, push forward, boxing, writing, brushing teeth, rotating, standing, eating, reading a book, waiving, walking, browsing phone, drinking, hands-up-down, phone call, side bend, check the wrist watch, washing hands and browsing laptop*, each lasting 300 s per session. The experimental setup followed the IEEE 802.11ac MU-MIMO standard, operating on channel 153 at 5.77 GHz with 80 MHz bandwidth, the system continuously captured *Channel Frequency Response* (CFR) from each STA through modified Nexmon-enabled network interface cards. The CFR measurements across each STA spanned 242 sub-channels, each representing the complex channel response across one transmit and one receive antenna links, forming a $1 \times 1 \times 242$ MIMO CFR matrix per frame [7].

For performance benchmarking, PULSE was evaluated alongside three state-of-the-art baselines: (i) *BeamSense* [7], a MU-MIMO based sensing system with few-shot adaptation; (ii) *SignFi* [8], a CFR-based sign gesture recognition framework; and (iii) *OneFi* [9], a cross-domain one-shot learning approach for gesture recognition. The dataset's extensive coverage, and multi-environment diversity, make it a robust benchmark for validating the temporal feature extraction and domain adaptation capabilities of PULSE.

IV. Performance Analysis of PULSE Framework

A. Performance of PULSE with Different Temporal Windows

In wireless sensing, the temporal window length (W) determines how many consecutive CFR frames are aggregated to capture motion dynamics. In PULSE, the windowing process transforms sequential CFR streams into fixed-length temporal segments, allowing the model to learn discriminative temporal embeddings within each segment.

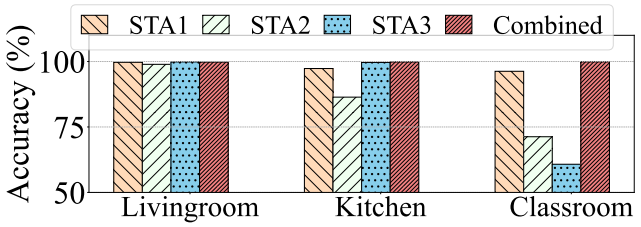


Fig. 4: PULSE accuracy comparison of individual against combined Wi-Fi stations across three indoor environments

We evaluate the impact of window size by varying $W = \{4, 8, 16, 32\}$ and report the model accuracy across three indoor environments—living room, kitchen, and classroom—as shown in Fig. 3a. The accuracy improves notably when increasing the window size from $W = 4$ to $W = 8$, improving around 6.16% in the living room, and 4.64% and 3.16% in the kitchen, and the classroom respectively, indicating that moderate temporal context enhances feature stability and discriminability. However, beyond $W = 8$, the performance gain becomes marginal (e.g., 99.72%–99.76%) while introducing higher inference latency. Therefore, $W = 8$ is selected as the optimal temporal window, achieving near-saturation accuracy with minimal delay, making it suitable for real-time edge sensing.

B. PULSE vs State-of-the-Art Approaches

We further benchmark PULSE against two established baselines—*BeamSense* and *SignFi*—across three indoor environments, as shown in Fig. 3b. PULSE achieves superior recognition accuracy of 99.78%, 99.82%, and 99.86% in the living room, kitchen, and classroom, respectively. In contrast, BeamSense records 98.72%, 95.69%, and 94.67%, while SignFi yields 92.64%, 94.64%, and 89.76% under identical settings. The results highlight that PULSE consistently outperforms both baselines, achieving up to 5.2% accuracy improvement over BeamSense and around 10% over SignFi. This gain stems from PULSE's ability to extract temporal embeddings from band-aware CFR features, capturing motion dynamics more effectively than static channel-based representations.

C. Spatial Diversity Analysis of Sensing Stations

As depicted in Fig 4, PULSE demonstrates superior performance when leveraging fused data from multiple Wi-Fi stations. Even if an individual station fails to sustain high accuracy—for example, STA2 and STA3 achieve only 86.39% and 60.76% in the Kitchen and Classroom— PULSE framework consistently extracts near-perfect accuracies of 99.82% and 99.86% from the combined signals. This robustness highlights PULSE's ability to intelligently integrate spatially diverse information, and effectively mitigating the limitations imposed by orientation of individual Wi-Fi stations. Consequently, even under unfavorable spatial configurations, PULSE ensures reliable and environment-invariant sensing performance.

D. PULSE Performance in Unseen Environment and Stations

As presented in Fig. 5, we evaluate the domain generalization performance of PULSE with data from unseen envi-

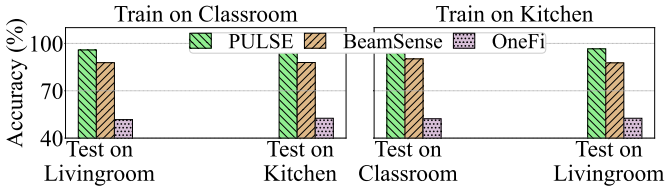


Fig. 5: Comparison of PULSE performance across environments with baseline BeamSense, and SignFi.

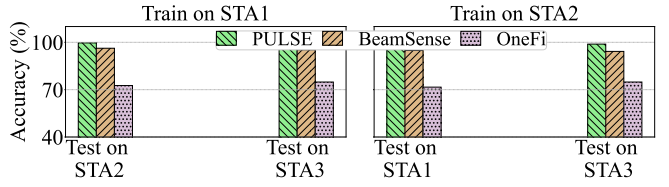


Fig. 6: Comparison of PULSE performance across Wi-Fi stations with baseline BeamSense, and SignFi.

ronment and sensing station. When trained on the classroom dataset and tested on unseen dataset of living room and kitchen, PULSE achieves an average accuracy of 95.84% and 95.05% respectively, outperforming BeamSense and OneFi by over 7.1%, and 40% respectively. Similarly, training on kitchen dataset and testing on the data from other unseen environments, yield similar accuracy, improving the performance of PULSE by up to 9% over its nearest competitor BeamSense. PULSE also demonstrates strong cross-station generalization performance following the similar trend, as depicted in Fig. 6.

Whereas, PULSE used contrastive learning with few-shots leveraging only 5s worth of data from unseen environment, BeamSense takes 15s worth of data per activity to fine-tune to new environment. Thus, PULSE not only generalizes effectively across unseen domains but also achieves breakthrough sensing performance under extreme data constraints.

E. Computational Efficiency and Model Compactness

Table I compares PULSE with state-of-the-art models– *BeamSense* and *SignFi*, highlighting its drastic reduction in computational burden. While BeamSense requires nearly 20 B FLOPs and SignFi about 1.7 B, PULSE performs the same sensing task with only 22.41 M FLOPs– over 900× and 75× fewer operations, respectively. PULSE also maintains a competitive parameter count (1.47 M) and achieves an exceptionally low inference latency of 0.012 ms/sample, compared to 22 ms for *BeamSense* and 0.62 ms for *SignFi*.

As presented in Table II, under varying window, FLOPs scale linearly with temporal window length ranging from 22.41 M—179.26 M for window length $W = 8$—64, while the inference latency remains below 0.025 ms. Contrarily, increasing the sub-channel group size ($k = 4$–32) adds negligible computational cost ranging from 21.89 M− − −25.50 M FLOPs. Thus, the configuration of $W = 8$ and $k = 8$ offers the best trade-off between accuracy, and computational burden, while taking only 0.012 ms of inference time.

Note that, in this analysis with $k = 8$, and number of sub-channels $F = 242$, the feature space is reduced from $W \times F \times 2$ for CFR based approach, to $W \times C = 8 \times 72$ for PULSE, resulting in an 85.12% reduction in dimensionality.

TABLE I: Analysis of computational efficiency of PULSE

Approach	FLOPs(M)	Params(M)	Inference (ms/samp)
PULSE	22.41	1.47	0.012
BeamSense	20000	4.25	22.0
SignFi	1700	2.10	0.62

TABLE II: Computational Efficiency of PULSE Under Varying Window and sub-channel Configurations

Window W ($K = 8$)	**FLOPs(M)**	**Params(M)**	**Inference (ms/samp)**
8	22.41	1.47	0.012
16	44.82	1.47	0.012
32	89.63	1.47	0.014
64	179.26	1.47	0.023
Sub-channel Group k ($W = 8$)	**FLOPs(M)**	**Params(M)**	**Inference (ms/samp)**
4	21.89	1.43	0.011
8	22.41	1.47	0.012
16	23.44	1.53	0.012
32	25.50	1.66	0.012

V. Conclusion

This letter introduced PULSE, a lightweight and domain-adaptive wireless sensing framework that transforms complex CFR streams into compact, interpretable temporal representations. By extracting physics-aware temporal features and leveraging contrastive few-shot learning, the framework achieves robust domain generalization across unseen environment and stations with only 5s of data from unseen domain. Experimental evaluations on the BeamSense dataset demonstrate that PULSE attains up to 96% of accuracy in cross-domain setup which is 9% higher than the state-of-the-art approaches. These results validate PULSE as a practical step toward real-time, generalizable framework, enabling efficient and scalable deployment of wireless sensing at the network edge.

References

[1] C. Tian, Y. Tian, X. Wang, Y. H. Kho, Z. Zhong, W. Li, and B. Xiao, "Human activity recognition with commercial wifi signals," *IEEE Access*, vol. 10, pp. 121 580–121 589, 2022.

[2] V. K. Singh, A. Walecha, A. Gera, R. Jay, A. Bhattacharya, and M. Maity, "Slim-sense: A resource efficient wifi sensing framework towards integrated sensing and communication," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 9, no. 1, pp. 1–33, 2025.

[3] B. Li, X. Jiang, Y. Du, Y. Yu, and R. Zhang, "Wi-ssr: Wi-fi based lightweight high-resolution model for human activity recognition," *IEEE Sensors Journal*, 2025.

[4] F. Meneghello, D. Garlisi, N. Dal Fabbro, I. Tinnirello, and M. Rossi, "SHARP: Environment and Person Independent Activity Recognition with Commodity IEEE 802.11 Access Points," *IEEE Transactions on Mobile Computing*, pp. 1–16, 2022.

[5] K. F. Haque, K. Rumman, A. Elyasi, F. Meneghello, and F. Restuccia, "Magic: Meta-learning adaptive gesture recognition with mmwave mimo csi," in *2025 IEEE 26th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2025, pp. 131–140.

[6] W. Yang, Z. Li, and S. Chen, "Environment independent gait recognition based on wi-fi signals," *IEEE Transactions on Mobile Computing*, 2025.

[7] K. F. Haque, M. Zhang, F. Meneghello, and F. Restuccia, "BeamSense: Rethinking Wireless Sensing with MU-MIMO Wi-Fi Beamforming Feedback," *Computer Networks*, vol. 258, p. 111020, 2025.

[8] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 1, Mar. 2018. [Online]. Available: https://doi.org/10.1145/3191755

[9] R. Xiao, J. Liu, J. Han, and K. Ren, "Onefi: One-shot recognition for unseen gesture via cots wifi," in *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 2021, pp. 206–219.