EQUIRFF: RADIO FREQUENCY FINGERPRINT IDENTIFICATION VIA PHYSICAL EQUIVARIANCE GUIDED REPRESENTATION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Radio frequency fingerprint identification (RFFI) seeks to preserve device-specific hardware cues under changing acquisition conditions, yet these cues are often entangled with domain factors in radio frequency (RF) signals. Overly strong invariance may suppress useful structural cues, while purely supervised models can exploit source-domain shortcuts. We propose EquiRFF, a two-stage framework that learns stable device identity from physical equivariance without explicit nuisance-factor separation. Stage I learns how structured physical transformations act in latent space and enforces consistency between sequential and composed transformations. Stage II extracts device-discriminative information that remains stable across these equivariant representations while retaining transformation information in the encoder. Physical views and condition parameters are used only during training. At inference, EquiRFF requires only a single I/Q observation, without physical-condition inputs or target-domain adaptation. Experiments on four public RF datasets show strong cross-domain performance against recent strong baselines.

1 INTRODUCTION

Radio frequency fingerprint identification (RFFI) identifies wireless devices from device-specific distortions introduced by transmitter hardware imperfections (Brik et al., 2008; Danev et al., 2012; Sankhe et al., 2019). Such hardware-induced distortions have long served as the basis for both classical and learning-based device identification across different wireless systems (Shen et al., 2021; Li et al., 2024; Zhang et al., 2025a). In practice, however, the received signal also reflects channel, receiver, and acquisition conditions, causing device-specific hardware cues to become entangled with domain-dependent variation (Zhang et al., 2025a). Simply suppressing this variation may remove information that remains useful for device identification, while purely supervised models can overfit shortcut features in the source domain (Wang et al., 2025). We therefore formulate cross-domain RFFI as a representation learning problem and propose EquiRFF, a two-stage framework that learns stable device identity from physical equivariance.

Rather than removing physical variation from the outset, we first model how it changes the latent representation. EquiMod (Devillers & Lefort, 2023) incorporates transformation parameters into representation learning, while V-JEPA 2-AC (Assran et al., 2025) conditions latent prediction on actions. CARE (Gupta et al., 2024) and STL (Yu et al., 2024) likewise preserve transformation-dependent structure rather than enforcing complete invariance. In radio frequency (RF) signals, physical transformations such as carrier phase offset, carrier frequency offset, and I/Q imbalance follow well-defined signal-domain structures (Huang et al., 2024; Sheng et al., 2025; Yao et al., 2024), but their effects in observed signals remain entangled with channel effects and device-specific hardware characteristics. EquiRFF therefore encodes this known structure before action-conditioned latent prediction and further enforces composition consistency so that sequential and composed transformations induce coherent latent transitions.

After learning how physical transformations affect the representation, the remaining challenge is to extract information that supports stable device identification across these changes. Directly aligning encoder representations may suppress both equivariant structure and condition-dependent cues that

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

remain useful for discrimination. We therefore introduce an RFF attention query head that uses a learnable query to extract a view-invariant and device-discriminative RFF representation from physically equivariant encoder features. A relative transformation recovery objective further encourages the encoder to retain the physical transformation structure learned in Stage I. Physical views and transformation parameters are used only during training. At inference, EquiRFF requires only a single I/Q observation, without physical-condition inputs, additional views, or target-domain adaptation. Our main contributions are summarized as follows:

- We introduce physics-structured action encoding and composition consistency to model how physical transformations interact with RF signal representations in latent space.
- We use an RFF attention query head to extract view-invariant and device-discriminative RFF representations from physically equivariant encoder features.
- EquiRFF supports source-only cross-domain identification from a single I/Q observation, without predefined nuisance-factor separation or target-domain adaptation.

2 Related Work

2.1 Transformation-Aware Representation Learning

Multi-view self-supervised learning commonly learns invariance by encouraging agreement between transformed views of the same sample. Methods such as BYOL (Grill et al., 2020) have shown the effectiveness of this paradigm, while later transformation-aware approaches including EquiMod (Devillers & Lefort, 2023), CARE (Gupta et al., 2024), STL (Yu et al., 2024), and SER (Lee et al., 2026) preserve transformation-dependent structure rather than removing it entirely. This line of work establishes that useful representations need not be fully invariant to view changes.

Predictive representation learning further models transformations directly in latent space. I-JEPA (Assran et al., 2023) establishes joint-embedding prediction without reconstructing the input. V-JEPA 2-AC (Assran et al., 2025) and seq-JEPA (Ghaemi et al., 2025) extend this idea by conditioning prediction on actions or relative transformations. Their action representations are learned without explicitly encoding the known mathematical structure of the RF transformations considered here. Their objectives also do not explicitly constrain successive latent transitions to follow the corresponding physical composition laws. EquiRFF addresses these two aspects through physics-structured action encoding and composition consistency along physically valid same-factor transformation trajectories.

2.2 Cross-Domain Radio Frequency Fingerprint Identification

Cross-domain RFFI has been studied through source-only generalization and methods that adapt using target-domain observations. RIEI (Zhang et al., 2025b) separates emitter-related and receiver-related information, while Avoiding Shortcuts (Wang et al., 2025) reduces reliance on channel shortcuts through source-domain augmentation and representation alignment. Both operate without target-domain data but introduce explicit assumptions about the domain-related variation addressed during training.

Distance-Invariant RFFI (Huang et al., 2026) and RFF-TTA (Li et al., 2026) instead use target-domain observations to adapt the source-trained model to deployment conditions. EquiRFF remains source-only and learns stable device representations from structured physical variation without predefined nuisance-factor separation or target-domain adaptation.

3 Method

3.1 Overview

EquiRFF is a two-stage representation learning framework for source-only cross-domain radio frequency fingerprint identification. Our main design principle is to separate the modeling of physical variation from the extraction of device identity. Rather than forcing RF observations collected under different physical conditions to share the same representation from the outset, EquiRFF first

learns how structured physical transformations change the latent representation and then extracts the device-discriminative information that remains stable across these physically equivariant representations.

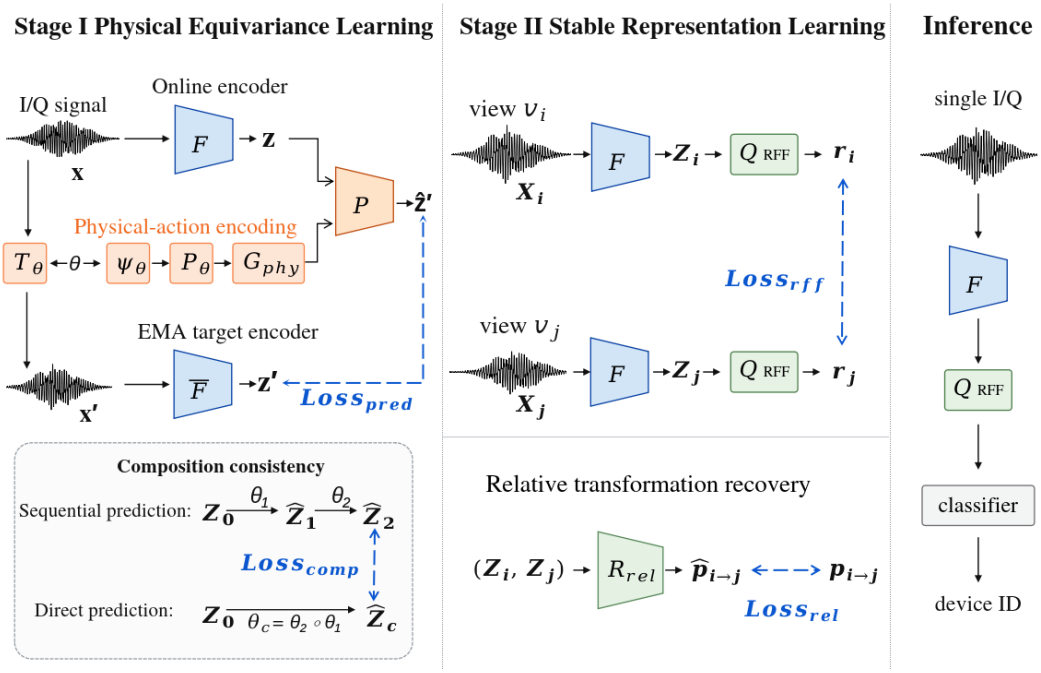


Figure 1: Overview of EquiRFF. Stage I learns physically equivariant latent dynamics through physics-structured action conditioning and composition consistency. Stage II learns a stable RFF representation from physically equivariant features while preserving transformation information through relative transformation recovery. At inference, only a single I/Q observation is required.

In Stage I, shown in the left panel of Figure 1, known RF transformation structure is encoded into action representations that condition latent prediction. Composition consistency further constrains successive transformations of the same physical factor and their equivalent composed transformation. In Stage II, shown in the middle panel, an RFF attention query head extracts a stable RFF representation from physically equivariant encoder features. The identity objective is applied to these RFF readouts rather than directly aligning the high-dimensional encoder representations across physical views. Relative transformation recovery further encourages the shared encoder to retain transformation information. The resulting RFF representation is view-invariant and device-discriminative. The right panel summarizes the single-observation inference pipeline.

To formalize this two-stage framework, let $x^{(0)} = x \in \mathbb{C}^N$ denote a length-N complex I/Q observation with device label y. Starting from $x^{(0)}$, we construct a trajectory of controlled physical actions,

$$x^{(k)} = \mathcal{T}_{\theta_k}\left(x^{(k-1)}\right), \qquad k = 1, \ldots, n, \tag{1}$$

where $\mathcal{T}_{\theta_k}$ denotes the signal-domain physical transformation parameterized by action θ_k. For composition supervision, all actions in a trajectory correspond to successive transformations of the same physical factor, while the other factors remain at the identity. The same physical endpoint can therefore be reached directly through the composed action $\Theta = \theta_n \circ \cdots \circ \theta_1$, where $\circ$ denotes transformation composition, yielding $x^{(c)} = \mathcal{T}_\Theta(x^{(0)})$. The original observation, intermediate states, and directly composed endpoint form the training views, while the sequential and direct paths provide the physical relation used for composition consistency. These physical views and transformation parameters are used only during training. At inference, EquiRFF requires only a single I/Q observation, which is processed by the encoder, RFF query head, and classifier.

3.2 Stage I: Physics-Structured Equivariant Dynamics Learning

Stage I learns physically equivariant latent dynamics through structured one-step prediction and multi-step composition consistency.

3.2.1 Structured Action Encoding and One-Step Latent Prediction

Changes in acquisition, propagation, and RF front-end conditions induce structured variations in I/Q signals. We consider carrier phase offset (CPO) (Huang et al., 2024; Kumar et al., 2025), carrier frequency offset (CFO) (Sheng et al., 2025; Gu et al., 2026), and I/Q imbalance (IQI) (Yao et al., 2024; Masoumi & Myers, 2026), which correspond to global phase rotation, progressive phase change, and linear I/Q distortion, respectively.

A straightforward conditioning scheme feeds the raw physical parameters $\theta = (\phi, \epsilon_f, \eta)$ directly into an MLP, where ϕ denotes carrier phase offset, $\epsilon_f = \Delta f / F_s$ the normalized carrier frequency offset, and η the I/Q imbalance parameters. This representation does not explicitly reflect the different transformation structures of these physical quantities. We therefore encode them as

$$p_\theta = \Psi_{\mathrm{RF}}(\theta), \tag{2}$$

where Ψ_{RF} is the physics-structured action encoding function and p_θ is the resulting physics-structured action representation.

CPO. CPO applies a periodic global rotation to the I/Q signal. We therefore encode the phase offset ϕ using identity-centered circular coordinates,

$$\Psi_{\mathrm{CPO}}(\phi) = [\cos\phi - 1,\ \sin\phi], \tag{3}$$

which preserve phase periodicity and map the identity action $\phi = 0$ to zero.

CFO. CFO induces a phase progression along the sampling sequence. For normalized offset $\epsilon_f = \Delta f / F_s$, the per-sample phase increment is $\omega_f = 2\pi\epsilon_f$, which we encode as

$$\Psi_{\mathrm{CFO}}(\epsilon_f) = [\cos\omega_f - 1,\ \sin\omega_f]. \tag{4}$$

This representation preserves the periodic structure of the phase increment governing signal evolution.

IQI. IQI jointly changes the amplitude and phase relationship between the I and Q components and is represented by a linear transformation $A(\eta)$ on the I/Q plane. We encode its deviation from the identity transformation as

$$\Psi_{\mathrm{IQI}}(\eta) = \mathrm{vec}\left(A(\eta) - I_2\right), \tag{5}$$

where I_2 is the 2×2 identity matrix and $\mathrm{vec}(\cdot)$ vectorizes the resulting matrix.

The three structured coordinates are concatenated as

$$p_\theta = [\Psi_{\mathrm{CPO}}(\phi), \Psi_{\mathrm{CFO}}(\epsilon_f), \Psi_{\mathrm{IQI}}(\eta)]. \tag{6}$$

We then use a lightweight MLP G_{phy} to project and expand p_θ for interaction with the signal latent representation. The one-step predictor can use the joint descriptor of CPO, CFO, and IQI, whereas composition consistency is applied only to same-factor trajectories as described below.

With the structured action representation established, we first model the latent transition induced by a single physical action. We use an online encoder F_ω and an EMA target encoder $F_{\bar{\omega}}$, whose parameters $\bar{\omega}$ are updated as an exponential moving average of ω. For a transition from $x^{(i)}$ to $x^{(j)}$,

$$Z_i = F_\omega(x^{(i)}), \qquad Z_j^{\mathrm{tgt}} = \mathrm{sg}\left(F_{\bar{\omega}}(x^{(j)})\right), \tag{7}$$

where Z_i and Z_j^{tgt} denote the corresponding online and target latent representations, and $\mathrm{sg}(\cdot)$ denotes stop-gradient.

The relative physical action is encoded as $p_{i\to j} = \Psi_{\mathrm{RF}}(\theta_{i\to j})$. An action-conditioned latent predictor P then estimates the latent representation after this transformation,

$$\widehat{Z}_j = P\left(Z_i, G_{\mathrm{phy}}\left(p_{i\to j}\right)\right), \tag{8}$$

where $\widehat{Z}_j$ denotes the predicted target latent representation.

The prediction objective is

$$\mathcal{L}_{\text{pred}} = d_z\left(\widehat{Z}_j, Z_j^{\text{tgt}}\right), \tag{9}$$

where $d_z(\cdot,\cdot)$ is the squared Euclidean distance between ℓ_2-normalized token representations, averaged over tokens and samples in the batch. Rather than directly aligning the two views, the predictor learns the one-step latent transition induced by the physical action. This transition model is then used for multi-step prediction.

3.2.2 Multi-Step Compositional Consistency

The one-step predictor models individual physical transitions, but fitting them independently does not guarantee coherent dynamics under successive transformations. For each composition trajectory, we successively vary only one physical factor and keep the other factors at the identity. The composed action therefore follows phase addition for CPO, frequency-offset addition for CFO, and ordered matrix multiplication for IQI:

$$\phi_\Theta = \text{wrap}\left(\sum_{k=1}^{n}\phi_k\right), \qquad \epsilon_\Theta = \sum_{k=1}^{n}\epsilon_k, \qquad A_\Theta = A_n \cdots A_1. \tag{10}$$

Here, $\text{wrap}(\cdot)$ reduces phase modulo 2π. For IQI, the composed operator is encoded directly as $\text{vec}(A_\Theta - I_2)$, preserving the ordered matrix product within the same four descriptor coordinates. Under this construction, the sequential and directly composed transformations reach the same signal endpoint.

Starting from $Z_0 = F_\omega(x^{(0)})$, repeated application of the one-step predictor gives

$$\widehat{Z}^{(0)} = Z_0, \qquad \widehat{Z}^{(k)} = P\left(\widehat{Z}^{(k-1)}, G_{\text{phy}}\left(p_{\theta_k}\right)\right), \qquad k = 1, \ldots, n. \tag{11}$$

The corresponding same-factor actions are composed into $\Theta = \theta_n \circ \cdots \circ \theta_1$ and applied directly from the same starting representation,

$$\widehat{Z}^{(\text{c})} = P\left(Z_0, G_{\text{phy}}\left(\Psi_{\text{RF}}(\Theta)\right)\right). \tag{12}$$

Since the sequential and direct paths reach the same physical endpoint, their predicted latent representations should remain consistent. Using the same normalized token distance as in the one-step prediction objective, we minimize

$$\mathcal{L}_{\text{comp}} = d_z\left(\widehat{Z}^{(n)}, \text{sg}\left(\widehat{Z}^{(\text{c})}\right)\right). \tag{13}$$

Here, the direct composed prediction serves as a stop-gradient reference for the sequential prediction in the composition objective.

The complete Stage I objective is

$$\mathcal{L}_{\text{S1}} = \mathcal{L}_{\text{pred}} + \mathcal{L}_{\text{comp}}, \tag{14}$$

Together, the two objectives train the model to capture individual physical transitions and their compositional structure, encouraging physically equivariant latent representations for Stage II.

3.3 Stage II: Stable RFF Representation Learning from Physical Equivariance

Stage II learns a stable RFF representation from the physically equivariant encoder representations while encouraging the shared encoder to retain the transformation information learned in Stage I. The encoder is initialized with the weights learned in Stage I.

3.3.1 Single-View RFF Extraction

Direct cross-view alignment can drive the encoder toward an overly invariant representation and suppress the transformation structure learned in Stage I. We therefore introduce an RFF attention query head Q_{RFF} on top of the shared encoder.

For each physical view v,

$$Z_v = F_\omega(v), \qquad r_v = Q_{\mathrm{RFF}}(Z_v), \tag{15}$$

where Z_v is the physically equivariant latent representation and r_v the extracted RFF representation. Q_{RFF} uses a learnable query to attend to Z_v, and is applied independently to each view without transformation parameters or cross-view aggregation.

Each r_v is supervised by standard cross-entropy classification and a multi-positive supervised contrastive objective over ℓ_2-normalized RFF representations. All other representations from the same device, including different physical views and observations, form positives, while representations from different devices form negatives.

The identity objective is

$$\mathcal{L}_{\mathrm{id}} = \mathcal{L}_{\mathrm{cls}} + \mathcal{L}_{\mathrm{rff}}, \tag{16}$$

Thus, view invariance is learned at the RFF representation rather than imposed directly on the physically equivariant encoder representation.

3.3.2 Relative Transformation Recovery

Because identity gradients still update the shared encoder, we additionally require relative physical transformations to remain recoverable from its latent representations. For two physically related views $x^{(i)}$ and $x^{(j)}$, a relative transformation recovery head R_{rel} predicts

$$\widehat{p}_{i\to j} = R_{\mathrm{rel}}\left(Z_i, Z_j\right). \tag{17}$$

The predicted descriptor $\widehat{p}_{i\to j}$ is trained against the same physics-structured action representation $p_{i\to j} = \Psi_{\mathrm{RF}}(\theta_{i\to j})$ used in Stage I using mean squared error (MSE), which defines the recovery loss $\mathcal{L}_{\mathrm{rel}}$.

The complete Stage II objective is

$$\mathcal{L}_{\mathrm{S2}} = \mathcal{L}_{\mathrm{id}} + \mathcal{L}_{\mathrm{rel}}, \tag{18}$$

This objective learns a view-invariant and device-discriminative RFF representation while encouraging the shared encoder to retain transformation information.

Physical views and transformation parameters are used only as training supervision. Since each RFF representation is extracted independently from a single view without condition input, deployment requires only one observed I/Q signal passed through the encoder, RFF query head, and classifier.

4 Experiments

We evaluate EquiRFF on four public RFFI datasets covering common RF signals collected from WiFi transmitters, UAVs, and Bluetooth devices. These datasets further span common cross-domain scenarios involving acquisition bursts, receivers, recording sessions, and transmission distances. All experiments follow a source-only protocol in which training and model selection use source-domain data and evaluation is performed on held-out target domains.

4.1 Experimental Setup

Datasets and protocols. Table 1 summarizes the four evaluation protocols. UAV (Soltani et al., 2020) contains seven DJI M100 UAVs recorded at 6, 9, 12, and 15 ft across multiple acquisition bursts. We use the first 16,384 complex I/Q samples of each observation. ManySig (Hanna et al., 2022) contains six WiFi transmitters from the WiSig collection, observed by multiple receivers over four recording days, with 256 complex I/Q samples per observation. We use ten receivers as the source domain and hold out receivers `7-7` and `8-8` as two unseen target receivers. ORACLE (Sankhe et al., 2019) contains over-the-air recordings from 16 USRP X310 transmitters, with 6,000 I/Q samples per observation. BLE31 (Albousayri et al., 2025) contains 31 Bluetooth devices recorded outdoors at 1, 1.5, 2, and 3 m, with 1,850 I/Q samples per observation.

The four datasets cover complementary deployment shifts. For UAV, the source domain consists only of bursts 1–3 at 6 and 9 ft, while burst 4 at 6 and 9 ft and all bursts at 12 and 15 ft are held out as target domains. ManySig evaluates cross-receiver generalization, ORACLE cross-session generalization, and BLE31 cross-distance generalization.

Table 1: Datasets and source-to-target evaluation protocols. (B denotes acquisition burst.)

Dataset	Devices	I/Q length	Domain shift	Source → target
UAV	7	16,384	Burst + distance	6/9 ft B1–3 → 6/9 ft B4, 12/15 ft B1–4
ManySig	6	256	Receiver	10 receivers → 2 unseen receivers
ORACLE	16	6,000	Recording session	Run1 → Run2 at 2 ft
BLE31	31	1,850	Distance	1 m → 1.5/2/3 m

Baselines. We compare EquiRFF against a set of recent strong baselines spanning predictive, invariant, transformation-aware, and RF-specific domain-generalization methods. Vanilla 1D-JEPA (Assran et al., 2023) and BYOL (Grill et al., 2020) provide predictive and invariant representation-learning references. EquiMod (Devillers & Lefort, 2023), CARE (Gupta et al., 2024), STL (Yu et al., 2024), V-JEPA 2-AC (Assran et al., 2025), seq-JEPA (Ghaemi et al., 2025), and SER (Lee et al., 2026) represent transformation-aware or equivariant learning strategies adapted to one-dimensional I/Q observations. Avoiding Shortcuts (Wang et al., 2025) serves as an RF-specific source-only domain-generalization baseline.

Implementation details. For controlled comparison, all baselines and EquiRFF use the same one-dimensional Transformer encoder. The encoder consists of six standard Transformer blocks with representation dimension $D = 256$, eight-head self-attention, and an FFN width of $4D$.

EquiRFF introduces three lightweight modules on top of the shared encoder. The physical action projector G_{phy} maps the 8-dimensional physics descriptor to D through a 128-dimensional hidden layer, with GELU activation and LayerNorm on the output. The action-conditioned predictor P contains three Transformer blocks, each with eight-head self-attention and an FFN width of $4D$. The RFF query head Q_{RFF} uses one learnable query and one cross-attention block with eight attention heads and an FFN width of $2D$, and outputs a D-dimensional RFF representation.

The physical action projector and action-conditioned predictor are used only for training-time physical supervision, while the RFF query head serves as the device representation head during identification. All methods use the same backbone encoder architecture and representation dimension.

We use three successive transformations of the same physical factor for composition supervision, with the remaining factors kept at the identity. CPO is sampled uniformly from $[-\pi, \pi]$, IQI uses amplitude imbalance in $[-1, 1]$ dB and phase imbalance in $[-5, 5]^\circ$, and the normalized CFO range is fixed to $\pm 5.0 \times 10^{-4}$. Both stages are optimized with AdamW using an initial learning rate of 1×10^{-4}, a batch size of 128, warmup, and cosine learning-rate decay. Stage I additionally uses an EMA target encoder. We train on ORACLE for 1000 epochs and on the other datasets for 100 epochs, which is sufficient for source-domain convergence. We report classification accuracy (%) for each target domain and source-domain training and validation accuracy to verify source-domain optimization. The best target result in each column is shown in bold. All methods are evaluated from a single I/Q observation.

4.2 Cross-Domain Identification Results

For all datasets, we report source-domain training and validation accuracy together with target-domain test accuracy. The source results confirm adequate optimization on the training distributions. We therefore evaluate cross-domain robustness primarily by target-domain test performance.

UAV evaluates generalization across both acquisition bursts and transmission distances. Table 2 shows that EquiRFF achieves the best accuracy in all ten target domains, with an average target accuracy of 93.66%. The gains remain consistent across the held-out burst at the source distances and the unseen transmission distances. In particular, EquiRFF reaches 87.72% on 6–4, 92.56% on 12–3, and 89.12% on 15–1, exceeding the strongest competing result in these domains by 7.72, 8.76, and 10.52 percentage points, respectively. These results indicate that the learned representation remains effective when burst and distance changes occur separately or jointly.

ManySig further evaluates cross-receiver generalization across four recording days. EquiRFF achieves the best accuracy on six of the eight target domains, including three of four days for each

Table 2: Cross-domain identification accuracy (%) on UAV. (Domain labels denote distance–burst.)

	Source		Target domains								
Method	**Train**	**Val**	**6–4**	**9–4**	**12–1**	**12–2**	**12–3**	**12–4**	**15–1**	**15–2**	**15–3**	**15–4**
Vanilla 1D-JEPA	99.87	92.26	80.00	99.38	94.66	85.09	83.80	85.73	73.80	77.37	88.77	98.43
EquiMod	99.41	91.89	76.65	98.24	94.24	88.51	79.02	82.70	71.58	71.16	77.04	95.75
CARE	99.09	88.12	57.66	81.64	58.15	77.87	66.53	86.99	62.11	54.51	71.85	82.05
STL	99.12	74.58	50.96	58.71	46.91	63.45	58.57	60.98	72.87	48.77	57.04	75.28
BYOL	99.97	86.19	62.94	82.26	89.19	80.44	52.72	55.93	61.99	61.08	50.12	68.19
V-JEPA 2-AC	98.68	90.60	71.88	93.15	96.49	93.03	61.89	71.84	64.44	71.16	63.70	71.50
Avoiding Shortcuts	98.17	87.10	62.94	91.70	78.51	76.77	71.45	67.05	78.60	66.12	79.01	87.87
seq-JEPA	99.53	84.15	79.29	73.34	84.41	80.68	70.12	56.82	66.90	72.10	59.51	71.34
SER	99.77	90.70	75.03	98.65	98.31	78.36	75.43	71.34	71.81	70.69	72.96	96.54
EquiRFF (Ours)	99.71	94.20	**87.72**	**99.59**	**99.72**	**96.33**	**92.56**	**91.41**	**89.12**	**85.23**	**95.19**	**99.69**

Table 3: Cross-domain identification accuracy (%) on ManySig.

	Source		Target `7-7`				Target `8-8`			
Method	**Train**	**Val**	**03/01**	**03/08**	**03/15**	**03/23**	**03/01**	**03/08**	**03/15**	**03/23**
Vanilla 1D-JEPA	99.65	99.43	74.78	78.39	72.06	63.28	78.06	79.00	73.33	87.22
EquiMod	99.55	99.43	84.83	73.89	78.89	68.39	79.83	82.06	60.00	69.39
CARE	99.50	99.35	74.94	61.44	64.72	64.67	75.50	82.11	84.39	**94.44**
STL	99.71	99.40	72.33	79.72	68.11	72.67	79.39	80.28	75.11	85.33
BYOL	99.60	99.42	75.94	75.56	70.72	66.89	69.44	83.11	79.06	80.33
V-JEPA 2-AC	99.93	99.53	**86.67**	81.44	80.17	65.94	70.17	83.89	75.17	80.94
Avoiding Shortcuts	99.73	98.78	80.72	86.61	85.06	88.50	83.61	83.72	81.22	87.89
seq-JEPA	97.96	97.70	74.94	60.56	63.00	59.33	74.22	81.11	63.00	75.00
SER	98.67	98.41	54.50	53.78	47.22	50.72	38.61	69.39	58.44	36.72
EquiRFF (Ours)	98.21	97.68	86.56	**86.89**	**85.71**	**89.37**	**85.56**	**92.83**	**92.94**	93.72

unseen receiver. For receiver `8-8`, it reaches 92.83% and 92.94% on 03/08 and 03/15, improving over the strongest competing results by 8.94 and 8.55 percentage points. On the two remaining domains, EquiRFF is close to the best competing method, reaching 86.56% versus 86.67% on receiver `7-7` on 03/01 and 93.72% versus 94.44% on receiver `8-8` on 03/23. The consistent performance across both held-out receivers and multiple recording days suggests that the learned RFF representation is not tied to a particular receiver condition.

ORACLE evaluates a different type of domain shift across independently repeated recording sessions. EquiRFF reaches 95.74% on Run2, compared with 93.19% for the strongest baseline, showing that the improvement extends beyond receiver and distance changes. On BLE31, where the source domain is collected at 1 m, EquiRFF achieves the best accuracy at 2 m and 3 m with 96.47% and 97.95%, respectively. At 1.5 m, where several methods already approach ceiling performance, EquiRFF reaches 99.37%, slightly below the best result of 99.98%. Across the 22 target domains in Tables 2–4, EquiRFF achieves the highest accuracy on 19, covering burst, receiver, recording-session, and distance shifts.

4.3 Ablation Study

We conduct component ablations on UAV to examine the contributions of the main design choices in EquiRFF. The first variant directly embeds the raw transformation parameters for one-step latent prediction. The second replaces them with the proposed physics-structured action representation. We then progressively introduce composition consistency, Stage-II RFF representation learning with an MLP, and finally the proposed RFF query head. All other training and evaluation settings are kept identical to those used in the cross-domain experiments.

Replacing the raw transformation parameters with the physics-structured action representation improves all ten target domains and raises the average accuracy from 61.62% to 80.25%. The broad

Table 4: Cross-domain identification accuracy (%) on ORACLE and BLE31.

	ORACLE			BLE31				
Method	**Train**	**Val**	**Run2**	**Train**	**Val**	**1.5 m**	**2 m**	**3 m**
Vanilla 1D-JEPA	97.08	84.51	76.13	99.51	99.68	99.44	91.95	97.11
EquiMod	96.13	76.53	76.25	96.96	99.92	99.74	95.31	87.95
CARE	98.54	90.14	80.38	99.37	99.71	**99.98**	92.66	92.48
STL	96.58	88.89	74.50	99.94	99.52	99.16	87.85	93.95
BYOL	93.96	84.65	63.25	97.80	99.68	98.74	88.68	92.60
V-JEPA 2-AC	92.38	77.08	71.25	93.10	99.11	96.39	86.34	88.65
Avoiding Shortcuts	95.09	97.78	93.19	99.57	99.44	98.03	84.90	89.06
seq-JEPA	91.28	71.60	67.69	93.47	92.42	66.87	50.10	52.90
SER	94.91	90.63	68.50	82.96	88.71	70.94	48.15	62.06
EquiRFF (Ours)	99.88	99.44	**95.74**	97.21	99.92	99.37	**96.47**	**97.95**

Table 5: Component ablation on UAV. (Domain labels denote distance–burst.)

Configuration	**6–4**	**9–4**	**12–1**	**12–2**	**12–3**	**12–4**	**15–1**	**15–2**	**15–3**	**15–4**
Raw action	58.07	80.08	77.53	68.58	64.01	66.29	53.80	46.89	43.58	57.32
Structured action	73.10	87.72	90.45	82.73	83.00	76.62	74.85	73.51	81.98	78.58
+ Composition	74.52	94.40	92.56	87.41	83.71	85.34	82.95	81.62	84.17	90.00
+ MLP	84.31	97.14	95.74	89.72	84.62	85.63	87.43	82.71	88.34	89.37
+ RFF query	**87.72**	**99.59**	**99.72**	**96.33**	**92.56**	**91.41**	**89.12**	**85.23**	**95.19**	**99.69**

improvement across domains shows that explicitly encoding the known structure of the physical transformations provides a substantially stronger conditioning signal than directly embedding their raw parameters. Adding composition consistency further improves every target domain and increases the average accuracy to 85.67%. This additional gain supports the role of multi-step physical supervision beyond independent one-step prediction.

Stage-II identity learning provides further improvements. Using an MLP readout increases the average accuracy from 85.67% to 88.50% and improves nine of the ten target domains, showing that extracting identity-stable information from the equivariant representation is beneficial even with a simple readout. Replacing the MLP with the proposed RFF attention query head improves all ten domains and raises the average accuracy to 93.66%. The gain from 88.50% to 93.66% indicates that the attention-based readout is more effective at extracting device-discriminative information from the high-dimensional encoder representation. Overall, the ablation results show progressive contributions from physics-structured action encoding, composition consistency, Stage-II identity learning, and the RFF query head.

5 Conclusion

We presented EquiRFF, a two-stage framework for source-only cross-domain RFFI that learns stable device identity from physical equivariance. Stage I combines physics-structured action encoding, action-conditioned latent prediction, and composition consistency to model RF transformation dynamics, while Stage II learns a stable RFF representation from physically equivariant features using an RFF attention query head. Experiments on four public datasets covering WiFi, UAV, and Bluetooth signals show strong generalization across acquisition-burst, receiver, recording-session, and distance shifts, and the ablations show incremental gains from structured action modeling, composition consistency, and the RFF query. At inference, EquiRFF requires only a single I/Q observation without target-domain adaptation. Future work will extend the framework to broader RF impairments and more complex open-world deployment conditions.

AI USE STATEMENT

Generative AI tools were used to aid or polish writing by checking the manuscript for typographical, grammatical, and wording issues, and for retrieval and discovery of potentially relevant literature and recent research developments. All identified literature was independently verified against the original sources, and all AI-assisted edits were reviewed by the authors. We take responsibility for the final content of this work.

REFERENCES

Haytham Albousayri, Bechir Hamdaoui, Weng-Keen Wong, and Nora Basha. Bluetooth fingerprint identification under domain shift through transient phase derivative. In *2025 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–9, 2025.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15619–15629, 2023.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, et al. V-jepa 2: Self-supervised video models enable understanding, prediction and planning. *arXiv preprint arXiv:2506.09985*, 2025.

Vladimir Brik, Suman Banerjee, Marco Gruteser, and Sangho Oh. Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, pp. 116–127, 2008.

Boris Danev, Davide Zanetti, and Srdjan Capkun. On physical-layer identification of wireless devices. *ACM Computing Surveys*, 45(1):1–29, 2012.

Alexandre Devillers and Mathieu Lefort. Equimod: An equivariance module to improve visual instance discrimination. In *International Conference on Learning Representations*, 2023.

Hafez Ghaemi, Eilif B. Muller, and Shahab Bakhtiari. seq-jepa: Autoregressive predictive learning of invariant-equivariant world models. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, koray kavukcuoglu, Remi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21271–21284, 2020.

Chaojie Gu, Jingze Zheng, Haoran Shi, Yichuan Yang, Shibo He, and Zhiguo Shi. Efficient sim2real deep learning for device-specific ofdm frequency offset calibration. *IEEE Transactions on Wireless Communications*, 25:19355–19366, 2026.

Sharut Gupta, Joshua Robinson, Derek Lim, Soledad Villar, and Stefanie Jegelka. Structuring representation geometry with rotationally equivariant contrastive learning. In *International Conference on Learning Representations*, 2024.

Samer Hanna, Samurdhi Karunaratne, and Danijela Cabric. Wisig: A large-scale wifi signal dataset for receiver and channel agnostic rf fingerprinting. *IEEE Access*, 10:22808–22818, 2022.

Sai Huang, Jiashuo He, Zheng Yang, Yuting Chen, Shuo Chang, Yifan Zhang, and Zhiyong Feng. Generalized automatic modulation classification for ofdm systems under unseen synthetic channels. *IEEE Transactions on Wireless Communications*, 23(9):11931–11941, 2024.

Shiyue Huang, Yuchen Su, Hongbo Liu, Zikang Ding, Xuewan He, Yanzhi Ren, and Haitao Jia. Towards distance-invariant radio frequency fingerprinting via augmented unsupervised learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(1):426–434, 2026.

Anil Kumar, Kumar Appaiah, and Sibi Raj B. Pillai. Carrier phase and frequency discriminators for receivers with 1-bit quantization. *IEEE Transactions on Communications*, 73(11):10509–10523, 2025.

Joohyung Lee, Changhun Kim, Hyunsu Kim, Kwanhyung Lee, and Juho Lee. Soft equivariance regularization for invariant self-supervised learning. In *International Conference on Learning Representations*, 2026.

Taotao Li, Yiyang Li, Zhenyu Wen, Jiahao Lin, Jinhao Wan, Jie Su, Cong Wang, and Zhen Hong. Rff-tta: Physical information-aware prototype for temporally varying RF fingerprinting online test-time-adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(1):686–694, 2026.

Yuepei Li, Kai Xu, Junqing Zhang, Chongyan Gu, Yuan Ding, George Goussetis, and Symon K. Podilchak. Puf-assisted radio frequency fingerprinting exploiting power amplifier active load-pulling. *IEEE Transactions on Information Forensics and Security*, 19:5015–5029, 2024.

Hamed Masoumi and Nitin Jonathan Myers. Compressed sensing-based sparse spatial channel estimation under iq imbalance. *IEEE Transactions on Signal Processing*, 74:3942–3955, 2026.

Kunal Sankhe, Mauro Belgiovine, Fan Zhou, Shamnaz Riyaz, Stratis Ioannidis, and Kaushik Chowdhury. Oracle: Optimized radio classification through convolutional neural networks. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pp. 370–378, 2019.

Guanxiong Shen, Junqing Zhang, Alan Marshall, Linning Peng, and Xianbin Wang. Radio frequency fingerprint identification for lora using deep learning. *IEEE Journal on Selected Areas in Communications*, 39(8):2604–2616, 2021.

Hanmin Sheng, Kai Chen, Wenlong Zeng, Hang Geng, Wenjian Ma, and Shafei Wang. Carrier frequency offset robust self-compensating neural network. *IEEE Transactions on Wireless Communications*, 24(5):4073–4085, 2025.

Nasim Soltani, Guillem Reus-Muns, Batool Salehi, Jennifer Dy, Stratis Ioannidis, and Kaushik Chowdhury. Rf fingerprinting unmanned aerial vehicles with non-standard transmitter waveforms. *IEEE Transactions on Vehicular Technology*, 69(12):15518–15531, 2020.

Yu Wang, Tomoaki Ohtsuki, Zhi Sun, Dusit Niyato, Xianbin Wang, and Guan Gui. Avoiding shortcuts: Enhancing channel-robust specific emitter identification via single-source domain generalization. *IEEE Transactions on Wireless Communications*, 24(4):3163–3176, 2025.

Zhisheng Yao, Yang Peng, Yu Wang, Congan Xu, Juzhen Wang, Yun Lin, and Guan Gui. A novel radio frequency fingerprint concealment method based on iq imbalance compensation and digital pre-distortion. *IEEE Transactions on Information Forensics and Security*, 19:7349–7361, 2024.

Jaemyung Yu, Jaehyun Choi, Dong-Jae Lee, HyeongGwon Hong, and Junmo Kim. Self-supervised transformation learning for equivariant representations. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Junqing Zhang, Francesco Ardizzon, Mattia Piana, Guanxiong Shen, and Stefano Tomasin. Physical layer-based device fingerprinting for wireless security: From theory to practice. *IEEE Transactions on Information Forensics and Security*, 20:5296–5325, 2025a.

Ying Zhang, Qiang Li, Hongli Liu, Liu Yang, and Jian Yang. Domain generalization for cross-receiver radio frequency fingerprint identification. *IEEE Internet of Things Journal*, 12(5):5207–5218, 2025b.