

OpenCSI: Self-Calibration Layer for Heterogeneous Mesh Wireless Sensor Networks

Karim Khamaisi, Simon Sigg, Bruno Rodrigues

Sensing Group SG, Institute of Computer Science in Vorarlberg ICV, University of St. Gallen HSG

/,Hintere Achmühlerstraße 1, 6850 Dornbirn, Austria

E-mail: {karim.khamaisi, bruno.rodrigues}@unisg.ch, simon.sigg@student.unisg.ch

Abstract—WiFi CSI sensing models trained in one environment usually fail in another, because standard normalization embeds chip- and room-specific artifacts into the feature space. We propose *OpenCSI*, an abstraction layer that hides those artifacts from the sensing logic above by exposing each link as a dimensionless Z-score against its own quiet-period temporal standard deviation. The denominator is learned online from a 120-second empty-room bootstrap and reported with a maturity tag, so downstream code can detect drift and abstain when the baseline is unreliable.

We evaluate OpenCSI across three rooms and three ESP32 generations (S3, C3, C6 at HT20 and HE20), with two deployments varying chip and room together and one controlled S3→C6 swap that holds positions fixed. A model trained on one cell holds binary F1 between 0.896 and 0.993 in seven of eight transfer cells without retraining, against 0.87 for warmup-normalized features; the eighth cell collapses on the noisiest silicon (ESP32-C3), whose compressed dynamic range breaks the ratio. The transfer is binary by construction: separating static from moving needs the magnitude that the temporal standard deviation absorbs. We release the source code and the evaluation dataset to support reproducible cross-environment CSI research.

Index Terms—WiFi CSI, calibration, temporal normalization, sensing, indoor localization

I. INTRODUCTION

Indoor occupancy sensing has no clean default. Cameras break privacy, PIR misses sitting occupants, and wearables require compliance [1]. HVAC, lighting, and emergency-response systems still over-ventilate empty rooms and miss occupied ones [5]. WiFi Channel State Information (CSI), the per-subcarrier complex frequency response that an OFDM receiver already extracts from every frame, senses motion and presence on hardware already on the wall, with no cameras or wearables [14], [26]. Low-cost ESP32 microcontrollers, typically under 3 USD, expose CSI directly from commodity 802.11 frames [6], so a mesh of dozens to hundreds of nodes is now within an undergraduate budget.

Portability is a major challenge for widespread deployment of wireless sensing, because a CSI model trained in one room rarely transfers to another, to a different node batch, or even across a single power cycle of the same node. CSI-Bench [31] is a work that codifies the barrier by reporting that state-of-the-art methods lose 33-70% of F1 under cross-device, cross-environment, and cross-user evaluation. A decade of work has thrown larger models, domain adaptation, and meta-learning at it [12], [28], [25],

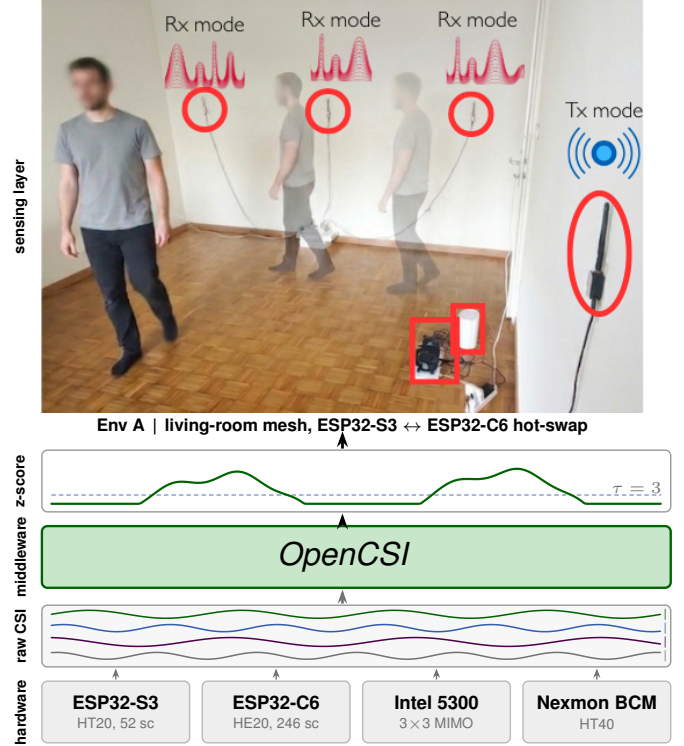


Fig. 1: OpenCSI normalises raw CSI from heterogeneous hardware (bottom) into per-link Z-scores (middle). The sensing layer (top, Env A living-room mesh) consumes one stable input across chip variants.

[30], [3], each still needing a fresh labeled or unlabeled capture from the new deployment.

The cost in this scenario compounds with scale of the network. A mesh of N nodes exposes $N(N-1)$ directed links, where each link carry its own per-boot phase, AGC (Automatic Gain Control) state, and chip-specific gain, and the parameter space multiplies again when chip families are mixed. The computational costs comes from standard preprocessing pipelines, which normalizes each recording against its own statistics, so the per-boot phase, AGC transient, and chip gain of that recording end up inside the features [16]. A model trained on those features cannot distinguish hardware variations from channel variation, and breaks the moment the hardware also changes.

We propose *OpenCSI*, an abstraction layer that decouples calibration from sensing logic by exposing each link as a single per-link signal (a z-score). While the room is empty, each link learns its own quiet-period mean and *temporal* standard deviation through a Welford accumulator [22], and once those statistics are warm every frame is converted to a dimensionless Z-score $z = |\bar{A} - \mu|/\sigma$.

Room geometry and chip gain enter σ multiplicatively, so the ratio cancels them. The additive noise-floor differences between chip families are not absorbed and compress the ratio in noisier hardware. The deviation $|\bar{A} - \mu|$ in the numerator carries the channel perturbation, so a 2σ value carries the same meaning across the chips, rooms, and reboots tested, and the sensing layer above consumes one stable per-link signal instead of re-solving calibration on every deployment.

We tested *OpenCSI* by running three real-world experiments. *Cross-room*: S3↔S3' in two different rooms, same PHY, different board batches. *Cross-chip*: S3↔C3 between rooms, same PHY but different silicon (the chip axis is confounded with the room axis here). *Cross-generation in the same room*: S3↔C6 in Env A, same geometry, different PHY (802.11n HT20 vs 802.11ax HE20, 52 vs 246 subcarriers). A single empty-versus-occupied threshold yields binary F1 ≥ 0.978 cross-room, F1 = 0.896–0.993 in three of four cross-chip directions (one fails at 0.524, §IV), and 0.985/0.896 across the two cross-generation directions on a shared 2.4 GHz carrier, with no per-deployment retuning (Fig. 1). Three-class macro F1 (empty/static/moving) falls below the binary threshold cross-room: the temporal-std denominator absorbs the magnitude information by design, so the transfer claim is scoped to *sign-thresholding* tasks (§V).

This paper makes three contributions.

- **A single normalized signal per link.** OpenCSI normalizes per-link amplitude by its own quiet-period temporal standard deviation. The abstraction holds for binary sign-thresholding in seven of eight cross-deployment cells (F1 ≥ 0.89 in seven, with one direction failing at 0.524 from C3's compressed dynamic range, §IV) and stops short of magnitude-sensitive tasks (three-class macro F1 falls below 0.7 cross-room).
- **Calibration drift as a monitorable mesh property.** A per-link reliability score and a frame-level calibration-maturity signal flag stale baselines within seconds; 30 s of fresh quiet observation restores F1 from 0.25 to 0.98 (§IV-E).
- **Open-source.** Apache-2.0 library with the Welford pipeline, maturity tag, and reliability scoring; the three-environment CSI traces (S3, C3, C6 at HT20 and HE20); and the evaluation scripts that reproduce every cell of Tables II–III.¹

TABLE I: Comparison with cross-environment approaches for WiFi CSI sensing.

Approach	Per-deployment cost		Capability	
	No tgt. data	No re-training	Cross-chip	On-device
Warmup-normalized amplitude	✓	✓	~ [†]	✓
CSI-ratio [27]	✓	✓	✗	✓
Domain adapt. [12], [3]	✗	✗	✗	✗
Per-deploy. calib. [31]	✗	✗	✗	✓
Foundation [11]	✗	~	✗	✗
OpenCSI	✓	✓	✓ [*]	✓

[†] 0.87 F1 cross-environment, but session-specific scale is embedded. ^{*} Sign-thresholding; cross-chip confounded with room; C3–S3' asymmetric. See §IV.

II. BACKGROUND AND RELATED WORK

Channel State Information is what an OFDM receiver already produces in the act of decoding a frame. After demodulating the long training field the receiver holds a per-subcarrier complex vector $H_k = A_k e^{j\phi_k}$ that records amplitude and phase across the band, and that richer description is what separates fine-grained sensing from RSSI-style coarse motion [14], [26], [9]. Pulling that vector off commodity hardware once required research-grade NICs and firmware patches [9], [13], [8], [4], and IEEE 802.11bf will eventually standardize the path [10]; ESP32 SoCs already expose CSI directly from radio firmware [6], [7] at a price that puts dense meshes within reach.

Multipath propagation (*i.e.*, radio signals travel multiple paths to a receiver, bouncing off objects) turns that vector into a sensor. Frames reach the receiver along a line-of-sight component and many reflected ones, their interference is what H_k encodes, and a moving body rearranges the paths within seconds. The **Fresnel zones** around each transmitter-receiver pair fix which spatial regions perturb the channel and by how much, so device placement is a first-order design parameter [20], [21]. A mesh of N ESP32 nodes exposes $N(N-1)$ directed links, each an independent spatial slice of the sensing volume [6], [2], with 52 subcarriers per link on HT20 and 246 on the ESP32-C6 through the 802.11ax HE-LTF path [7], [4]. Adjacent commodity-radio sensing modalities (ZigBee [23], [15], BLE [17]) share the same passive premise but lack CSI's per-subcarrier resolution.

Calibration is in this scenario the main challenge that prevents a mesh from transferring across deployments. Prior work attacks calibration three ways (Table I): algebraic per-frame cancellation, model-side adaptation, and foundation-model pretraining. We compare against each below.

The closest prior method is CSI-ratio [27], which differs along a complementary axis. CSI-ratio is an *algebraic* per-frame cancellation across the two antennas of a multi-antenna NIC, exact for shared receiver-side gain, CFO, and

¹<https://github.com/sensing-group/OpenCSI>

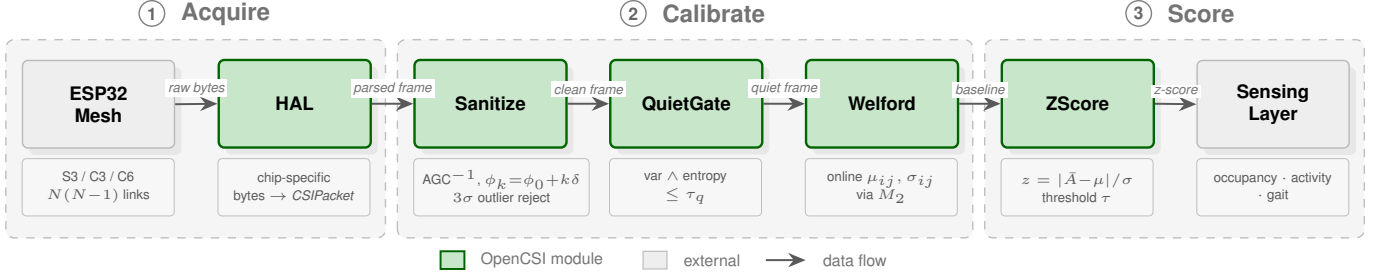


Fig. 2: OpenCSI pipeline. Raw CSI from a heterogeneous ESP32 mesh flows through three phase bands (Acquire, Calibrate, Score) and is emitted as the per-link tuple $(z, \text{reliability})$ to the sensing layer. The dashed arc inside Calibrate marks the QuietGate admitting only quiet-period frames into the Welford estimator.

SFO. OpenCSI is a *statistical* cancellation across each link’s own quiet time window, which costs a 120s bootstrap but works on single-antenna nodes and across mesh links, where no co-located antenna pair exists.

Most other cross-deployment work pushes the burden into the model. Adversarial domain adaptation [12], meta-learning [28], compressed feature learners [25], augmentation-based generalization [30], body-coordinate-velocity features [29], and end-to-end deep models [19] all assume target-environment samples, and the calibration step itself remains per-deployment across the SenseFi [24] and CSI-Bench [31] suites. Foundation-model pretraining moves that cost from per-deployment to one-time by learning environment-invariant representations from millions of samples [11], but the compute is substantial and cross-hardware generalization remains undemonstrated.

III. OPENCSI DESIGN

OpenCSI emits one normalized signal per link to the sensing layer above (Fig. 2): a dimensionless Z-score paired with a scalar reliability tag, designed to carry consistent meaning across chip variants, rooms, and reboots. The library parses chip-specific packet formats, sanitizes hardware artifacts, and updates per-link baselines during quiet periods.

A. Hardware Abstraction

Raw CSI packets differ across chip families in subcarrier count, AGC state encoding, phase noise, and per-boot randomization (S3 HT20: 52 subcarriers in a 240-byte packet; C6 HE20: 246 subcarriers via the IDF v5.5 HE-LTF path [7], following AX-CSI [4]). A Hardware Abstraction Layer separates parsing from calibration: each chip family implements a *CSIExtractor* returning a canonical *CSIPacket*, and the downstream pipeline is chip-agnostic at the API level. We ship and validate extractors for the full ESP32 family (S3, C3, C6); the same interface admits parsers for Intel 5300, Nexmon-patched chipsets, or PicoScenes NICs, but no non-ESP32 traces are reported here, so the empirical claim is scoped to ESP32 silicon.

B. Sanitization

Sanitization runs three steps before baseline learning. AGC correction inverts the per-packet gain factor recovered from *rx_state*. A linear phase trend $\phi_k = \phi_0 + k\delta$ is fit and removed. Packets drifting beyond 3σ from the short-term amplitude median are discarded as retransmits or partial captures.

C. Quiet Period Detection and Baseline Learning

Calibration learns what the channel looks like at rest by automatically identifying unoccupied periods. **Bootstrap (first 120s after power-on)** unconditionally treats incoming frames as quiet so the per-link Welford estimators reach stable mean/std; the room *must* be empty during this window or the occupant’s perturbation is absorbed into the baseline. **Active detection** then gates on two complementary signals over a sliding window: *temporal variance* (low variance = stable channel) and *spectral entropy* of the per-subcarrier amplitude distribution (gates out periodic HVAC/electronic interference that confounds variance-only detectors). A frame is quiet only if neither signal exceeds its learned threshold; in known-quiet phases (our protocol) the gate can be forced for a specified duration to avoid starving the learner in noisy environments.

When a frame is classified as quiet, each link’s sanitized amplitude and phase update a Welford online accumulator [22] that maintains a running mean and standard deviation without storing raw samples. Baselines are frequency-bucket-aware (six packet-rate buckets, 2 Hz to 60+ Hz) and time-of-day-aware (four 6-hour buckets per day to absorb diurnal thermal drift).

D. Z-score and Reliability: The Per-Link Metric

Let (μ_{ij}, σ_{ij}) denote the per-link mean and standard deviation of the cross-subcarrier amplitude average, learned during quiet periods. The calibrated amplitude Z-score for link $(i \rightarrow j)$ at time t is:

$$z_{ij}^{\text{amp}}(t) = \frac{|\bar{A}_{ij}(t) - \mu_{ij}|}{\sigma_{ij}}, \quad (1)$$

where $\bar{A}_{ij}(t)$ is the mean amplitude across subcarriers at time t . The deviation is *unsigned*: a person can increase

or decrease received amplitude depending on Fresnel-zone position, so only its magnitude carries presence information.

An analogous phase Z-score is computed from the standard deviation of phase differences across subcarriers, and the primary output *combined_change* is the SNR-weighted combination

$$z_{ij}^{\text{comb}} = (1 - w_\phi) \cdot z_{ij}^{\text{amp}} + w_\phi \cdot z_{ij}^{\text{phase}}, \quad (2)$$

where $w_\phi = \sigma((\text{SNR}_{ij} - 15)/5)$ up-weights phase on high-SNR links and down-weights it on low-SNR links. The differential-phase formulation cancels per-boot phase offsets and timing-offset trends that make absolute ESP32 phase unusable. The phase channel is exposed end-to-end in the API; on single-antenna ESP32 nodes the SNR weighting attenuates it (amplitude carries most of the signal, §IV), but it is retained for multi-antenna platforms where ratio cancellation [27] sharpens the phase statistic.

Each link ($i \rightarrow j$) carries a scalar *reliability* score $r_{ij} \in [0, 1]$ that combines four saturating factors:

$$r_{ij} = r_{ij}^{(\text{age})} \cdot r_{ij}^{(N)} \cdot r_{ij}^{(\text{SNR})} \cdot r_{ij}^{(\text{rate})}, \quad (3)$$

with $r_{ij}^{(\text{age})} = \exp(-\Delta t_{ij}/\tau_{\text{age}})$ decaying exponentially since the last quiet-period update, $r_{ij}^{(N)} = \min(1, N_{ij}^{\text{quiet}}/N_{\text{min}})$ saturating once the Welford accumulator has N_{min} quiet samples, $r_{ij}^{(\text{SNR})}$ a sigmoid in instantaneous SNR, and $r_{ij}^{(\text{rate})} = \min(1, f_{ij}/f_{\text{min}})$ saturating at a minimum packet rate. Uninitialized links carry $r_{ij} = 0$. The frame-level *calibration_maturity* is the weighted mean of r_{ij} across active links; §V shows accuracy plateaus once it exceeds ~ 0.65 . The drift-recovery experiment in §IV-E bounds the dominant terms empirically: 30 s of fresh quiet observation restores both $r^{(N)}$ and $r^{(\text{age})}$ to saturation and recovers F1 from 0.25 to 0.98, so under normal operation the deployment threshold is governed by these two factors rather than SNR or packet rate, which change more slowly.

E. Z-score Robustness

Eq. (1) divides each link's per-frame amplitude by its own quiet-period standard deviation. The carrier frequency, per-subcarrier gain, and chip AGC all multiply both the numerator and the denominator, so they cancel. The ratio therefore does not depend on band, PHY format, or subcarrier count, and the Z-score should transfer when any of them changes. We test this prediction in §IV-C by swapping an 802.11n HT20 mesh (52 subcarriers) for an 802.11ax HE20 mesh (246 subcarriers) at identical positions on the same 2.4 GHz carrier.

Additive noise behaves differently as it enters σ as $\sigma_{ij}^2 \approx \sigma_{\text{channel}}^2 + \sigma_n^2$, and inflates the denominator on chips with a higher noise floor. For example, the ESP32-C3, which compresses the ratio and degrades cross-chip transfer (§IV).

IV. REAL-WORLD EVALUATION

A. Setup

Hardware. Each environment is instrumented with eight ESP32 nodes in a perimeter topology (corners and mid-walls), yielding 56 directed links. Env A and Env B use *different* ESP32-S3 board batches (802.11n HT20, 52 subcarriers); we denote the Env B batch as ESP32-S3' to keep the two distinguishable in tables and prose. All nodes transmit raw CSI via UDP to a central collection host.

Environments.

- **Env A:** 3.28×6.06 m living room, training environment, ESP32-S3.
- **Env B:** 7.02×4.44 m robotics laboratory in a different building, zero-shot test, ESP32-S3'.
- **Env C:** 2.91×4.43 m office, held-out cross-chip test, ESP32-C3.

Env C uses ESP32-C3 (same 802.11n HT20 PHY as S3, different silicon: AGC, noise floor, dynamic range), isolating silicon-level effects at fixed PHY; §IV-C additionally swaps Env A's S3 mesh for eight ESP32-C6 nodes (802.11ax HE20, 2.4 GHz, 246 useful subcarriers) at identical positions. Floor-plans and photographs are in Fig. 3.

Data collection. Each 12-minute session is recorded at 20 Hz with phases 5 min quiet baseline (forced-quiet, encompassing the 120 s bootstrap of §III) / 3 min static / 2 min moving / 2 min trailing empty, with timestamped class labels (empty, static, moving). Binary evaluation merges static and moving into "occupied." Raw UDP payloads are saved for offline replay.

Baselines. *Warmup-normalized amplitude:* per-link amplitude $A_k = \sqrt{I_k^2 + Q_k^2}$ normalized by the mean/std of the warmup window of the same session, with source statistics applied zero-shot at the cross-environment boundary. We use this as a competence floor for the cross-deployment setting; it is not a claim about the field's strongest preprocessor (e.g., per-subcarrier z-scoring or PCA-on-amplitude), and we revisit that scope in §V. *Ablation (cross-sc-std):* cross-subcarrier standard deviation only, without OpenCSI's temporal denominator, quiet gating, or reliability weighting; isolates the contribution of the temporal-std denominator.

Feature extraction. For both OpenCSI and the standard baseline, per-link scalar features are aggregated across the L active links to produce a fixed-size feature vector:

$$\mathbf{f} = \left[\max_l z_l, \text{mean}_l z_l, \text{std}_l z_l, \sum_l \mathbf{1}[z_l > \tau] \right], \quad (4)$$

where $\tau = 2.0$ for calibrated Z-scores and $\tau = 1.5$ for normalized amplitude. OpenCSI additionally provides *reliability-weighted* variants.

Models. A Random Forest (100 trees) and a two-layer MLP (64-64, ReLU, Adam) trained in scikit-learn, with 5-fold cross-validation within the source environment for hyperparameters. We report RF throughout; MLP is noted where relevant.

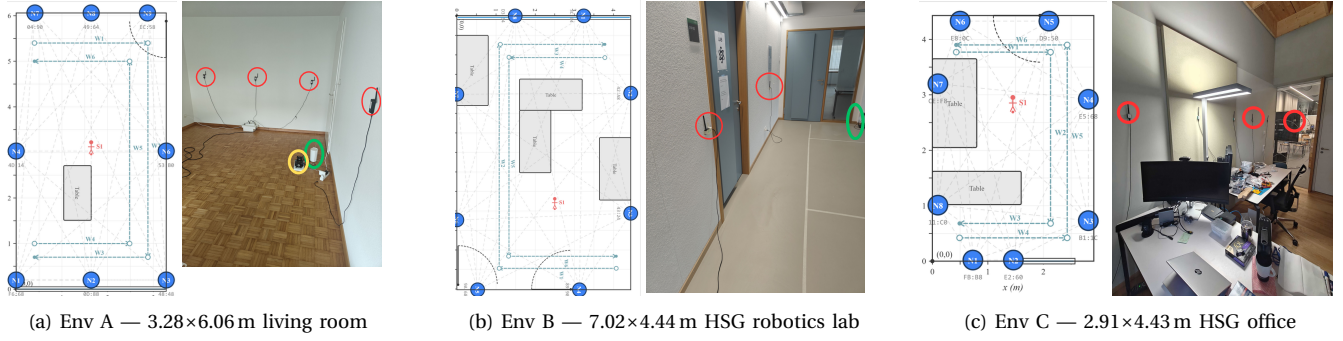


Fig. 3: Floorplans (left of each pair) and photographs (right) of the three deployment environments (chip variants in Table II). Blue • = mesh node; S = static pose; dashed teal = walking path (W1–W6). Photo overlays: red = node, green = router, yellow = backend host.

TABLE II: Binary cross-environment, cross-chip, and cross-generation transfer (F1-score, RF). Highlighted cells mark the load-bearing zero-shot transfers.

Features	Test on A (S3)				Test on B (S3')				Test on C (C3)				Test on A (C6)			
	A	B	C	A(C6)	A	B	C	A(C6)	A	B	C	A(C6)	A	B	C	A(C6)
Calibrated (OpenCSI)	1.000	0.982	0.896	0.985	0.978	1.000	0.524	0.990	0.993	0.974	1.000	0.592	0.896	0.882	0.580	1.000
Warmup-normalized	1.000	0.874	0.693	0.090	0.923	1.000	0.608	0.414	0.876	0.870	1.000	0.487	0.284	0.435	0.631	1.000
Ablation (cross-sc-std)	1.000	0.848	0.655	0.105	0.917	1.000	0.690	0.451	0.876	0.789	1.000	0.437	0.623	0.429	0.446	1.000

B. Experiment 1: Binary Cross-Environment Transfer

We train on one environment and evaluate zero-shot on the other with the same model weights. Table II reports binary F1 (RF) on three transfers: cross-environment at fixed PHY (Env A $\leftrightarrow$ Env B, S3 $\leftrightarrow$ S3'), cross-chip at fixed PHY (S3 $\leftrightarrow$ C3 via Env C), and cross-generation in Env A (S3 $\leftrightarrow$ C6, §IV-C).

Calibrated features achieve high cross-environment transfer in the same-variant directions (A $\leftrightarrow$ B, S3 $\leftrightarrow$ S3', F1 $\geq$ 0.978) and in the cross-variant S3 $\rightarrow$ C3 direction (A,B $\rightarrow$ C, F1 = 0.993 and 0.974), while warmup-normalized amplitude drops by 8–20 percentage points in the same settings. The ablation feature set performs no better than warmup-normalized amplitude in several directions, consistent with the temporal-std denominator carrying most of the transfer benefit on the tasks studied. Figure 4 visualizes the distribution-level effect: standard session-normalized Z-scores have deployment-specific ranges (and in Env C fail to separate empty from occupied at all), while OpenCSI's calibrated Z-scores retain a single decision threshold ($z \approx 3$) across rooms, chip variants, and WiFi generations. The cross-generation panel (rightmost, ESP32-C6) is discussed in §IV-C.

One direction is a structural failure: C(C3) $\rightarrow$ B(S3') drops to 0.524, below both warmup-normalized amplitude (0.608) and the ablation (0.690). Env C's C3 nodes produce calibrated Z-scores at a compressed scale ($z_{\max} = 7.69$ vs. 20.47 in A and 30.37 in B), so a classifier trained on the narrower scale does not generalize to the wider one, whereas B $\rightarrow$ C transfers well. Env C is the only ESP32-C3 deployment and also the smallest room, so chip and room axes are confounded; an S3-in-Env-C control is left to future work.

On these single-antenna nodes the SNR weighting in

Eq. (2) attenuates the noisier phase channel, so amplitude carries most of the signal (§III).

Classifier choice matters. A two-layer MLP (64-64) at the same feature dimensionality on the same calibrated features drops to F1 = 0.675 cross-environment (A $\rightarrow$ B), a 30-point gap behind the Random Forest's 0.978. The MLP overfits in-domain absolute-scale cues that the Random Forest's piecewise partitioning resists. The transfer property holds for the Random Forest, not for deep classifiers on the same features. Recovering it on an MLP is a separate question we do not address.

C. Cross-Generation Test: 802.11n HT20 $\rightarrow$ 802.11ax HE20

Section III-E predicts that the dimensionless ratio in Eq. (1) is invariant to subcarrier count and PHY format. A clean test requires changing those two quantities without touching the room. We do so by replacing the eight ESP32-S3 nodes of Env A with eight ESP32-C6 nodes at identical (x, y) positions: the C6 is a 2.4 GHz-only WiFi 6 part, so the carrier is unchanged and the only physical-layer differences are 802.11n HT20 $\rightarrow$ 802.11ax HE20 and 52 $\rightarrow$ 246 useful subcarriers. The same 12-minute protocol is re-run, and the model is reused across the boundary in both directions.

The abstraction holds in both directions, with C6 $\rightarrow$ S3 cleaner than S3 $\rightarrow$ C6. In the A(C6) row and column of Table II, the C6-trained classifier transfers to S3 at F1 = 0.985 (within 0.024 of the in-session 5-fold reference on A(C6), F1 = 0.961), and the S3-trained classifier transfers to C6 at F1 = 0.896. Both directions are an order of magnitude above warmup-normalized amplitude (0.09 and 0.28). The standard top row of Figure 4 fails entirely in both cross-generation directions (F1 < 0.30), pointing to the temporal-std denominator as the component that carries the signal

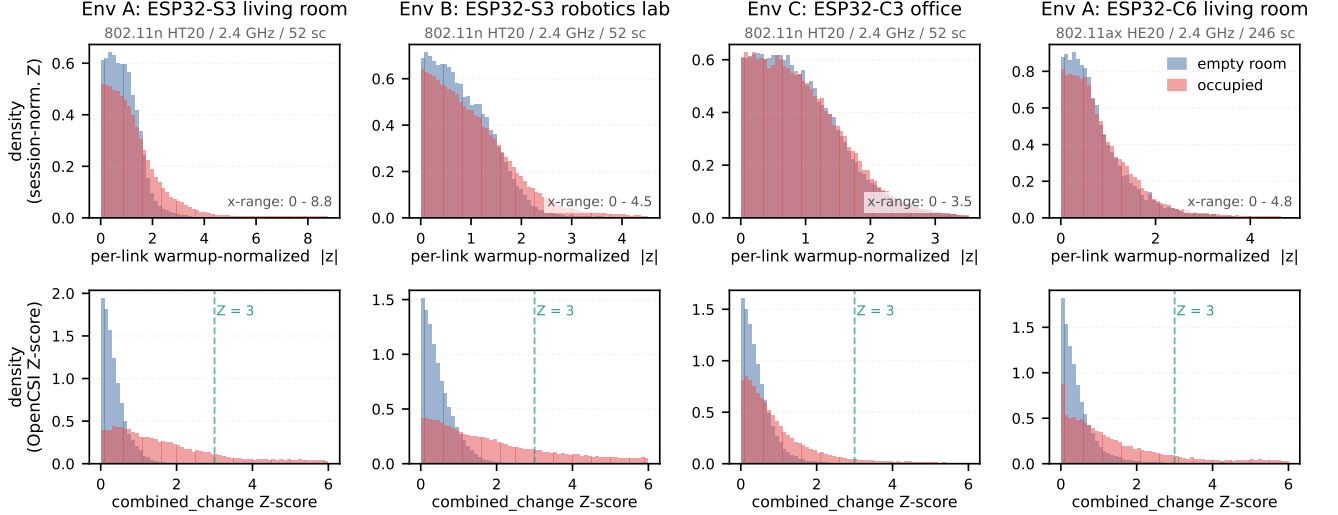


Fig. 4: Z-score distributions across deployments (one column per session). Top: standard session-normalized amplitude Z-score, per-deployment range. Bottom: OpenCSI *combined_change* Z-score, with the empty-vs-occupied split at $z \approx 3$ preserved in every panel. Distributions clipped to the 99.5th percentile.

across the PHY change rather than any incidental cross-PHY similarity.

The asymmetry between the two directions is residual and visible in the rightmost panel of Figure 4: C6 empty-room *combined_change* has a heavier upper tail than S3 ($p_{99.5} = 16.4$ vs. 3.2), so an S3-trained decision boundary occasionally misclassifies C6 quiet frames as occupied; the occupied clusters overlap closely, so the $z \approx 3$ split holds in both directions despite the $4.7\times$ subcarrier-count change. We report this as a single-session pilot.

D. Per-Activity Binary Breakdown

Table III disaggregates binary detection by activity type, revealing where calibration provides the largest gain. Same-PHY $A \rightarrow B$ detects both static and moving occupancy reliably with calibrated features (Cal F1 ≥ 0.959 on all six cells); warmup-normalized amplitude degrades most on empty-vs-moving (0.869 vs. 0.964 in $A \rightarrow B$), consistent with movement-induced variance being more environment-sensitive.

The cross-generation direction $C6 \rightarrow \{A, B\}$ is uniformly strong (Cal F1 ≥ 0.964 on all three tasks); the reverse $\{A, B\} \rightarrow C6$ is asymmetric, with the largest gap on empty-vs-moving (Cal $A \rightarrow C6 = 0.611$, $B \rightarrow C6 = 0.806$), the residual S3 $\rightarrow C6$ scale shift already discussed in §IV-C. On S3 $\rightarrow C3$, calibrated $B \rightarrow C$ empty-vs-moving F1 falls to 0.467 (vs. 0.593 standard) under the same C3-narrow-range mechanism; $A \rightarrow C$ and the reverse directions transfer cleanly.

The reach of the abstraction stops at sign-thresholding: three-class (empty/static/moving) macro F1 falls below the binary threshold cross-environment because that task requires the absolute magnitude of the deviation, which the

denominator in Eq. (1) absorbs by design (§V, sign-vs-magnitude).

E. Calibration Drift: Detection and Online Recovery

A deployed mesh's calibration can go stale (a node restart, furniture motion, or a baseline migrated from a structurally similar room). A baseline-swap ablation on $A \rightarrow B$ replacing Env B's per-link baseline with a pooled Env A one collapses binary F1 from 0.98 to 0.25 with empty-class recall at 0.5%: the baseline mean offset swamps any real perturbation, so every B frame looks uniformly active.

The collapse is detectable on-device: *calibration_maturity* (§III) drops below the ~ 0.65 operational threshold (§V) within seconds, so the mesh rejects on failure rather than silently producing wrong outputs. Recovery is fast: rebuilding the per-link baseline from 30 s of fresh quiet observation restores F1 to 0.98 (15 s already gives 0.976).

Per-deployment calibration is required but recoverable in 30 s, so staleness is a transient fault under *calibration_maturity* monitoring. Because each of the 56 directed links carries its own *reliability* score (§III), and *calibration_maturity* is their weighted mean, a drop also pinpoints *which* links have degraded.

V. DEPLOYMENT ANALYSIS

A. Calibration Maturity and Warmup Time

How long must a node observe before its Z-scores are usable? Fig. 5 plots cross-environment F1 ($A \rightarrow B$) against warmup duration alongside the frame-level *calibration_maturity* score.

Maturity reaches ~ 0.65 after 5 minutes (A: 0.67, B: 0.65, C: 0.62), once the per-link Welford accumulators are stable.

TABLE III: Per-activity binary F1-score (RF) along three transfer axes: cross-environment same-PHY ($A \leftrightarrow B$), cross-generation in the same room ($A \leftrightarrow A(C6)$), and joint cross-environment + cross-generation ($B \leftrightarrow A(C6)$). Bold cells mark the strongest calibrated cross-generation entries.

Task	A $\leftrightarrow$ B (cross-env, S3 $\leftrightarrow$ S3')				A $\leftrightarrow$ A(C6) (cross-generation)				B $\leftrightarrow$ A(C6) (cross-env + cross-gen)			
	Cal A $\rightarrow$ B	Cal B $\rightarrow$ A	Std A $\rightarrow$ B	Std B $\rightarrow$ A	Cal A $\rightarrow$ C6	Cal C6 $\rightarrow$ A	Std A $\rightarrow$ C6	Std C6 $\rightarrow$ A	Cal B $\rightarrow$ C6	Cal C6 $\rightarrow$ B	Std B $\rightarrow$ C6	Std C6 $\rightarrow$ B
Empty vs. static	0.986	0.987	0.900	0.956	0.717	0.968	0.512	0.409	0.935	0.993	0.421	0.412
Empty vs. moving	0.964	0.959	0.869	0.880	0.611	0.964	0.659	0.572	0.806	0.979	0.668	0.482
Empty vs. all occupied	0.982	0.978	0.874	0.923	0.703	0.980	0.575	0.359	0.880	0.990	0.457	0.370

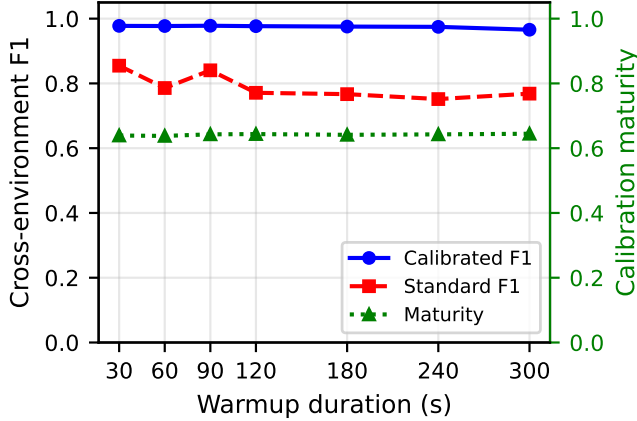


Fig. 5: Cross-environment F1 and OpenCSI calibration maturity vs. quiet warmup duration. The 0.65 maturity plateau is empirical to our three environments.

Calibration state persists to disk, so warmup is paid once per deployment, not per restart.

B. Sign vs. Magnitude

The temporal-std denominator in Eq. (1) absorbs the gain that a particular chip and geometry apply to channel fluctuations, so what crosses any fixed threshold on the ratio is whether a perturbation occurred rather than how large it was in absolute units. Empty-vs-occupied detection therefore reduces to a sign-style threshold and transfers across chips and rooms, whereas static-vs-moving discrimination depends on residual magnitudes that still scale with the chip noise floor and room geometry and does not.

The empirical pattern matches the prediction: empty $z_{\max}$ holds at 1.17, 1.85, and 1.90 across environments A, B, and C, well below any occupied state, while static and moving $z_{\max}$ scale with room and chip by up to a factor of four yet preserve the static > moving ordering within each deployment. Magnitude-sensitive tasks such as per-subcarrier vector matching, quantile alignment, and target-side re-anchoring remain for future work.

C. Deployment and Scaling

Bringing OpenCSI online takes three steps: power on the mesh in an empty room, wait until maturity climbs above 0.65 (about 5 minutes), and let the library checkpoint to disk. The sensing layer then reads (z, r) tuples per link. A maturity drop flags drift (a node restart, a moved baseline,

rearranged furniture), and 30 s of fresh quiet observation rebuilds the affected statistics (§IV-E).

Quiet periods are the resource the method consumes. Office and residential deployments provide them at night and during lunch hours; the time-of-day buckets absorb diurnal thermal drift. Server rooms and factories lack natural quiet windows and would need opportunistic low-motion gating or a pre-loaded baseline. Slow large-object motion (an open door, moved furniture) sits between the variance and spectral-entropy gates and can be absorbed into the baseline; the maturity drop flags it but does not recover it.

Compute scales with link count, not subcarrier count. Per-frame work is $O(K)$ to average across K subcarriers plus a constant per-link update, so per-node cost grows as $O(K(N-1))$ at mesh size N . Memory is a few kilobytes at $N=8$ and stays under 1 MB at $N=50$. Per-link statistics update locally; the bottleneck at scale is mesh airtime, not calibration, and how to allocate that airtime across links has its own dilution-vs-coverage tradeoff that we examine separately [18].

The per-link signal does not depend strongly on chip-specific physics. Weights swap across the S3 $\leftrightarrow$ C6 hot-swap at identical positions, with C6 $\rightarrow$ S3 transferring at F1 = 0.985 and the reverse at 0.896 (§IV-C). Because drift is on-device-detectable, deployments can declare a maturity threshold below which the sensing layer abstains rather than fail silently.

D. Limitations

Our empirical evaluation covers binary presence and three-class activity on ESP32 nodes at 2.4 GHz, using tree-based learners over amplitude-dominated single-antenna features. Richer outputs like multi-person counting and localization remain open, and the HAL admits parsers for Intel 5300, Nexmon, and PicoScenes that we have not yet exercised in this work.

5/6 GHz operation and multi-antenna platforms are next. On the latter, the differential phase becomes useful as antenna ratios cancel its noise sources [27]. The 120 s bootstrap assumes a known-empty room, and the calibration-maturity signal reports when that assumption is violated, with longer-horizon characterization of multi-day stability and run-to-run variance across source-target pairings left to longer-trace deployments.

VI. CONCLUSION

OpenCSI is an abstraction layer between hardware and sensing logic: each link is exposed as a single dimensionless

Z-score (amplitude divided by its own quiet-period temporal standard deviation), so the sensing layer above need not model chip- or room-specific physics. The abstraction is approximate. The temporal-std denominator divides out chip gain but not noise-floor differences, which compresses the ratio on noisier silicon (§IV). The transfer covers binary sign-thresholding (seven of eight cells, $F1 \geq 0.89$), not tasks needing the absolute magnitude of the deviation. Released under Apache 2.0 with full evaluation data.

Future work fills the experimental grid, runs non-ESP32 silicon, and bounds multi-day stability. Magnitude-sensitive tasks (per-subcarrier vectors, quantile matching, target-side re-anchoring) need complementary mechanisms that the dimensionless Z-score does not provide.

ACKNOWLEDGMENTS

AI tools were used for spelling, grammar, style, and code assistance in prototype development. The conceptualization, methodological design, and interpretation of results originated from the authors, while AI tools served only as auxiliary instruments for language refinement and implementation support.

REFERENCES

- [1] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, and A. Javed, "Occupancy detection in non-residential buildings: A survey and novel privacy preserved occupancy monitoring solution," *Applied Sciences*, Vol. 11, No. 24, p. 11856, 2021.
- [2] M. Atif, S. Muralidharan, H. Ko, and B. Yoo, "ESP32-based low-cost WiFi sensing platform," in *Proc. International Conference on Information Networking (ICOIN)*. IEEE, 2020, pp. 772–777.
- [3] Z. Chen, L. Zhang, C. Jiang, J. Cao, and W. Cui, "Cross-domain WiFi sensing with channel state information: A survey," *ACM Computing Surveys*, Vol. 56, No. 1, pp. 1–35, 2023.
- [4] M. Cominelli, F. Gringoli, and F. Restuccia, "AX-CSI: Enabling CSI extraction on commercial 802.11ax Wi-Fi platforms," in *Proc. 15th ACM Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (WiNTECH)*. ACM, 2021, pp. 46–53.
- [5] B. Dong, V. Prakash, F. Feng, and Z. O'Neill, "A review of smart building sensing system for better indoor environment control," *Energy and Buildings*, Vol. 199, pp. 29–46, 2019.
- [6] Espressif Systems, "ESP-CSI: ESP32 Wi-Fi channel state information," <https://github.com/espressif/esp-csi>, 2023.
- [7] —, "ESP-IDF v5.5 Wi-Fi CSI API: HE-LTF Acquisition on ESP32-C6," <https://docs.espressif.com/projects/esp-idf/en/v5.5/esp32c6/api-guides/wifi.html>, 2025.
- [8] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *Proc. 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (WiNTECH)*. ACM, 2019, pp. 21–28.
- [9] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, p. 53, 2011.
- [10] IEEE Standards Association, "IEEE 802.11bf: WLAN sensing," IEEE 802.11 Task Group bf, 2024, https://www.ieee802.org/11/Reports/tgb_f_update.htm.
- [11] C. Jiang, Y. Yan, Y. Wang, C. T. Chou, and W. Hu, "Scale what counts, mask what matters: Evaluating foundation models for zero-shot cross-domain Wi-Fi sensing," *arXiv preprint arXiv:2511.18792*, 2025.
- [12] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition," in *Proc. 24th Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM, 2018, pp. 289–304.
- [13] Z. Jiang, T. H. Luan, X. Ren, D. Lv, H. Hao, J. Wang, K. Zhao, W. Xi, Y. Xu, and R. Li, "PicoScenes: A versatile Wi-Fi channel state information toolkit for IEEE 802.11n/ac/ax sensing applications," *ACM Transactions on Sensor Networks*, Vol. 18, No. 2, pp. 1–36, 2022.
- [14] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Computing Surveys*, Vol. 52, No. 3, pp. 1–36, 2019.
- [15] K. O. E. Müller, D. Datsomor, D. Schumm, B. Rodrigues, and B. Stiller, "Big Brother is Watching You: Non-Intrusive ZigBee User Profiling," in *Proc. 20th International Conference on Network and Service Management (CNSM)*. IEEE, 2024, pp. 1–7.
- [16] V. V. Ratnam, H. Chen, H. H. Chang, A. Sehgal, and J. C. Zhang, "Optimal preprocessing of WiFi CSI for sensing applications," *IEEE Transactions on Wireless Communications*, Vol. 23, No. 9, pp. 10820–10833, 2024, arXiv:2307.12126.
- [17] B. Rodrigues, C. Halter, M. Franco, E. J. Scheid, C. Killer, and B. Stiller, "BluePIL: a Bluetooth-based Passive Localization Method," in *Proc. IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2021, pp. 28–36.
- [18] B. Rodrigues and K. Khamaisi, "Less is More: The Dilution Effect in Multi-Link Wireless Sensing," 2026.
- [19] C. Turetta, G. Skenderi, L. Capogrosso, F. Demrozi, P. H. Kindt, A. Masrur, F. Fummi, M. Cristani, and G. Pravadelli, "Towards deep learning-based occupancy detection via WiFi sensing in unconstrained environments," in *Proc. Design, Automation and Test in Europe Conference (DATE)*. IEEE, 2023.
- [20] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial WiFi devices," *IEEE Journal on Selected Areas in Communications*, Vol. 35, No. 5, pp. 1118–1131, 2017.
- [21] X. Wang, K. Niu, J. Xiong, B. Qian, Z. Yao, T. Lou, and D. Zhang, "Placement matters: Understanding the effects of device placement for WiFi sensing," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 6, No. 1, pp. 1–25, 2022.
- [22] B. P. Welford, "Note on a method for calculating corrected sums of squares and products," *Technometrics*, Vol. 4, No. 3, pp. 419–420, 1962.
- [23] J. Wilson and N. Patwari, "Radio Tomographic Imaging with Wireless Networks," *IEEE Transactions on Mobile Computing*, Vol. 9, No. 5, pp. 621–632, 2010.
- [24] J. Yang, X. Chen, D. Wang, H. Zou, C. X. Lu, S. Sun, L. Xie, and Y. Lu, "SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing," *Patterns*, Vol. 4, No. 3, p. 100703, 2023.
- [25] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression," *IEEE Internet of Things Journal*, Vol. 9, No. 15, pp. 13 086–13 095, 2022.
- [26] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, Vol. 55, No. 10, pp. 98–104, 2017.
- [27] D. Zhang, Y. Zeng, D. Wu, and J. Xiong, "Boosting WiFi sensing performance via CSI ratio," *IEEE Pervasive Computing*, Vol. 20, No. 1, pp. 62–70, 2021.
- [28] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "CrossSense: Towards cross-site and large-scale WiFi sensing," in *Proc. 24th Annual International Conference on Mobile Computing and Networking (MobiCom)*. ACM, 2018, pp. 305–320.
- [29] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0: Zero-effort cross-domain gesture recognition with WiFi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 11, pp. 8671–8688, 2022.
- [30] R. Zhou, Y. Cheng, S. Li, H. Zhang, and C. Liu, "DGSense: A Domain Generalization Framework for Wireless Sensing," *arXiv preprint arXiv:2502.08155*, 2025.
- [31] G. Zhu, Y. Hu, W. Gao, W.-H. Wang, B. Wang, and K. J. R. Liu, "CSI-Bench: A large-scale in-the-wild dataset for multi-task WiFi sensing," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2025, arXiv:2505.21866.