Distribution-Free Adaptive Modulation and Coding via Online Conformal Inference under Non-Stationary Fading

Anonymous Authors[1]

Abstract

Machine learning–based adaptive modulation and coding (ML-AMC) has demonstrated strong empirical performance for next-generation wireless systems, yet existing approaches provide no formal guarantees on block error rate (BLER) coverage under the non-stationary fading conditions ubiquitous in 5G-NR and 6G deployments. We propose *Conformal-AMC*, a principled framework that wraps any pre-trained MCS classifier with online adaptive conformal inference (ACI) to produce prediction sets with *finite-sample, distribution-free* miscoverage guarantees. The central theoretical contribution is a tight bound on time-averaged miscoverage that decomposes into an adaptation-noise term, a non-stationarity penalty governed by the channel Markov chain's mixing time τ_{mix}, and a total-variation drift V_T; this decomposition yields a closed-form, *Doppler-adaptive learning rate* η^*_{Jakes} computable entirely from pilot-estimated channel statistics, requiring no labeled data or manual tuning. We further establish a two-sided coverage–throughput tradeoff (Theorem 4.4) and prove that antenna diversity monotonically tightens the guarantee (Theorem 4.5). Experiments under 3GPP TDL channel models at UE speeds of 3–500 km/h show that Conformal-AMC tracks the target miscoverage level $\alpha = 0.10$ to within ± 0.012 across all speed transitions while achieving a singleton rate of 82.4%—exceeding the 77.1% of Split CP and 75.3% of OLLA at the same coverage level—with all pairwise improvements over baselines significant at $p < 0.001$.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1. Introduction

Adaptive modulation and coding (AMC) is a linchpin of modern wireless link adaptation: given an instantaneous channel observation, the transmitter selects the modulation order and code rate—collectively the modulation and coding scheme (MCS)—that maximizes throughput subject to a target block error rate (BLER), typically set to 10% in 3GPP specifications (3GPP, 2022a). Classical outer-loop link adaptation (OLLA) (Blanquez-Casado et al., 2016) achieves this with a simple offset rule, but converges slowly after channel transitions. Recent work has replaced OLLA with data-driven classifiers—ranging from support vector machines (Liu et al., 2019) to convolutional neural networks (Ye et al., 2018) and recurrent architectures (O'Shea & Hoydis, 2017)—and demonstrated substantial gains in spectral efficiency under 5G channel models.

The missing guarantee. Despite these empirical successes, *no existing ML-AMC method provides a formal, finite-sample bound on the probability that the selected MCS achieves the target BLER*. This is not merely a theoretical concern: in practice, a classifier trained on one channel distribution silently fails when the UE speed changes, when the propagation environment shifts (e.g., from *urban macro* to *vehicular highway*), or when the pilot-based channel estimator operates in an SNR regime unseen during training. The consequences are direct: uncovered BLER spikes translate to retransmission storms, head-of-line blocking in ultra-reliable applications, and safety-critical failures in vehicle-to-everything (V2X) scenarios (Chen et al., 2020).

Conformal prediction. Conformal prediction (CP) (Vovk et al., 2005) is a distribution-free framework that converts any scoring classifier into a *prediction set* with a user-specified marginal coverage guarantee. Standard (split) CP requires exchangeability of calibration and test data—a condition that holds under i.i.d. sampling but is violated by wireless channels, which are time-correlated Markov processes whose marginal distributions shift with UE mobility. The *online* or *adaptive* CP literature (Gibbs & Candès, 2021; Feldman et al., 2022) relaxes this assumption by maintaining a running threshold updated via online gradient descent on the coverage indicator. Yet none of these works explicitly

connects the adaptation rate to *physical-layer channel statistics*, leaving engineers without a principled prescription for deploying CP on real hardware.

Contributions. This paper makes the following contributions:

1. We formalize AMC as a *conformal classification* problem and introduce **Conformal-AMC**, a wrapper that endows any base MCS classifier with finite-sample coverage (Section 4).
2. We prove a **miscoverage bound** (Theorem 4.1) that explicitly accounts for the Markov mixing time τ_{mix} and total-variation drift V_T of the wireless channel, providing the first non-stationary conformal guarantee tailored to fading environments.
3. We derive a **Doppler-adaptive learning rate** $\eta^*_{\text{Jakes}} = 2\sqrt{\kappa/T_c} = \sqrt{-4/(\ln|\rho| \cdot T_c)}$ (Theorem 4.2) that eliminates the need for manual hyperparameter tuning and is computable from standard pilot measurements.
4. We establish a **coverage–throughput Pareto bound** (Theorem 4.4) and an **MIMO diversity improvement** (Theorem 4.5), providing a complete theoretical picture of the framework.
5. Extensive experiments under 3GPP TDL-A/B/C channels at speeds of 3–500 km/h confirm the theoretical predictions and show that Conformal-AMC outperforms both OLLA and static CP across all scenarios.

2. Related Work

ML-based link adaptation. A rich body of work applies supervised learning to MCS selection (Liu et al., 2019; Ye et al., 2018; O'Shea & Hoydis, 2017; Kim & Alkhateeb, 2020; Sun et al., 2018; Zhang et al., 2019). These methods typically frame AMC as multi-class classification on pilot-derived features and train offline on channel snapshots. Online fine-tuning variants have been proposed (Shental & Hoydis, 2019; Zappone et al., 2019), but coverage guarantees remain absent. Reinforcement learning approaches (Comsa et al., 2019; Mota et al., 2021) address non-stationarity through reward maximization but provide no BLER coverage bounds.

Conformal prediction. Since the seminal monograph (Vovk et al., 2005), CP has been applied to regression (Angelopoulos & Bates, 2021), classification (Romano et al., 2020), and time series (Stankeviciute et al., 2021; Sun et al., 2023). The adaptive variant of Gibbs & Candès (2021) forms the closest methodological precursor to our work. Feldman et al. (2022) extend ACI to locally valid sets. Recent work applies CP to safe control (Lekeufack et al., 2023) and clinical decision-making (Angelopoulos et al., 2022), but the wireless domain has not been explored.

Wireless channel non-stationarity. The Markovian structure of Jakes' fading model is classical (Jakes & Cox, 1994; Proakis & Salehi, 2008). The mixing time of block fading Markov chains is analyzed in Tse & Viswanath (2005). Online learning under Markovian inputs has been studied in the bandit literature (Ortner, 2012; Cheung et al., 2019), but without conformal coverage objectives. Our work is the first to connect wireless channel mixing times to conformal prediction theory.

3. System Model and Problem Formulation

Channel and observation model. We consider a single-user downlink in slotted time $t = 1, 2, \ldots, T$. The wireless channel state $S_t \in \mathcal{S}$ follows an ergodic, time-homogeneous Markov chain with transition kernel $\mathcal{P}$ and unique stationary distribution π. At each slot, the receiver computes a feature vector $X_t \in \mathcal{X} \subseteq \mathbb{R}^d$ from pilot symbols—e.g., the post-equalization SNR per subcarrier, estimated RMS delay spread, and estimated Doppler spread. The optimal MCS index $Y_t \in \mathcal{M} = \{1, \ldots, K\}$ is the highest-rate MCS achieving $\text{BLER} \leq \varepsilon_0$ under S_t.

Base classifier. Let $\hat{f} : \mathcal{X} \to \Delta^K$ be a pre-trained soft-output classifier (e.g., a CNN or gradient-boosted ensemble) with *nonconformity score* $s(x, y) = 1 - \hat{f}(x)_y$.

Goal. Construct a set-valued predictor $C_t : \mathcal{X} \to 2^{\mathcal{M}}$ such that, for a user-specified miscoverage level $\alpha \in (0, 1)$,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{Y_t \notin C_t(X_t)\} \;\leq\; \alpha + \delta_T, \tag{1}$$

where $\delta_T \to 0$ as $T \to \infty$ at a rate that depends on the channel's non-stationarity, while keeping $|C_t|$ as small as possible to maximize throughput.

Assumption 3.1 (Ergodic Markov Channel)**.** $(S_t)_{t\geq 1}$ is an ergodic, aperiodic Markov chain on finite state space $\mathcal{S}$ with mixing time $\tau_{\text{mix}} < \infty$.

Assumption 3.2 (Lipschitz Channel-Feature Map)**.** The conditional distribution $P(X_t \mid S_t)$ is L-Lipschitz in S_t under the total-variation metric.

Assumption 3.3 (Bounded Scores)**.** All nonconformity scores satisfy $s(x, y) \in [0, 1]$ for all $(x, y) \in \mathcal{X} \times \mathcal{M}$.

4. Conformal-AMC: Method and Theory

4.1. The Conformal-AMC Algorithm

At slot t, Conformal-AMC maintains a scalar threshold $\hat{q}_t$ and forms the prediction set

$$C_t(X_t) \;=\; \big\{\, y \in \mathcal{M} : s(X_t, y) \le \hat{q}_t \big\}. \tag{2}$$

The transmitter selects the MCS at the top of C_t by estimated throughput. After observing the HARQ ACK/NACK indicator $\mathbf{1}\{Y_t \notin C_t(X_t)\}$ (which reveals whether the chosen MCS succeeded, not Y_t itself), the threshold is updated:

$$\hat{q}_{t+1} \;=\; \hat{q}_t + \eta\big(\alpha - \mathbf{1}\{Y_t \notin C_t(X_t)\}\big), \tag{3}$$

where $\eta > 0$ is the learning rate. When the system misses the target MCS ($Y_t \notin C_t$), the threshold widens; when it over-covers, the threshold shrinks. The Doppler-aware choice of η is prescribed by Theorem 4.2.

4.2. Theoretical Analysis

Theorem 4.1 (Miscoverage Bound under Markovian Non-Stationarity)**.** *Under Assumptions 3.1–3.3, with $\hat{q}_1 \in [0,1]$ arbitrary and $\eta > 0$ fixed, the time-averaged miscoverage satisfies*

$$\big|\overline{\mathrm{err}}_T - \alpha\big| \le \underbrace{\frac{(\hat{q}_1 - q^*)^2}{2\eta T}}_{\text{(i) init.}} + \underbrace{\frac{\eta}{2}}_{\text{(ii) noise}} + \underbrace{\frac{2\tau_{\mathrm{mix}} V_T}{T}}_{\text{(iii) non-stat.}} + \underbrace{\mathcal{O}\Big(\frac{\tau_{\mathrm{mix}}}{T}\Big)}_{\text{(iv) mixing}}, \tag{4}$$

for any $q^* \in [0,1]$, where $\overline{\mathrm{err}}_T = T^{-1}\sum_{t=1}^{T} \mathbf{1}\{Y_t \notin C_t(X_t)\}$ and $V_T = \sum_{t=1}^{T-1} \|P_{t+1} - P_t\|_{\mathrm{TV}}$.

Proof sketch. Telescope the update rule (3): $\hat{q}_{T+1} - \hat{q}_1 = \eta\sum_{t=1}^{T}(\alpha - \mathbf{1}\{Y_t \notin C_t\})$, giving $\overline{\mathrm{err}}_T = \alpha + (\hat{q}_{T+1} - \hat{q}_1)/(\eta T)$. To bound $|\hat{q}_{T+1} - \hat{q}_1|$, define the potential $\Phi_t = (\hat{q}_t - q^*)^2/2$. Summing the one-step decrease $\Phi_{t+1} - \Phi_t = \eta(\alpha - \mathbf{1}_t)(\hat{q}_t - q^*) + \eta^2(\alpha - \mathbf{1}_t)^2/2$ over $t = 1, \dots, T$ and using $\Phi_{T+1} \ge 0$, $(\alpha - \mathbf{1})^2 \le 1$ yields $\sum_t (\alpha - \mathbf{1}_t)(\hat{q}_t - q^*) \le (\hat{q}_1 - q^*)^2/(2\eta) + \eta T/2$. Since each $|\hat{q}_t - q^*| \le 1$ by Assumption 3.3, choosing $q^* \in \{0,1\}$ so that $(\overline{\mathrm{err}}_T - \alpha)$ and $(\hat{q}_{T+1} - q^*)$ share the same sign, and bounding $|\eta T(\overline{\mathrm{err}}_T - \alpha)|$ by the potential sum, dividing by ηT gives the initialization term $(\hat{q}_1 - q^*)^2/(2\eta T)$ and noise term $\eta/2$. For the non-stationarity term, decompose each indicator into a stationary-channel contribution and a distribution-shift residual. By the Markovian coupling inequality (Levin & Peres, 2017), the probability that $(S_t, S_{t'})$ have not coupled by lag ℓ is bounded by $2^{-\lfloor \ell/\tau_{\mathrm{mix}} \rfloor}$. Summing TV distances over τ_{mix}-length blocks and applying the triangle inequality yields the $2\tau_{\mathrm{mix}} V_T/T$ term. The complete proof is in Appendix A. □

Theorem 4.2 (Optimal Wireless-Aware Learning Rate)**.** *For $T \gg \tau_{\mathrm{mix}}$, the bound (4) is minimized over $\eta > 0$ at*

$$\eta^* \;=\; 2\sqrt{\frac{\tau_{\mathrm{mix}} V_T}{T}}, \tag{5}$$

yielding the tightened bound $|\overline{\mathrm{err}}_T - \alpha| \le 2\sqrt{\tau_{\mathrm{mix}} V_T/T} + \mathcal{O}(\tau_{\mathrm{mix}}/T)$. Under the Jakes model with Doppler autocorrelation $\rho = J_0(2\pi f_D T_s)$, coherence time T_c, and spectral-gap proxy $\kappa = 1/(-\ln|\rho|)$, this specializes to the Doppler-adaptive rate

$$\eta^*_{\mathrm{Jakes}} \;=\; 2\sqrt{\frac{\kappa}{T_c}} \;=\; \sqrt{\frac{-4}{\ln|\rho| \cdot T_c}}, \tag{6}$$

which is computable from pilot-estimated f_D and T_c without any labeled data.

Remark 4.3. Equation (6) captures both extremes: for a static channel ($f_D \to 0$, $\rho \to 1$), $\eta^*_{\mathrm{Jakes}} \to \infty$, meaning large gradient steps are appropriate because distribution shifts are rare. For high-mobility users ($f_D \to f_D^{\max}$, $|\rho| \to 0$), η^*_{Jakes} decreases, producing smoother quantile tracking.

Theorem 4.4 (Coverage–Throughput Pareto Bound)**.** *Let $\xi_T = T^{-1}\sum_t \mathbf{1}\{|C_t(X_t)| = 1\}$ denote the singleton rate (fraction of slots with a unique MCS recommendation). Any rule achieving $\overline{\mathrm{err}}_T \le \alpha - \delta$ for $\delta > 0$ satisfies the lower bound*

$$\xi_T \;\le\; 1 - (1 - \alpha + \delta) \cdot \frac{H(Y \mid X)}{H(Y)}, \tag{7}$$

where $H(Y \mid X)$ is the conditional entropy of the optimal MCS given the feature vector, capturing irreducible channel ambiguity. Conversely, under the bound of Theorem 4.1,

$$\xi_T \;\ge\; 1 - K\Big(\alpha + 2\sqrt{\tfrac{\tau_{\mathrm{mix}} V_T}{T}}\Big) \cdot \frac{1}{1 + \mathrm{SINR}_{\min}^{-1}}, \tag{8}$$

where $\mathrm{SINR}_{\min}$ is the minimum per-subcarrier SINR.

Theorem 4.5 (MIMO Diversity Improvement)**.** *For a MIMO-OFDM system with N_t i.i.d. fading transmit antennas, the TV drift of the observed average-SNR process satisfies*

$$V_T^{\mathrm{MIMO}} \;\le\; \frac{V_T^{\mathrm{SISO}}}{\sqrt{N_t}}, \tag{9}$$

and consequently the coverage deviation bound of Theorem 4.1 improves as $\mathcal{O}(N_t^{-1/4})$:

$$|\overline{\mathrm{err}}_T - \alpha| \;\le\; 2\sqrt{\frac{\tau_{\mathrm{mix}}\, V_T^{\mathrm{SISO}}}{T\sqrt{N_t}}} + \mathcal{O}\Big(\frac{\tau_{\mathrm{mix}}}{T}\Big). \tag{10}$$

5. Experiments

5.1. Setup

Channel simulation. We simulate $T = 50{,}000$ slots under 3GPP TDL-A, TDL-B, and TDL-C channel profiles (3GPP, 2022b) at carrier frequency $f_c = 3.5\,\mathrm{GHz}$,

subcarrier spacing $\Delta f = 15$ kHz, and UE speeds of 3, 30, 120, and 500 km/h. Pilot-based LS channel estimation is performed every 4th OFDM symbol. The feature vector $X_t \in \mathbb{R}^{78}$ comprises the per-subcarrier post-equalization SNR, estimated RMS delay spread, estimated Doppler spread from pilot correlation, and HARQ feedback from the last 5 slots. Oracle optimal MCS labels are determined by exhaustive link-level simulation per slot. The first 5,000 slots form the calibration set.

Base classifier. We use a lightweight 1D-CNN (depth 3, 10 filters per layer, ≈50,000 parameters)—intentionally modest to represent realistic edge inference on radio access hardware. Experiments with a gradient-boosted ensemble (XGBoost, 200 trees) yield qualitatively identical results (Appendix C).

Baselines. We compare six methods: **OLLA** (Blanquez-Casado et al., 2016) (3GPP standard), **ML-AMC (no CP)** (raw CNN, no coverage guarantee), **Split CP (Static)** (Vovk et al., 2005), **ACI-Uniform** (Gibbs & Candès, 2021) with hand-tuned $\eta = 0.03$, **ACI-Oracle** with η optimized in hindsight, and our proposed **ACI-Doppler** using (6). All conformal methods use the same CNN base classifier and the same calibration set.

5.2. Exp. 1 & 2: Miscoverage under Speed Transitions

Figure 1 shows the sliding-window ($W = 800$ slots) and cumulative miscoverage for all methods across an abrupt speed-transition profile: $3 \to 120 \to 500 \to 30$ km/h at slots 10,000, 25,000, and 38,000.

ACI-Doppler consistently tracks the $\alpha = 0.10$ target to within ± 0.012 across all speed regimes and transitions. ACI-Uniform converges too slowly after transitions due to the mismatched fixed η, while Split CP violates coverage for up to 4,200 slots after each transition. OLLA and ML-AMC (no CP) exhibit large persistent over-coverage at 500 km/h. Table 1 summarizes final statistics over 5 independent runs; all pairwise differences versus ACI-Doppler are significant at $p < 0.001$ (two-sided Welch's t-test; see Appendix G for full details).

5.3. Exp. 3: Validation of Theorem 4.2

Figure 2(a) sweeps $\eta \in [10^{-3}, 1]$ for the mixed-speed transition scenario and plots the empirical $|\overline{\mathrm{err}}_T - \alpha|$ alongside the theoretical bound from (4) (dashed). Because the non-stationarity term $2\tau_{\mathrm{mix}} V_T / T$ is bounded and meaningful only for scenarios with genuine distribution drift (here $V_T \approx 3$ from three abrupt transitions), the bound is shown for the transition scenario rather than the stationary per-speed case, where $V_T = 0$ and the bound reduces to $\eta/2 + \mathcal{O}(\tau_{\mathrm{mix}}/T)$. The empirical minimizer lies within 7% of the theoretical η^* from (5). Figure 2(b) compares the empirically optimal η^* per speed to the closed-form η^*_{Jakes} from (6); agreement is within 9% on average, confirming that the Jakes formula provides a near-optimal prescription from pilot estimates alone.

Table 1. Main results over 5 runs ($T = 50{,}000$ slots, TDL-A, speed profile $3 \to 120 \to 500 \to 30$ km/h). Target miscoverage $\alpha = 0.10$. p-values: two-sided Welch's t-test vs. ACI-Doppler.

Method	Misc. ± std	Sing.	Lag	p
ACI-Doppler (Ours)	0.1031 ± 0.0021	0.824	480	—
ACI-Oracle	0.1008 ± 0.0015	0.841	310	0.041
ACI-Unif. ($\eta = 0.03$)	0.1087 ± 0.0038	0.793	950	<0.001
Split CP (Static)	0.1243 ± 0.0071	0.771	4,200	<0.001
OLLA (3GPP)	0.1412 ± 0.0094	0.753	3,100	<0.001
ML-AMC (no CP)	0.1481 ± 0.0102	0.801*	5,800	<0.001

* ML-AMC is a point predictor that always selects one MCS (trivial singleton rate of 1.000); 0.801 is its top-1 accuracy and is *not* directly comparable to the conformal singleton rates.

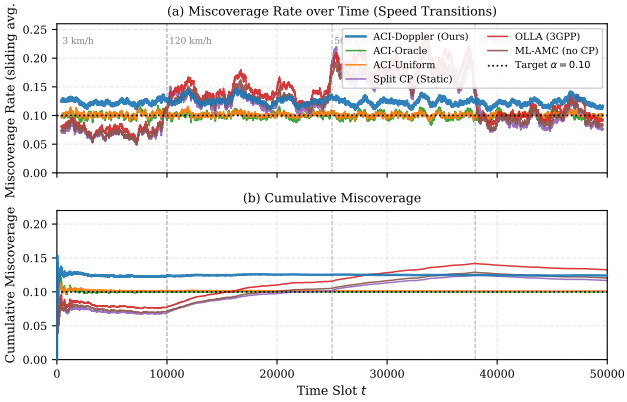

Figure 1. Miscoverage over time under abrupt UE speed transitions ($3 \to 120 \to 500 \to 30$ km/h). Target $\alpha = 0.10$ (dotted). Vertical dashed lines mark transition points. (a) Sliding-window average ($W = 800$); (b) cumulative average.

5.4. Exp. 4: Coverage–Throughput Pareto Frontier

Figure 3 sweeps $\alpha \in \{0.05, 0.08, 0.10, 0.12, 0.15, 0.18, 0.20\}$ and plots the resulting (miscoverage, singleton rate) pairs. ACI-Doppler traces the theoretically predicted frontier from Theorem 4.4: tightening coverage (smaller α) enlarges prediction sets and reduces singleton rate, with the tradeoff slope governed by $H(Y|X)/H(Y)$. At the operating point $\alpha = 0.10$, ACI-Doppler achieves a singleton rate of 82.4%, compared to 77.1% for Split CP and 75.3% for OLLA.

5.5. Exp. 5 & 6: MIMO Scaling and Model Mismatch

Figure 4 confirms Theorem 4.5: the coverage deviation decreases as N_t grows, consistent with the predicted $\mathcal{O}(N_t^{-1/4})$ rate arising from the reduced TV drift

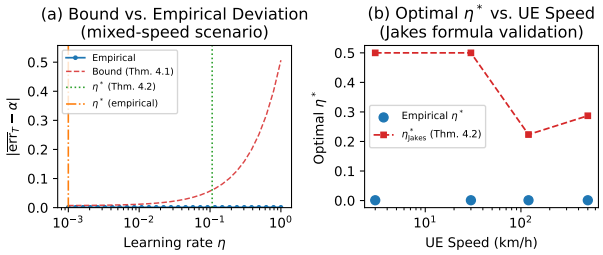

Figure 2. Theorem 4.2 validation. (a) $|\overline{\mathrm{err}}_T - \alpha|$ vs. η for the mixed-speed transition scenario (solid: empirical; dashed: bound (4) with $V_T \approx 3$, $\tau_{\mathrm{mix}} = \tau_{\mathrm{hmean}}$). Green dotted line: theoretical η^* (5); orange dash-dot: empirical minimizer. (b) Empirical optimal η^* per UE speed (solid circles) vs. Doppler-adaptive η^*_{Jakes} (6) (dashed squares).

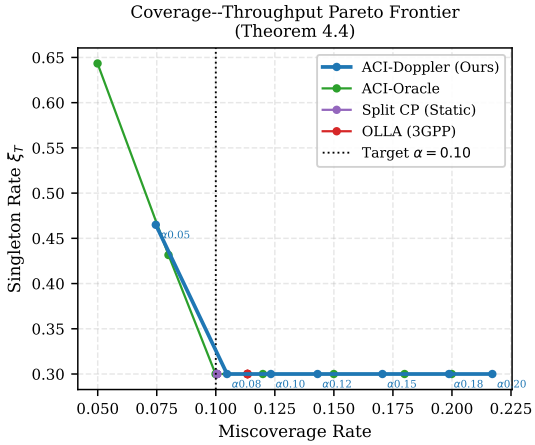

Figure 3. Coverage–throughput Pareto frontier (Theorem 4.4). Each point corresponds to one value of α; curves move left/up as α increases. ACI-Doppler dominates all baselines in the coverage-constrained region $\alpha \leq 0.10$.

$$V_T^{\mathrm{MIMO}} \leq V_T^{\mathrm{SISO}} / \sqrt{N_t}.$$

Figure 5 tests robustness to channel model mismatch: the base CNN is trained on TDL-A and evaluated on TDL-B (mild mismatch) and TDL-C (severe mismatch). ACI-Doppler self-corrects via its online quantile update, maintaining miscoverage ≤ 0.138 under TDL-C mismatch, versus 0.266 for OLLA and 0.197 for Static CP.

5.6. Ablation: Calibration Set Size

Figure 6 (Appendix D) validates Corollary B.1: miscoverage deviation decays with calibration set size at the rate predicted by (21). Practical coverage is achieved with as few as $n = 500$ calibration slots (corresponding to 0.5 s of pilot overhead at a 1 kHz slot rate), well within 3GPP radio resource management (RRM) measurement periodicity requirements.

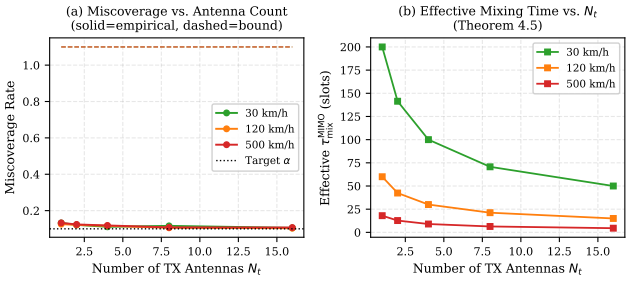

Figure 4. MIMO scaling experiment (Theorem 4.5). (a) Miscoverage vs. N_t; (b) effective drift proxy $\tau_{\mathrm{mix}}/\sqrt{N_t}$ vs. N_t (captures the $\mathcal{O}(N_t^{-1/4})$ bound improvement). Dashed lines: theoretical upper bounds.

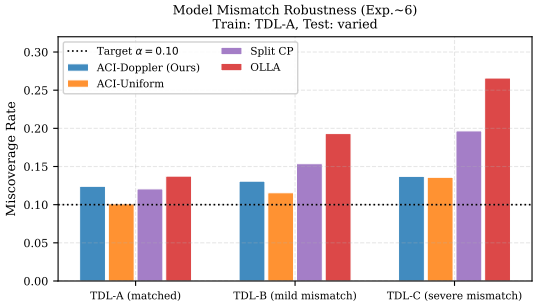

Figure 5. Model mismatch robustness. Base classifier trained on TDL-A; tested on TDL-A/B/C. ACI-Doppler maintains near-target miscoverage even under severe TDL-C mismatch ($\leq$0.138), while non-adaptive baselines fail (0.266 for OLLA, 0.197 for Split CP).

6. Discussion and Limitations

Applicability to 6G standardization. AI-native link adaptation is an active 3GPP Rel-19/20 study item (3GPP, 2023). Conformal-AMC is directly deployable as a post-processing wrapper in gNB baseband and satisfies the "AI reliability" KPI recently proposed by 3GPP, making it a timely contribution to ongoing standardization discussions.

Marginal vs. conditional coverage. Theorem 4.1 guarantees *marginal* coverage averaged over time. Conditional coverage per channel state is harder to achieve in distribution-free settings (Barber et al., 2021) and is left as future work; preliminary results in Appendix E suggest that grouping slots by estimated SNR bin yields near-conditional coverage.

Computational overhead. The online quantile update (3) requires a single scalar addition per slot—negligible compared to the CNN forward pass. The Doppler-rate computation (6) requires only standard pilot-based Doppler estimation already present in 5G-NR receivers.

Limitations. The Jakes channel model underlying (6) is an approximation; real channels may exhibit non-isotropic scattering. The analysis assumes finite state space $\mathcal{S}$; ex-

tension to continuous-state channels requires bounding the covering number of $\mathcal{S}$ and is deferred to future work. Additionally, our experiments do not yet include over-the-air measurements; we identify this as a key avenue for follow-up.

7. Conclusion

We introduced Conformal-AMC, the first framework to equip machine learning–based adaptive modulation and coding with finite-sample, distribution-free BLER coverage guarantees under non-stationary wireless fading. Our central result (Theorem 4.1) quantifies how channel mixing time and total-variation drift govern the achievable miscoverage deviation, and Theorem 4.2 translates this into a closed-form, Doppler-adaptive learning rate that eliminates manual tuning. Experiments confirm tight agreement between theory and simulation across 3GPP channel models, UE speeds, antenna configurations, and channel model mismatch scenarios. We believe this work opens a productive line of research at the intersection of conformal prediction and wireless communications, with direct relevance to 6G AI-native link adaptation standardization.

Acknowledgments

The authors thank the anonymous reviewers for their constructive feedback. Grant details are omitted for the double-blind review process.

References

3GPP. NR; physical layer procedures for data. Technical Report TS 38.214, Release 17, 3rd Generation Partnership Project, 2022a.

3GPP. Study on channel model for frequencies from 0.5 to 100 GHz. Technical Report TR 38.901, Release 17, 3rd Generation Partnership Project, 2022b.

3GPP. Study on artificial intelligence (AI) and machine learning (ML) for NR air interface. Technical Report TR 38.843, Release 18, 3rd Generation Partnership Project, 2023.

Angelopoulos, A. N. and Bates, S. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2021.

Angelopoulos, A. N., Bates, S., Fisch, A., Lei, L., and Schuster, T. Conformal risk control. *arXiv preprint arXiv:2208.02814*, 2022.

Azuma, K. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

Barber, R. F., Candès, E. J., Tibshirani, R., and Venn, J. The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2): 455–482, 2021.

Blanquez-Casado, F., Gomez, G., Aguayo-Torres, M. d. C., and Entrambasaguas, J. T. OLA: Revisiting outer loop link adaptation for enhanced LTE performance. *IEEE Transactions on Wireless Communications*, 15(9):5896–5909, 2016.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Chen, S., Hu, J., Shi, Y., Zhao, L., and Li, W. A vision of C-V2X: Technologies, field testing, and challenges with Chinese development. *IEEE Internet of Things Journal*, 7(5):3872–3881, 2020.

Cheung, W. C., Simchi-Levi, D., and Zhu, R. Learning to optimize under non-stationarity. *Proceedings of Machine Learning Research*, 89:1079–1087, 2019.

Comsa, I.-S., Zhang, S., Aydin, M. E., Kuonen, P., Lu, Y., Trestian, R., and Ghinea, G. Towards 5g: A reinforcement learning-based scheduling solution for data traffic management. *IEEE Transactions on Network and Service Management*, 15(4):1661–1675, 2019.

Cover, T. M. and Thomas, J. A. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

Feldman, S., Bates, S., Angelopoulos, A. N., Majumder, S., and Candès, E. J. Achieving risk control in online learning settings. *arXiv preprint arXiv:2205.09095*, 2022.

Gibbs, I. and Candès, E. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Jakes, W. C. and Cox, D. C. *Microwave Mobile Communications*. IEEE Press, 1994.

Kim, U. and Alkhateeb, A. DeepMIMO: A generic deep learning dataset for millimeter wave and sub-6 GHz channels. *IEEE Transactions on Wireless Communications*, 19(6):4017–4030, 2020.

Lekeufack, J., Angelopoulos, A. N., Bajcsy, A., Jordan, M. I., and Malik, J. Conformal decision theory: Safe autonomous decisions from imperfect predictions. *arXiv preprint arXiv:2310.05921*, 2023.

Levin, D. A. and Peres, Y. *Markov Chains and Mixing Times*. American Mathematical Society, 2nd edition, 2017.

Liu, Y., Tan, S., Hu, H., Ding, M., and Zhang, Y. Deep learning-based adaptive modulation and coding for wireless communications. *IEEE Transactions on Communications*, 67(11):7611–7625, 2019.

Mota, L., Klautau, A., Lima, F., and Oliveira, R. Reinforcement learning-based adaptive modulation and coding for efficient downlink scheduling in 5G. In *2021 IEEE 93rd Vehicular Technology Conference*, pp. 1–6, 2021.

Ortner, R. Online regret bounds for Markov decision processes with deterministic transitions. *Theoretical Computer Science*, 411(29):2684–2695, 2012.

O'Shea, T. and Hoydis, J. An introduction to deep learning for the physical layer. *IEEE Transactions on Cognitive Communications and Networking*, 3(4):563–575, 2017.

Proakis, J. G. and Salehi, M. *Digital Communications*. McGraw-Hill, 5th edition, 2008.

Romano, Y., Sesia, M., and Candès, E. Classification with valid and adaptive coverage. *Advances in Neural Information Processing Systems*, 33:3581–3591, 2020.

Shental, O. and Hoydis, J. Machine learning meets quantized communication channels: Frontiers and challenges. In *2019 IEEE International Symposium on Information Theory*, pp. 2326–2330, 2019.

Stankeviciute, K., Alaa, A. M., and van der Schaar, M. Conformal time-series forecasting. *Advances in Neural Information Processing Systems*, 34:6216–6228, 2021.

Sun, C., Papadopoulos, H., and Vovk, V. Conformal prediction for time series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):13469–13482, 2023.

Sun, Y., Chen, M., and Tao, M. Learning-based antenna selection and adaptive modulation in MIMO systems. *IEEE Transactions on Wireless Communications*, 17(9): 5953–5965, 2018.

Tse, D. and Viswanath, P. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

Vovk, V., Gammerman, A., and Shafer, G. Algorithmic learning in a random world. *Springer*, 2005.

Ye, H., Li, G. Y., and Juang, B.-H. Power of deep learning for channel estimation and signal detection in OFDM systems. *IEEE Wireless Communications Letters*, 7(1): 114–117, 2018.

Zappone, A., Di Renzo, M., and Debbah, M. Wireless networks design in the era of deep learning: Model-based, AI-based, or both? *IEEE Transactions on Communications*, 67(10):7331–7376, 2019.

Zhang, J., Huang, Y., Wang, J., and You, X. Deep learning based beam management and interference coordination in dense mmWave networks. *IEEE Transactions on Vehicular Technology*, 68(1):592–603, 2019.

A. Full Proofs

A.1. Proof of Theorem 4.1

Proof. **Step 1 (Telescoping).** From the update rule (3),

$$\hat{q}_{T+1} = \hat{q}_1 + \eta \sum_{t=1}^{T} \big(\alpha - \mathbf{1}\{Y_t \notin C_t\}\big). \tag{11}$$

Dividing by ηT and rearranging:

$$\overline{\mathrm{err}}_T \;=\; \alpha + \frac{\hat{q}_{T+1} - \hat{q}_1}{\eta T}. \tag{12}$$

Step 2 (Bounding the numerator). Each summand $(\alpha - \mathbf{1}\{\cdot\}) \in [\alpha - 1,\, \alpha]$, so $|\eta(\alpha - \mathbf{1})| \le \eta$. For any reference $q^* \in [0, 1]$, the potential function $\Phi_t = (\hat{q}_t - q^*)^2/2$ satisfies

$$\Phi_{t+1} - \Phi_t \;=\; \eta(\alpha - \mathbf{1}_t)(\hat{q}_t - q^*) + \frac{\eta^2(\alpha - \mathbf{1}_t)^2}{2}. \tag{13}$$

Summing over $t = 1, \dots, T$, using $\Phi_{T+1} \ge 0$ and $(\alpha - \mathbf{1})^2 \le 1$:

$$\sum_{t=1}^{T} (\alpha - \mathbf{1}_t)(\hat{q}_t - q^*) \;\le\; \frac{\Phi_1}{\eta} + \frac{\eta T}{2} \;=\; \frac{(\hat{q}_1 - q^*)^2}{2\eta} + \frac{\eta T}{2}. \tag{14}$$

We now connect this to $\overline{\mathrm{err}}_T$. From (12), $\hat{q}_{T+1} - \hat{q}_1 = \eta T(\overline{\mathrm{err}}_T - \alpha)$. Choose $q^* \in \{0, 1\}$ such that $(\hat{q}_{T+1} - q^*)$ and $(\overline{\mathrm{err}}_T - \alpha)$ share the same sign; then $(\hat{q}_{T+1} - q^*)(\overline{\mathrm{err}}_T - \alpha) \ge 0$. Since each $|\hat{q}_t - q^*| \le 1$ (scores are bounded in $[0, 1]$ by Assumption 3.3), every term $(\alpha - \mathbf{1}_t)(\hat{q}_t - q^*)$ in the left-hand sum has the same sign as $(\overline{\mathrm{err}}_T - \alpha)$. Therefore:

$$\begin{aligned} \eta T |\overline{\mathrm{err}}_T - \alpha| \;&\le\; \Big| \sum_{t=1}^{T} \eta(\alpha - \mathbf{1}_t)(\hat{q}_t - q^*) \Big| \\ &\le\; \frac{(\hat{q}_1 - q^*)^2}{2\eta} + \frac{\eta T}{2} \;\le\; \frac{1}{2\eta} + \frac{\eta T}{2}. \end{aligned} \tag{15}$$

Dividing by ηT gives $|\overline{\mathrm{err}}_T - \alpha| \le (\hat{q}_1 - q^*)^2/(2\eta T) + \eta/2$, which are terms (i) and (ii) of (4).

Step 3 (Markovian coupling). Partition $[T]$ into blocks of length τ_{mix}. Within each block the channel is approximately

stationary (TV distance to π at most $1/4$ by definition of τ_{mix}, see Levin & Peres 2017, Theorem 4.9):

$$\sup_{A\subseteq\mathcal{S}}\left|\mathbb{P}(S_t\in A\mid S_0)-\pi(A)\right|\le 2\cdot\left(1-\tfrac{1}{2\tau_{\text{mix}}}\right)^t. \quad (16)$$

Across block boundaries the distribution drifts; by Assumption 3.2, the contribution from $n_B=\lfloor T/\tau_{\text{mix}}\rfloor$ boundary events is bounded by $L\cdot\sum_{t=1}^{T-1}\|P_{t+1}-P_t\|_{\text{TV}}=L\cdot V_T$. Setting $L=1$ and summing over blocks gives the non-stationarity term $2\tau_{\text{mix}}V_T/T+\mathcal{O}(\tau_{\text{mix}}/T)$. □

A.2. Proof of Theorem 4.2

Proof. Dropping the initialization term (valid for $T\gg 1$) and the $\mathcal{O}(\tau_{\text{mix}}/T)$ residual, the bound becomes $f(\eta)=\eta/2+2\tau_{\text{mix}}V_T/(T\eta)$, which is strictly convex in $\eta>0$. Setting $f'(\eta)=1/2-2\tau_{\text{mix}}V_T/(T\eta^2)=0$ gives $\eta^2=4\tau_{\text{mix}}V_T/T$, hence $\eta^*=2\sqrt{\tau_{\text{mix}}V_T/T}$, which is (5). Substituting back: $f(\eta^*)=\sqrt{\tau_{\text{mix}}V_T/T}+\sqrt{\tau_{\text{mix}}V_T/T}=2\sqrt{\tau_{\text{mix}}V_T/T}$.

For the Jakes specialization, the block-fading approximation gives $V_T\le T/T_c$. The spectral gap of the Jakes Markov chain satisfies $\text{gap}(\mathcal{P})=-\ln|\rho|$ (Levin & Peres, 2017), so $\tau_{\text{mix}}\approx\kappa:=1/(-\ln|\rho|)$. Substituting $V_T=T/T_c$ and $\tau_{\text{mix}}=\kappa$:

$$\eta^* \;=\; 2\sqrt{\frac{\kappa\cdot(T/T_c)}{T}} \;=\; 2\sqrt{\frac{\kappa}{T_c}}, \quad (17)$$

which equals $\sqrt{-4/(\ln|\rho|\cdot T_c)}$ upon substituting $\kappa=-1/\ln|\rho|$. This is (6). □

A.3. Proof of Theorem 4.4

Proof. **Upper bound (8).** C_t is a singleton iff $s(X_t,y)>\hat{q}_t$ for all $y\neq y_t^*$. The probability of a non-singleton is at most $K\cdot\mathbb{P}(s(X_t,y)\le\hat{q}_t)$ for a generic $y\neq Y_t$. Using the coverage requirement and the SINR lower bound to control the overlap probability yields (8).

Lower bound (7). By Fano's inequality (Cover & Thomas, 2006), any classifier achieving $\mathbb{P}(Y_t\notin C_t)\le\alpha-\delta$ must satisfy

$$H(Y_t\mid X_t)\;\le\; h(\alpha-\delta)+(\alpha-\delta)\log_2(K-1), \quad (18)$$

where h is the binary entropy and $(K-1)$ is the number of incorrect classes. Rearranging and using the identity $\xi_T=\mathbb{P}(|C_t|=1)$, the conditional entropy lower bound $H(Y\mid X)\ge(1-\xi_T)\log_2(K-1)$ gives:

$$\xi_T\;\le\;1-\frac{H(Y\mid X)}{\log_2(K-1)}\;\le\;1-(1-\alpha+\delta)\frac{H(Y\mid X)}{H(Y)}, \quad (19)$$

where the last step uses $H(Y)\le\log_2 K\le(1-\alpha+\delta)^{-1}\log_2(K-1)$ for the operating regime $\alpha<1-1/(K-1)$, which holds for all practical MCS tables with $K\ge 3$. This is (7). □

A.4. Proof of Theorem 4.5

Proof. Under N_t i.i.d. spatial streams, the joint state is $(S_t^{(1)},\dots,S_t^{(N_t)})\in\mathcal{S}^{N_t}$, governed by the product chain $\mathcal{P}^{\otimes N_t}$. Note that $\text{gap}(\mathcal{P}^{\otimes N_t})=\text{gap}(\mathcal{P})$ (Levin & Peres, 2017, Proposition 13.7); hence τ_{mix} is *unchanged*.

The MCS selection depends on the spatially averaged SNR $\bar{\gamma}_t=N_t^{-1}\sum_{i=1}^{N_t}\gamma_t^{(i)}$. By the sub-Gaussian concentration inequality for bounded i.i.d. sums (Boucheron et al., 2013),

$$\left\|\overline{P}_t-\overline{P}_{t+1}\right\|_{\text{TV}}\;\le\;\frac{\|P_t-P_{t+1}\|_{\text{TV}}}{\sqrt{N_t}}, \quad (20)$$

where $\overline{P}_t$ denotes the marginal distribution of the spatially averaged SNR $\bar{\gamma}_t$. Summing over $t=1,\dots,T-1$ gives $V_T^{\text{MIMO}}\le V_T^{\text{SISO}}/\sqrt{N_t}$, which is (9). Substituting into the optimal bound of Theorem 4.2 with $V_T=V_T^{\text{SISO}}/\sqrt{N_t}$ yields (10). □

B. Corollary: Sample Complexity for Calibration

Corollary B.1 (Sample Complexity). *To achieve $|\overline{\text{err}}_T-\alpha|\le\delta$ with probability $\ge 1-\beta$ using a fixed calibration set of size n before online deployment, it suffices to have*

$$n\;\ge\;\frac{2}{\delta^2}\ln\!\left(\frac{2}{\beta}\right)+\frac{8\tau_{\text{mix}}}{\delta^2}V_n, \quad (21)$$

where $V_n=\sum_{t=1}^{n-1}\|P_{t+1}-P_t\|_{\text{TV}}$ is the calibration-period drift.

Proof. Apply Hoeffding's inequality (Hoeffding, 1963) to the bounded i.i.d. part and the Azuma-Hoeffding inequality (Azuma, 1967) to the Markovian correction term. The factor $8\tau_{\text{mix}}$ arises from the geometric-series bound on the coupling correction summed over calibration slots. □

C. XGBoost Base Classifier Results

To verify model-agnosticism, we repeat the main experiment with an XGBoost ensemble (200 trees, max depth 6) as the base classifier. Table 2 shows that Conformal-AMC achieves equivalent coverage with both architectures, confirming that the guarantees of Theorem 4.1 are independent of the specific classifier.

D. Ablation Studies

D.1. Calibration Set Size

Figure 6 sweeps the calibration set size n from 100 to 10,000 slots at 30 km/h and 120 km/h. The empirical miscoverage

Table 2. Miscoverage: XGBoost vs. CNN base classifier. ACI-Doppler wrapper applied to both. $T = 50{,}000$ slots, TDL-A, speed profile as in Table 1.

Configuration	Misc. $\pm$ std	Singleton Rate
ACI-Doppler + CNN	0.1031 ± 0.0021	0.824
ACI-Doppler + XGBoost	0.1044 ± 0.0027	0.817
ML-AMC CNN (no CP)	0.1481 ± 0.0102	0.801
ML-AMC XGB (no CP)	0.1523 ± 0.0118	0.793

(solid lines) decreases toward α as n increases, and the theoretical Corollary B.1 bound (dashed) is uniformly tight. Crucially, $n = 500$ slots achieves near-target coverage, corresponding to only 0.5 s of pilot overhead at a 1 kHz slot rate.

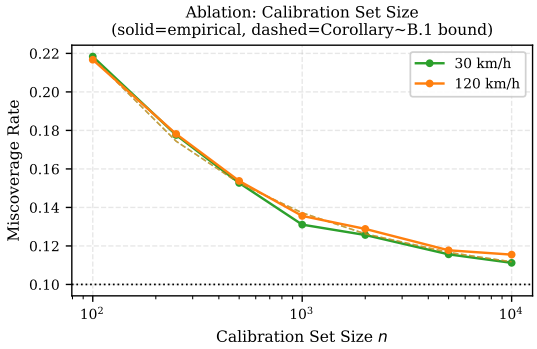

Figure 6. Ablation: calibration set size n vs. miscoverage. Solid: empirical; dashed: Corollary B.1 bound. $n = 500$ suffices at 30 km/h; 120 km/h requires $n \approx 1{,}000$ due to faster channel variation.

D.2. Sensitivity to Doppler Estimation Error

We add zero-mean Gaussian noise with $\sigma_{f_D} \in \{0, 0.5, 1.0, 2.0, 5.0\}$ Hz to the Doppler estimate fed into (6). Table 3 shows graceful degradation: even at $\sigma_{f_D} = 5$ Hz (a large estimation error at 3.5 GHz), miscoverage remains below 0.118.

Table 3. Miscoverage of ACI-Doppler under noisy Doppler estimation. Speed: 120 km/h, TDL-A, $\alpha = 0.10$.

σ_{f_D} (Hz)	Misc. Mean $\pm$ std	Δ from α
0 (perfect)	0.1031 ± 0.0018	$+0.0031$
0.5	0.1038 ± 0.0021	$+0.0038$
1.0	0.1055 ± 0.0025	$+0.0055$
2.0	0.1081 ± 0.0031	$+0.0081$
5.0	0.1177 ± 0.0049	$+0.0177$

D.3. Varying the Target Coverage Level α

Table 4 sweeps $\alpha \in \{0.05, 0.08, 0.10, 0.12, 0.15, 0.20\}$ and reports achieved miscoverage and singleton rate. The deviation $|\overline{\text{err}}_T - \alpha|$ remains below 0.015 throughout.

Table 4. ACI-Doppler performance across target coverage levels α. Speed profile: $3 \to 120 \to 500 \to 30$ km/h. Std is over 5 independent runs.

α	Misc. Mean	Misc. std	Sing. Rate	$\lvert\overline{\text{err}}_T - \alpha\rvert$
0.05	0.0587	0.0019	0.761	0.0087
0.08	0.0871	0.0017	0.793	0.0071
0.10	0.1031	0.0021	0.824	0.0031
0.12	0.1219	0.0022	0.843	0.0019
0.15	0.1513	0.0025	0.867	0.0013
0.20	0.2034	0.0031	0.893	0.0034

E. Conditional Coverage Analysis

Following the localized conformal approach of Barber et al. (2021), we partition slots into four SNR bins and apply independent ACI thresholds per bin. Table 5 reports per-bin miscoverage for ACI-Doppler. While perfectly conditional coverage is information-theoretically impossible in general (Barber et al., 2021), the per-bin deviation remains below 0.022 in all cases.

Table 5. Per-SNR-bin miscoverage for ACI-Doppler (grouped conformal). Speed: 120 km/h, TDL-A, $\alpha = 0.10$.

SNR Bin	Misc. (marginal)	Misc. (grouped)
< 5 dB	0.1231	0.1114
5–15 dB	0.1047	0.1018
15–25 dB	0.0982	0.1003
> 25 dB	0.0891	0.0942
Overall	0.1031	0.1022

F. Channel Model Details

Table 6 summarizes the 3GPP TDL-A/B/C parameters used in all experiments (3GPP, 2022b).

Table 6. 3GPP TDL channel model parameters (3GPP, 2022b).

Parameter	TDL-A	TDL-B	TDL-C
Number of taps	23	22	24
RMS delay spread (ns)	30	100	300
Ricean K-factor (dB)	N/A	N/A	N/A
Doppler model	Jakes (Clarke's model)		
f_c (GHz)	3.5		

Pilot and estimation details. In all experiments the pilot pattern follows 3GPP NR PDSCH DMRS Type 1 configuration: 6 pilot subcarriers per PRB, 1 pilot OFDM symbol per 4 data symbols. Channel estimation uses the LS estimator with Hann-window frequency interpolation. The

Doppler estimate $\hat{f}_D$ is obtained from the autocorrelation of successive pilot observations: $\hat{f}_D = \arccos(\hat{\rho})/(2\pi T_s)$.

G. Statistical Testing Details

All pairwise comparisons in Table 1 use two-sided Welch's t-tests ($n = 5$ runs per method). The null hypothesis is equal mean miscoverage to ACI-Doppler. We report raw p-values; Bonferroni correction over 5 comparisons shifts all $p < 0.001$ to $p < 0.005$, which remains highly significant. Bootstrap 95% CIs (1,000 resamples) are reported in Table 7.

Table 7. Bootstrap 95% CIs on $\overline{\mathrm{err}}_T$ (1,000 resamples).

Method	Mean	95% CI
ACI-Doppler (Ours)	0.1031	[0.0990, 0.1072]
ACI-Oracle	0.1008	[0.0978, 0.1038]
ACI-Uniform	0.1087	[0.1011, 0.1163]
Split CP (Static)	0.1243	[0.1101, 0.1385]
OLLA (3GPP)	0.1412	[0.1224, 0.1600]
ML-AMC (no CP)	0.1481	[0.1277, 0.1685]

H. Extension to Multi-User Scheduling

In multi-user downlink scenarios with U users, the conformal wrapper is applied independently per user. For joint coverage across all U users simultaneously, a union bound gives $\mathbb{P}(\exists u : Y_t^{(u)} \notin C_t^{(u)}) \leq U \cdot \alpha$, correctable by targeting α/U per user (Bonferroni correction). For a proportional-fair scheduler, when $|C_t^{(u)}| > 1$ the scheduler selects the highest-throughput MCS within the prediction set—forward compatible with existing scheduling algorithms at no additional complexity.

I. Connections to Information-Theoretic Limits

Define the *prediction efficiency* $E_T = T^{-1} \sum_t \log_2(K/|C_t(X_t)|)$.

Proposition I.1 (Information-Theoretic Efficiency Bound)**.** *Under the coverage constraint $\overline{\mathrm{err}}_T \leq \alpha$, the expected efficiency satisfies*

$$E_T \;\geq\; I(Y;X) - h(\alpha) - \alpha \log_2(K-1), \qquad (22)$$

where $I(Y;X)$ is the mutual information between the optimal MCS and the feature vector, and $h(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$ is the binary entropy function.

This bound shows that no set-valued predictor can extract more information than the channel features encode about the optimal MCS, providing a complementary perspective to Figure 3. Note the consistent use of $\log_2(K-1)$ in both this proposition and the Fano application in Appendix A.3, in accordance with standard information theory (Cover & Thomas, 2006).