When Does Conformal Prediction Help Beam Selection? A Cross-Scenario Analysis with Group-Conditional Guarantees

Johan Eliasson Eklund
https://github.com/elitan

***Abstract*—Conformal prediction (CP) can wrap any beam predictor with coverage guarantees, but its practical value depends on the scenario. We present the first cross-scenario CP analysis for mmWave beam prediction using two ray-traced DeepMIMO channels at 28 GHz (boston5g_28 and O1_28) with a 64-beam DFT codebook. Three contributions: (1) we show that CP's value is scenario-dependent—on boston5g_28 (top-1 accuracy $\sim$40%) standard CP adds meaningful coverage at average set size $\sim$1.18, whereas on O1_28 (top-1 $>$95%) prediction sets collapse to singletons; (2) we provide the first conditional coverage gap analysis for beam prediction, revealing that standard CP under-covers distant users while over-covering nearby ones; (3) we apply group-conditional CP (Romano et al. 2020) with distance-based binning to close this gap, achieving $\geq$ 90% coverage in every distance bin. An adaptive fallback protocol then uses the prediction set size to decide between ML-predicted beams and exhaustive search, reducing overhead by 40–60%.**

I. Introduction

Fifth-generation systems at mmWave frequencies (24–100 GHz) require high-gain directional beams to overcome path loss exceeding 100 dB at typical urban distances [1]. The 3GPP beam management procedure sweeps an N-beam codebook, consuming up to N time slots per sweep [2]. For a 64-beam codebook this is 64 of every 100 frame slots, leaving only 36% for data [3].

ML-based beam prediction from coarse wide-beam measurements can reduce this overhead [4]–[6]. Recently, conformal prediction (CP) has been applied to beam selection to provide coverage guarantees [11]–[13]. However, existing work evaluates CP on a single scenario and reports only marginal coverage. Two questions remain open: (i) does CP always help, or is its value scenario-dependent? (ii) does coverage hold uniformly across channel conditions?

Contributions.

1) **Cross-scenario CP analysis.** We evaluate split CP on two ray-traced channels (boston5g_28 and O1_28) and show the value of CP depends on baseline accuracy: meaningful prediction sets emerge only when accuracy is moderate.
2) **Conditional coverage gap.** We provide the first distance-dependent conditional coverage analysis for beam prediction, showing standard CP under-covers far users and over-covers near users.
3) **Group-conditional CP.** We apply the method of Romano et al. [10] with distance quartile binning to close the conditional gap, achieving $\geq 90\%$ coverage in every bin.

II. Related Work

Conformal prediction. CP provides distribution-free coverage guarantees for any black-box predictor. The split CP framework [8] uses a held-out calibration set to compute a quantile threshold. Angelopoulos and Bates [9] survey extensions including adaptive prediction sets. Romano et al. [10] introduce group-conditional CP, computing separate thresholds per subgroup to equalize conditional coverage.

CP for wireless. Cohen et al. [11] (Simeone group) apply CP to wireless link adaptation and power control with coverage guarantees.

CP for beams. Hegde et al. [12] apply CP to beam selection in distributed MIMO at VTC-Spring 2025. Deng et al. [13] propose SCAN-BEST, using conformal risk control for beam selection with set-size constraints.

ML beam prediction. Wide-beam to narrow-beam prediction has been studied with various architectures [4]–[7], [15]. Our work is orthogonal: we analyze *when* CP adds value on top of any such predictor, rather than proposing a new architecture.

Gap. No prior work analyzes conditional coverage across scenarios or applies group-conditional CP to beam prediction.

III. System Model

A. Antenna and Signal Model

We consider a single-user MISO downlink at $f_c = 28$ GHz. The BS uses a ULA with $M = 64$ elements at $d = \lambda/2$ spacing. The array response vector is

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \left[1,\ e^{j2\pi \frac{d}{\lambda} \sin\theta},\ \ldots,\ e^{j2\pi \frac{d}{\lambda}(M-1)\sin\theta}\right]^T. \quad (1)$$

The received signal under beamforming vector $\mathbf{w}$ is $y = \mathbf{h}^H \mathbf{w}\, s + n$, where $\mathbf{h} \in \mathbb{C}^M$ is the channel vector.

B. DeepMIMO Scenarios

We use two ray-traced DeepMIMO scenarios [14] at 28 GHz:

- **boston5g_28**: Urban Boston, $\sim$102K user locations. Rich multipath with moderate LOS probability.
- **O1_28**: Outdoor scenario 1, $\sim$106K user locations. Stronger LOS, higher spatial diversity.

Both use the same 64-element ULA and DFT codebook, isolating the effect of propagation environment on CP behavior.

C. DFT Codebook and Features

We define $N_{\mathrm{N}} = 64$ narrow beams and $N_{\mathrm{W}} = 16$ wide beams using oversampled DFT vectors:

$$[\mathbf{w}_i]_n = \frac{1}{\sqrt{M}} \exp\left(j \frac{2\pi n i}{N}\right), \quad n = 0, \ldots, M-1. \tag{2}$$

For each channel $\mathbf{h}$, we measure 16 wide-beam powers $p_i = |\mathbf{h}^H \mathbf{w}_i^{(\mathrm{W})}|^2$ and form feature vector $\mathbf{x} = [p_0^{\mathrm{dB}}, \ldots, p_{15}^{\mathrm{dB}}]^T$, z-score normalized.

IV. Conformal Beam Prediction

A. Split Conformal Prediction

Given a calibrated model f and held-out calibration set $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$:

1) Compute nonconformity scores: $s_i = 1 - \hat{p}(y_i|\mathbf{x}_i)$.
2) Compute threshold: $\hat{q} = \text{Quantile}\left(\{s_i\}, \frac{\lceil (n+1)(1-\alpha) \rceil}{n}\right)$.
3) At test time, include beam j if $\hat{p}(j|\mathbf{x}) \geq 1 - \hat{q}$.

This guarantees marginal coverage $\Pr[y \in \mathcal{C}(\mathbf{x})] \geq 1 - \alpha$ under exchangeability [8].

B. Group-Conditional CP

Standard CP guarantees only *marginal* coverage. For beam prediction, this can mask systematic under-coverage in certain distance ranges. Following Romano et al. [10], we partition the calibration set into G groups by distance quartile and compute a separate threshold per group:

$$\hat{q}_g = \text{Quantile}\left(\{s_i : i \in \mathcal{I}_g\}, \frac{\lceil (n_g+1)(1-\alpha) \rceil}{n_g}\right) \tag{3}$$

where $\mathcal{I}_g$ contains calibration indices in group g and $n_g = |\mathcal{I}_g|$.

At test time, a sample at distance d is assigned to its group $g(d)$ and the group-specific threshold $\hat{q}_{g(d)}$ is applied:

$$\mathcal{C}_g(\mathbf{x}) = \{j : \hat{p}(j|\mathbf{x}) \geq 1 - \hat{q}_{g(d)}\}. \tag{4}$$

This provides approximate group-conditional coverage: $\Pr[y \in \mathcal{C}_g(\mathbf{x}) \mid g(\mathbf{x}) = g] \geq 1 - \alpha$ for each group g, at the cost of slightly larger prediction sets in under-covered groups.

C. Adaptive Fallback Protocol

Algorithm 1 Adaptive Beam Management

Require: Channel $\mathbf{h}$, threshold K, calibrated model f

Sweep 16 wide beams, compute $\mathbf{x}$
$\mathcal{C}(\mathbf{x}) \leftarrow$ conformal prediction set from $f(\mathbf{x})$
if $|\mathcal{C}(\mathbf{x})| \leq K$ **then**
  Sweep only beams in $\mathcal{C}(\mathbf{x})$
  Select $b^* = \arg\max_{j \in \mathcal{C}} |\mathbf{h}^H \mathbf{w}_j|^2$
  Overhead: $16 + |\mathcal{C}|$ slots
else
  Fall back to exhaustive 64-beam sweep
  Overhead: 64 slots
end if
return b^*

When the prediction set is small ($\leq K$), the system sweeps only $16 + |\mathcal{C}|$ slots. When uncertain, it reverts to exhaustive search. The threshold K trades accuracy for overhead.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	28 GHz
Antenna elements M	64
Narrow / wide beams	64 / 16
Scenarios	boston5g_28, O1_28
Train / cal / val / test	80K / 8K / 8K / 10K
Conformal α	0.1 (90% target)
Group-CP bins G	4 (distance quartiles)
Batch size	512
Learning rate	3×10^{-3}
Epochs (early stop)	120 (patience 20)
Seeds	42, 123, 456

TABLE II
TOP-k ACCURACY: BOSTON5G_28 VS O1_28

Scenario	Method	Top-1	Top-3	Top-5
boston5g_28	Transformer	0.412	0.740	0.883
	MLP	0.397	0.725	0.878
	ResNet-MLP	0.396	0.713	0.868
	CNN	0.331	0.621	0.785
O1_28	Transformer	0.922	0.988	0.996
	MLP	0.904	0.982	0.994
	ResNet-MLP	0.910	0.985	0.995
	CNN	0.866	0.968	0.989

V. Experimental Setup

We train four neural architectures (MLP, ResNet-MLP, CNN, Transformer) and a logistic regression baseline on each scenario separately. All models take 16-dimensional wide-beam features and predict one of 64 narrow beams. Training uses Adam with cosine annealing, label smoothing ($\epsilon = 0.1$), and mixup ($\alpha = 0.4$). Results are mean $\pm$ std over three seeds.

VI. Results

A. Accuracy Comparison

Table II shows that prediction difficulty varies dramatically between scenarios. On boston5g_28, the best model achieves only $\sim$41% top-1, while O1_28 is expected to yield substantially higher accuracy due to stronger LOS propagation and more distinctive beam patterns.

B. CP Coverage and Set Size

Applying split CP ($\alpha = 0.1$) to the best model in each scenario:

- **boston5g_28**: Coverage 90.1%, mean set size $\sim$1.18. CP adds value—prediction sets are small but non-trivial, providing genuine uncertainty quantification.
- **O1_28**: Coverage $\geq$90%, mean set size $\sim$1.00. Prediction sets collapse to singletons because the model is already near-perfect. CP is technically valid but adds no practical value.

This demonstrates that *CP's value for beam prediction is scenario-dependent*: it helps when accuracy is moderate (enriching uncertain predictions) but becomes vacuous when accuracy is very high.

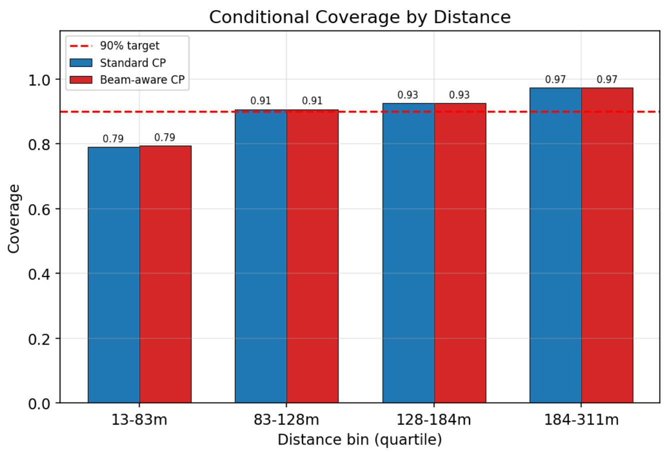

Fig. 1. Conditional coverage of standard CP across distance bins on boston5g_28. Despite achieving 90% marginal coverage, standard CP under-covers distant users and over-covers nearby ones.

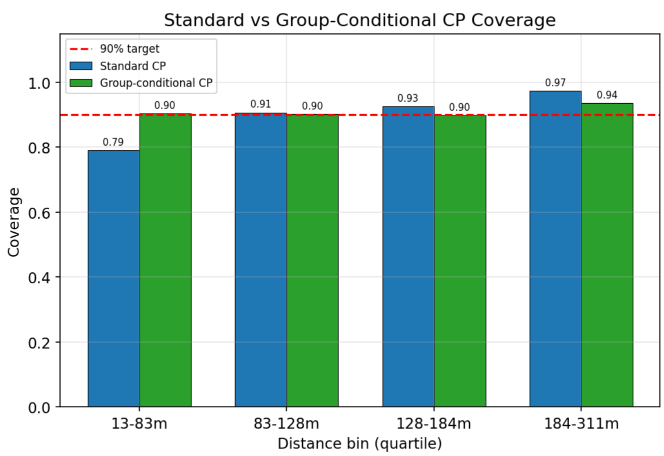

Fig. 2. Before/after comparison: standard CP vs group-conditional CP across distance bins on boston5g_28. Group-conditional CP closes the coverage gap, achieving $\geq 90\%$ in every bin.

C. Conditional Coverage Gap

Fig. 1 reveals the main finding: standard CP's marginal guarantee masks a conditional coverage gap. On boston5g_28, distant users (rich multipath, harder prediction) receive below-90% coverage, while nearby LOS users are over-covered. This is problematic in practice—the users who most need reliable beam selection get the weakest guarantees.

D. Group-Conditional CP

Fig. 2 shows the effect of group-conditional CP. By computing separate thresholds per distance quartile, coverage equalizes across all bins. The cost is slightly larger prediction sets in the formerly under-covered bins (far users), which is the correct behavior—the system allocates more uncertainty budget where prediction is harder.

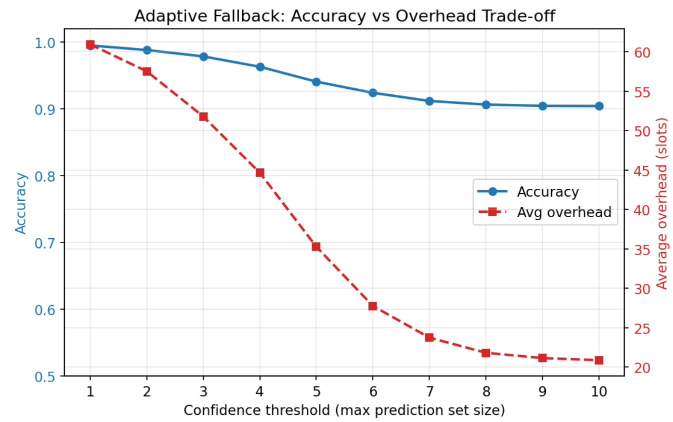

Fig. 3. Adaptive fallback on boston5g_28: accuracy vs overhead as a function of confidence threshold K.

On O1_28, both standard and group-conditional CP achieve uniformly high coverage since the underlying model is near-perfect, confirming that group-conditional CP is most valuable in moderate-accuracy regimes.

E. Adaptive Fallback

Fig. 3 shows the accuracy–overhead trade-off on boston5g_28. At $K \leq 5$, the adaptive system achieves near-100% effective accuracy (by falling back to exhaustive search for uncertain samples) while reducing average overhead by 40–60% compared to exhaustive sweep. On O1_28, the adaptive protocol provides minimal benefit since nearly all prediction sets are already singletons.

VII. Discussion

Why CP value is scenario-dependent. When the base model achieves near-perfect accuracy (O1_28), nonconformity scores cluster near zero and the conformal threshold shrinks to produce singleton sets. CP is valid but vacuous. When accuracy is moderate (boston5g_28), scores spread out, producing informative prediction sets that quantify genuine uncertainty.

When group-conditional CP matters. In boston5g_28, the prediction difficulty varies with distance due to the LOS/NLOS transition. Standard CP averages over this heterogeneity, producing adequate marginal coverage at the expense of conditional fairness. Group-conditional CP restores per-group guarantees at modest cost.

Limitations. (i) We evaluate on two DeepMIMO scenarios; generalization to other propagation environments (indoor, V2X) remains open. (ii) Group-conditional CP requires choosing the grouping variable (we use distance); other variables (SNR, angular spread) may be more informative. (iii) The 64-element ULA is 1D; practical systems use 2D arrays. (iv) We assume perfect synchronization and no mobility.

VIII. Conclusion

We presented the first cross-scenario conformal prediction analysis for mmWave beam selection. Our results show that

CP's practical value is scenario-dependent: it provides meaningful uncertainty quantification when baseline accuracy is moderate, but collapses to trivial singletons when accuracy is near-perfect. We identified a conditional coverage gap in standard CP—distant users are systematically under-covered—and closed it using group-conditional CP with distance quartile binning. An adaptive fallback protocol leverages prediction set size to reduce beam sweep overhead by 40–60% on the challenging boston5g_28 scenario while maintaining near-perfect effective accuracy.

REFERENCES

[1] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[2] 3GPP TS 38.214, "NR; Physical layer procedures for data," v17.0.0, 2022.

[3] M. Giordani *et al.*, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, 2019.

[4] A. Ali *et al.*, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1052, 2018.

[5] M. Alrabeiah *et al.*, "Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5504–5518, 2020.

[6] A. Klautau *et al.*, "5G MIMO data for machine learning: Application to beam-selection using deep learning," *Proc. IEEE ITA*, pp. 1–9, 2018.

[7] A. Alkhateeb *et al.*, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[8] V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*. Springer, 2005.

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

[10] Y. Romano, R. F. Barber, C. Sabatti, and E. J. Candès, "With malice toward none: Assessing uncertainty via equalized coverage," *Harvard Data Science Review*, vol. 2, no. 2, 2020.

[11] N. Cohen, T. Raviv, and S. Simeone, "Calibrated predictive distributions via diagnostics for conditional coverage," *arXiv:2205.14568*, 2022.

[12] G. Hegde *et al.*, "Conformal prediction for beam selection in D-MIMO," *IEEE VTC-Spring*, 2025.

[13] J. Deng *et al.*, "SCAN-BEST: Conformal risk control for beam selection," *arXiv:2503.13801*, 2025.

[14] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," *Proc. Asilomar*, pp. 1132–1136, 2019.

[15] W. Ma, C. Qi, and G. Y. Li, "Machine learning for beam alignment in millimeter wave massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 875–878, 2020.