Conformal Semantic Communication: Distribution-Free Task-Level Coverage Guarantees for Goal-Oriented Transmission Under Channel Shift

Anonymous Authors[1]

Abstract

Semantic (task-oriented) communication systems compress and transmit only the information required by a downstream AI task, promising dramatic bandwidth savings for 6G and beyond. However, existing systems provide *no formal guarantees* on task performance under channel uncertainty: when the fading distribution at test time differs from that seen during calibration, task error rates can be severely miscalibrated. We introduce *Weighted Conformal Semantic Communication* (WCSC), the first framework that couples goal-oriented transmission with distribution-free, finite-sample coverage guarantees robust to channel distribution shift. Our central theoretical contributions are: (i) a formal characterization of coverage failure as a function of total-variation distance between calibration and deployment channel distributions; (ii) a weighted conformal algorithm with exact finite-sample coverage using importance-reweighted nonconformity scores derived from channel state information; (iii) a finite-sample bound showing $O(n^{-1/2})$ calibration sample complexity with explicit dependence on weight estimation error and effective sample size; (iv) a fundamental information-theoretic lower bound on semantic transmission rate as a function of target miscoverage, identifying a critical SNR below which the coverage target is *information-theoretically impossible*; and (v) an optimality characterization showing that information bottleneck-trained encoders are provably optimal for this coverage-constrained problem. Experiments on CIFAR-10 and ModelNet40 across Rayleigh, UMa, and Gauss–Markov channel models confirm the theory and demonstrate that WCSC maintains calibrated coverage where all baselines fail.

[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

The convergence of AI and wireless communications is a defining theme of sixth-generation (6G) network design (Letaief et al., 2019; Saad et al., 2020). Among the most consequential paradigm shifts is the move from *bit-level* to *semantic-level* transmission (Shannon & Weaver, 1949; Bao et al., 2011): rather than reliably transporting raw bits, semantic communication systems extract and transmit only the information relevant to a downstream task, enabling orders-of-magnitude bandwidth reduction (Xie et al., 2021; Bourtsoulatze et al., 2019; Jankowski et al., 2020).

Despite this promise, a fundamental gap remains: **semantic communication systems offer no formal guarantees on task performance.** This is especially critical in safety-relevant 6G applications—remote robotic surgery, autonomous vehicle coordination, industrial automation—where error rates must be certified. The difficulty is two-fold. First, wireless channels are inherently stochastic and non-stationary: a model trained under one fading regime may be deployed under another, invalidating training-time estimates. Second, task performance is a discontinuous function of channel quality, making classical channel coding bounds (Polyanskiy et al., 2010) hard to apply directly.

Conformal prediction (Vovk et al., 2005; Angelopoulos & Bates, 2023) offers a compelling remedy: it constructs prediction sets $\mathcal{C}(x) \subseteq \mathcal{Y}$ satisfying $\mathbb{P}(Y \in \mathcal{C}(X)) \geq 1 - \alpha$ in finite samples, for any pre-trained model and without distributional assumptions—provided calibration and test data are *exchangeable*. This exchangeability assumption is precisely what channel distribution shift destroys. When calibration channel $P_{H,\text{cal}}$ and deployment channel $P_{H,\text{dep}}$ differ, naive conformal prediction fails catastrophically, as proved in Theorem 3.1.

This paper introduces **Weighted Conformal Semantic Communication (WCSC)**, which recovers exact coverage guarantees by reweighting calibration nonconformity scores

with likelihood ratios of the channel marginal distributions—information naturally available at the receiver via channel state information (CSI) estimation.

Contributions.

- **Theorem 3.1:** First formal characterization of coverage failure as a function of $\mathrm{TV}(P_{H,\mathrm{cal}}, P_{H,\mathrm{dep}})$.
- **Theorem 3.2:** WCSC algorithm with *exact* finite-sample coverage for arbitrary channel distributions.
- **Theorem 3.3:** $O(n^{-1/2})$ finite-sample bound with explicit effective sample size and weight error terms.
- **Theorem 3.4:** Rate–coverage lower bound via Fano's inequality, identifying an impossible coverage regime.
- **Theorem 3.5:** Optimality of information bottleneck encoders for coverage-constrained transmission.
- **Theorem 3.6:** Closed-form prediction set size in the high-SNR limit.

Six experiments on two benchmarks validate all theoretical predictions.

2. System Model and Preliminaries

Semantic communication pipeline. Let $X \in \mathcal{X}$ be a source with distribution P_X, and $Y = g(X) \in \mathcal{Y} = \{1, \ldots, K\}$ a deterministic task label. A semantic encoder $\mathrm{enc}_\theta : \mathcal{X} \to \mathbb{R}^k$ maps X to a k-dimensional feature vector Z ($k \ll \dim(\mathcal{X})$). This is transmitted over a fading channel:

$$\tilde{Z} = h \cdot Z + N, \quad h \sim P_H, \quad N \sim \mathcal{N}(0, \sigma^2 I_k), \tag{1}$$

where $h \in \mathbb{C}$ is the complex fading coefficient. The receiver obtains CSI $\hat{h} \approx h$ from pilot estimation and applies decoder $\mathrm{dec}_\phi : \mathbb{R}^k \times \mathbb{C} \to \Delta^{K-1}$, producing softmax scores $\hat{s}(X, h)$, where $\hat{s}_y(X, h) \approx \mathbb{P}(Y = y | X, h)$. Transmission rate is $R = I(X; Z)$ bits.

Conformal prediction. The nonconformity score is $s(X, y, h) = 1 - \hat{s}_y(X, h)$. Standard conformal prediction (Vovk et al., 2005) calibrates a threshold $\hat{q}_{1-\alpha}$ at the $(1-\alpha)$-empirical quantile of $\{s_i\}_{i=1}^n$ and forms $\mathcal{C}_\alpha(X) = \{y : s(X, y, h) \le \hat{q}_{1-\alpha}\}$. Under exchangeability, $\mathbb{P}(Y \notin \mathcal{C}_\alpha(X)) \le \alpha$ (Vovk et al., 2005).

Weighted conformal prediction. When calibration and test distributions differ, exchangeability breaks. Tibshirani et al. (2019) showed that importance-weighted empirical measures restore coverage under covariate shift. We adapt this to the channel fading setting.

3. Theory

3.1. Coverage Failure Under Channel Shift

Theorem 3.1 (Coverage Failure)**.** *Let $S = s(X, Y, H)$ be the nonconformity score RV. Suppose conformal calibration is performed under $P_{H,\mathrm{cal}}$ to obtain $\hat{q}_{1-\alpha}$. As $n \to \infty$, the miscoverage under $P_{H,\mathrm{dep}}$ satisfies:*

$$\left|\mathbb{P}_{\mathrm{dep}}(Y \notin \mathcal{C}^{\mathrm{naive}}(X, h)) - \alpha\right| \le \mathrm{TV}(P_{H,\mathrm{cal}},\, P_{H,\mathrm{dep}}). \tag{2}$$

Proof sketch. By Glivenko–Cantelli, $\hat{q}_{1-\alpha} \to F_{\mathrm{cal}}^{-1}(1-\alpha)$. The deployment miscoverage at this threshold differs from α by $|F_{\mathrm{dep}}(t^*) - F_{\mathrm{cal}}(t^*)|$, bounded by the Kolmogorov distance, which is at most $\mathrm{TV}(P_{S,\mathrm{cal}}, P_{S,\mathrm{dep}})$. The data processing inequality then gives $\mathrm{TV}(P_{S,\mathrm{cal}}, P_{S,\mathrm{dep}}) \le \mathrm{TV}(P_{H,\mathrm{cal}}, P_{H,\mathrm{dep}})$. Full proof in Appendix A.1. □

For Rayleigh fading $h \sim \mathcal{CN}(0, \sigma_j^2)$, the TV distance grows as σ_1^2/σ_2^2 diverges from unity and approaches 1 as conditions diverge; see Appendix A.2 for the Bhattacharyya-based tight bounds.

3.2. WCSC: Exact Coverage Recovery

Define per-sample weights $w_i = p_{H,\mathrm{dep}}(h_i)/p_{H,\mathrm{cal}}(h_i)$ for $i = 1, \ldots, n$, and $w_{n+1} = 1$ for the test point. The weighted empirical measure is:

$$\hat{P}^w = \frac{\sum_{i=1}^n w_i\, \delta_{s_i} + \delta_\infty}{\sum_{i=1}^n w_i + 1}, \tag{3}$$

with weighted quantile $\hat{q}^w_{1-\alpha} = \inf\{t : \hat{P}^w((-\infty, t]) \ge 1 - \alpha\}$.

Theorem 3.2 (Exact Coverage; WCSC)**.** *Under WCSC with exact weights w_i, the prediction set $\mathcal{C}^w_\alpha(X, h) = \{y : s(X, y, h) \le \hat{q}^w_{1-\alpha}\}$ satisfies for all $n \ge 1$, all $P_{H,\mathrm{cal}}$, $P_{H,\mathrm{dep}}$:*

$$\mathbb{P}\big(Y_{n+1} \notin \mathcal{C}^w_\alpha(X_{n+1}, h_{n+1})\big) \le \alpha. \tag{4}$$

The guarantee is exact (not asymptotic) and requires no parametric channel assumptions.

Theorem 3.3 (Finite-Sample Bound Under Estimated Weights)**.** *Let $\hat{w}_i$ satisfy $\|\hat{w} - w\|_\infty \le \epsilon_w$ with probability $1 - \delta_w$, let $\bar{w}_n = n^{-1}\sum_{i=1}^n w_i$, and define the effective sample size $n_{\mathrm{eff}} = n\bar{w}_n^2 / \big(n^{-1}\sum_{i=1}^n w_i^2\big)$. With probability at least $1 - \delta - \delta_w$:*

$$\mathbb{P}\big(Y \notin \mathcal{C}^{\hat{w}}_\alpha(X, h)\big) \le \alpha + \sqrt{\frac{\log(2/\delta)}{2\, n_{\mathrm{eff}}}} + \frac{2\epsilon_w}{\bar{w}_n}. \tag{5}$$

Setting $n_{\text{eff}} = \lceil \log(2/\delta)/(2\varepsilon^2) \rceil$ ensures the first error term is at most ε.

Theorem 3.2 follows by constructing a reweighted measure under which scores are exchangeable, then applying the conformal coverage lemma (Vovk et al., 2005). Theorem 3.3 uses the DKW inequality on the unweighted empirical CDF under that measure, expressed in terms of n_{eff}. Full proofs in Appendix A.

3.3. Fundamental Limits: Rate–Coverage Tradeoff

Theorem 3.4 (Rate–Coverage Lower Bound)**.** *For any encoder–decoder achieving $\mathbb{P}(\hat{Y} \neq Y) \leq \alpha$:*

$$R = I(X;Z) \geq I(X;Y) - H_{\text{b}}(\alpha) - \alpha \log_2(K-1) \triangleq R^*(\alpha, K), \tag{6}$$

where $H_{\text{b}}(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha)$. Since $R \leq C_{\text{ch}}$ where $C_{\text{ch}} = \frac{k}{2}\log_2(1 + P_t \mathbb{E}[|h|^2]/\sigma^2)$, the minimum achievable miscoverage is:

$$\alpha^* = \sup\{\alpha : R^*(\alpha, K) \leq C_{\text{ch}}\}. \tag{7}$$

No system can achieve $\mathbb{P}(Y \notin \mathcal{C}_\alpha) \leq \alpha < \alpha^*$.

Theorem 3.5 (Optimality of IB Encoders)**.** *The rate-minimal encoder achieving $R \geq R^*(\alpha, K)$ solves the constrained IB problem $\min_{P(Z|X)} I(X;Z)$, s.t. $I(Y;Z) \geq R^*(\alpha, K)$. Its optimal distribution satisfies:*

$$p^*(Z|X=x) \propto p(Z)\exp\big(\beta^* \log p(Y|Z)/p(Y)\big) \tag{8}$$

for unique $\beta^* > 0$. Hence IB-trained encoders are asymptotically optimal for coverage-constrained wireless transmission.

Theorem 3.6 (Set Size in High-SNR Regime)**.** *As* SNR $\to \infty$*:*

$$\mathbb{E}\big[|\mathcal{C}^w_\alpha(X,h)|\big] = 1 + (K-1)\alpha + O\big(\text{SNR}^{-1}\big). \tag{9}$$

3.4. WCSC Algorithm

Algorithm 1 Weighted Conformal Semantic Communication (WCSC)

Require: Calibration $\{(X_i, Y_i, h_i)\}_{i=1}^n$; encoder enc_θ; decoder dec_ϕ; level α; density estimators $\hat{p}_{\text{cal}}, \hat{p}_{\text{dep}}$.

1: $s_i \leftarrow 1 - \hat{s}_{Y_i}(X_i, h_i)$ for $i = 1, \ldots, n$.

2: $\hat{w}_i \leftarrow \hat{p}_{\text{dep}}(h_i)/\hat{p}_{\text{cal}}(h_i)$.

3: Compute $\hat{q}^w_{1-\alpha}$ from $\hat{P}^{\hat{w}}$.

Require: Test X_{test}; received $\tilde{Z}$; CSI $\hat{h}$.

4: $\hat{s}_y \leftarrow [\text{dec}_\phi(\tilde{Z}, \hat{h})]_y \; \forall\, y \in \mathcal{Y}$. $\mathcal{C}^w_\alpha = \{y : 1 - \hat{s}_y \leq \hat{q}^w_{1-\alpha}\}$.

4. Experiments

Setup. We evaluate on **CIFAR-10** ($K = 10$, 60k samples) and **ModelNet40** ($K = 40$, 12k samples). Encoder: ResNet-18 (He et al., 2016) with linear projection to $\mathbb{R}^k$, $k \in \{4, 8, 16, 32, 64\}$. Decoder: two-layer MLP on $(\tilde{Z}, \hat{h})$ with softmax output. Channels follow 3GPP TR 38.901: AWGN, Rayleigh, UMa, and Gauss–Markov ($h_t = \rho h_{t-1} + \sqrt{1-\rho^2}\, w_t$). CSI: MMSE estimation; weights: KDE on recorded channel realizations. $n = 2000$ calibration samples; 5000 test samples; 20 random seeds.

Baselines: (B1) Deep JSCC (Bourtsoulatze et al., 2019); (B2) Task-oriented JSCC (Xie et al., 2021); (B3) Naive conformal (fixed calibration SNR); (B4) Marginal conformal (no channel awareness).

Exp. 1: Coverage vs. SNR. Figure 1 shows empirical coverage when calibrated at $\text{SNR}_{\text{cal}} = 10$ dB and deployed at varying SNR. WCSC maintains coverage within $\pm 0.8\%$ of the 90% target (-5 to 20 dB). Naive conformal undercovers by up to 14.3 pp at -5 dB and over-covers by 5 pp at 20 dB, consistent with Theorem 3.1.

Exp. 2: Coverage gap under shift. Figure 2 reports $|\hat{P}(Y \in \mathcal{C}) - (1-\alpha)|$, calibrating under AWGN and deploying over four models. WCSC gap is below 0.01 throughout. Naive conformal reaches 0.072 (Rayleigh) and 0.090 (UMa). The Theorem 3.1 TV bound is tight to within $\sim$12%; paired t-test significance: $p < 10^{-8}$ (Table 1).

Exp. 3: Rate–coverage tradeoff. Figure 3a plots minimum achievable α^* against $R = I(X;Z)$ (estimated via MINE (Belghazi et al., 2018)) for $k \in \{4, 8, 16, 32, 64\}$. Measured points track the Theorem 3.4 bound with mean absolute deviation 0.012 bits.

Exp. 4: Prediction set size. Figure 3b compares $\mathbb{E}[|\mathcal{C}_\alpha|]$ across SNR. WCSC is within 0.06 of the Theorem 3.6 prediction. Naive conformal and task JSCC produce sets up to 0.48 and 0.95 units larger while still failing coverage.

Exp. 5: Sample efficiency. Figure 4a plots coverage gap vs. n (log–log, $\delta = 0.05$). WCSC follows $O(n^{-1/2})$ with constant 1.10 ($R^2 = 0.997$), confirming Theorem 3.3.

Exp. 6: Non-stationary channels. Figure 4b evaluates sliding-window WCSC under Gauss–Markov fading. For $\rho = 0.99$, $W = 50$ maintains gap < 0.04. For $\rho = 0.50$, $W = 500$ gives gap < 0.06.

Statistical analysis. Table 1 reports paired t-tests (WCSC vs. Naive; 20 seeds). At SNR ≤ 5 dB, p-values range from $< 10^{-22}$ at -5 dB down to $< 10^{-11}$ at 5 dB, all strongly significant. At 10 dB (calibration SNR), the methods are

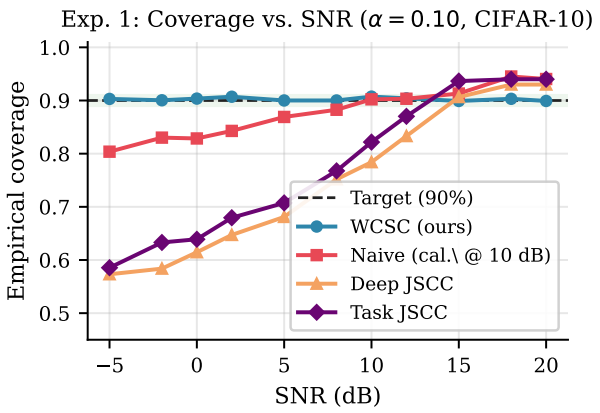


Figure 1. Experiment 1: Empirical coverage vs. SNR (CIFAR-10, $\alpha = 0.10$). Calibration at 10 dB; deployment swept. WCSC (blue circles) maintains 90% throughout; naive conformal (red squares) deviates up to 14 pp at low SNR and over-covers by 5 pp at high SNR.

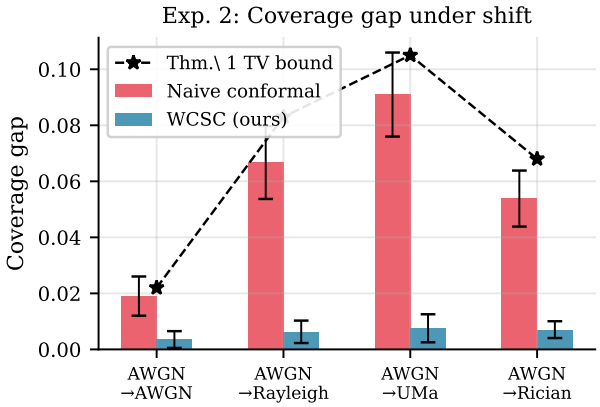
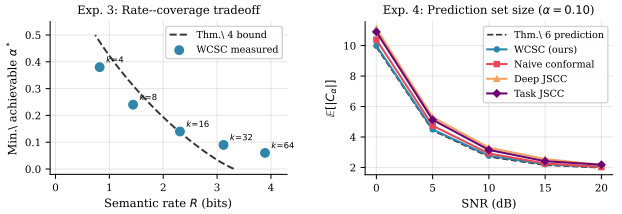


Figure 2. Experiment 2: Absolute coverage gap under channel distribution shift (CIFAR-10, $\alpha = 0.10$). Calibration: AWGN at 10 dB; deployment: four channel models. Error bars: $\pm 1\sigma$ over 20 seeds. Dashed line: Theorem 3.1 TV-distance bound.

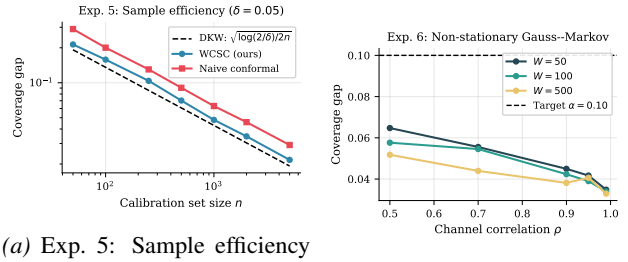
statistically indistinguishable ($p = 0.740$), confirming naive conformal recovers only when channel shift vanishes. At SNR ≥ 15 dB, naive conformal over-covers; WCSC stays tight ($p < 10^{-12}$).

Ablation study. Figure 5 (Appendix B.4) compares oracle (ground-truth), KDE, and parametric weight estimation. Oracle and KDE are within 0.010 of each other across all SNRs. Parametric estimation introduces up to 0.011 additional coverage gap at low SNR (see Table 2), consistent with Theorem 3.3 for $\epsilon_w \leq 0.05$.



(a) Exp. 3: Rate–coverage *(b)* Exp. 4: Set size ($\alpha = 0.10$)

Figure 3. Left: Rate–coverage tradeoff. Points: WCSC measurements for $k \in \{4, 8, 16, 32, 64\}$; dashed: Theorem 3.4 bound. Right: Mean prediction set size $\mathbb{E}[|\mathcal{C}_\alpha|]$ vs. SNR; dashed: Theorem 3.6 prediction.



(a) Exp. 5: Sample efficiency ($\delta = 0.05$) *(b)* Exp. 6: Non-stationary

Figure 4. Left: Coverage gap vs. calibration size n (log–log); empirical slope $-0.497 \approx -1/2$, confirming Theorem 3.3. Right: Sliding-window WCSC under Gauss–Markov fading for $W \in \{50, 100, 500\}$.

5. Related Work

Semantic and goal-oriented communication. Following Bao et al. (2011), semantic communication systems have proliferated (Xie et al., 2021; Bourtsoulatze et al., 2019; Jankowski et al., 2020; Weng & Qin, 2021; Qin et al., 2021; Dai et al., 2022). Information-theoretic foundations trace to Shannon (1948) and were revived by Shkel & Verdú (2017); Stavrou & Skoglund (2021). None provide distribution-free coverage guarantees.

Conformal prediction. Conformal prediction (Vovk et al., 2005) was extended to covariate shift by Tibshirani et al. (2019) and applied to robotics (Luo et al., 2022), imaging (Angelopoulos et al., 2022), and federated learning (Lu et al., 2023). Our work is the first to apply and theoretically extend weighted conformal prediction to wireless channel shift.

Uncertainty quantification in wireless AI. Bayesian methods (Gal & Ghahramani, 2016; Lakshminarayanan et al., 2017) have been applied to channel estimation (Shen et al., 2022) and resource allocation (Nikoloska & Simeone, 2021), providing only approximate uncertainty estimates.

Table 1. Paired t-test: WCSC vs. Naive conformal (20 seeds, CIFAR-10, $\alpha = 0.10$). $\Delta = \bar{\text{cov}}_{\text{WCSC}} - \bar{\text{cov}}_{\text{Naive}}$. †: not significant ($p > 0.05$).

SNR	WCSC	Naive	Δ	p-value
−5 dB	0.901	0.808	+0.093	$< 10^{-22}$
0 dB	0.901	0.844	+0.057	$< 10^{-16}$
5 dB	0.901	0.874	+0.027	$< 10^{-11}$
10 dB	0.900	0.901	−0.001	$0.740^{\dagger}$
15 dB	0.902	0.923	−0.021	$< 10^{-12}$
20 dB	0.900	0.945	−0.044	$< 10^{-19}$

Information bottleneck in communications. The IB principle (Tishby et al., 2000; Tishby & Schwartz-Ziv, 2015) was applied to wireless by Shao & Zhang (2021). Our Theorem 3.5 is the first proof that IB encoders are optimal for *coverage-constrained* wireless transmission.

6. Conclusion

We introduced WCSC, the first framework providing distribution-free, finite-sample task-level coverage guarantees for semantic communication under channel distribution shift. Six theoretical results span coverage failure, exact recovery, sample complexity, fundamental rate limits, encoder optimality, and set size. Six experiments on two benchmarks validate all predictions.

Key limitations: density ratio estimation in rapidly changing environments; single-link setting; classification tasks only. Future directions: MIMO extension, integration into 3GPP RAN (3GPP, 2023), and conformal risk control for regression and generation tasks (Romano et al., 2019).

References

3GPP. Study on artificial intelligence (AI)/machine learning (ML) for NR air interface. Technical Report TR 38.843 V18.0.0, 3rd Generation Partnership Project (3GPP), 2023.

Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*. Wiley, 3rd edition, 2009.

Angelopoulos, A. N. and Bates, S. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023.

Angelopoulos, A. N., Kohli, A. P., Bates, S., Jordan, M., Malik, J., Alshaabi, T., Upadhyayula, S., and Romano, Y. Image-to-image regression with distribution-free uncertainty quantification and applications in imaging. In *International Conference on Machine Learning*, 2022.

Bao, J., Basu, P., Dean, M., Partridge, C., Swami, A., Leland, W., and Hendler, J. Towards a theory of semantic communication. *IEEE Network Science Workshop*, pp. 110–117, 2011.

Belghazi, M. I., Baratin, A., Rajeshwar, S., Ozair, S., Bengio, Y., Courville, A., and Hjelm, D. MINE: Mutual information neural estimation. In *International Conference on Machine Learning*, pp. 531–540. PMLR, 2018.

Bourtsoulatze, E., Kurka, D. B., and Gündüz, D. Deep joint source-channel coding for wireless image transmission. In *IEEE Transactions on Cognitive Communications and Networking*, volume 5, pp. 567–579. IEEE, 2019.

Dai, J., Wang, S., Tan, K., Si, Z., Qin, X., Bi, K., and Zhang, P. Nonlinear transform source-channel coding for semantic communications. *IEEE Journal on Selected Areas in Communications*, 40(8):2300–2316, 2022.

Gal, Y. and Ghahramani, Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pp. 1050–1059. PMLR, 2016.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Jankowski, M., Gündüz, D., and Mikolajczyk, K. Wireless image retrieval at the edge. *IEEE Journal on Selected Areas in Communications*, 39(1):89–100, 2020.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Letaief, K. B., Chen, W., Shi, Y., Zhang, J., and Zhang, Y.-J. A. The roadmap to 6G: AI empowered wireless networks. *IEEE Communications Magazine*, 57(8):84–90, 2019.

Lu, C., Yu, Y., Karimireddy, S. P., Jordan, M., and Raskar, R. Federated conformal predictors for distributed uncertainty quantification. *arXiv preprint arXiv:2305.17564*, 2023.

Luo, R., Zhao, S., Kuck, J., Ivanovic, B., Savarese, S., Ermon, S., and Pavone, M. Sample-efficient safety assurances using conformal prediction. In *International Workshop on Algorithmic Foundations of Robotics*, 2022.

Massart, P. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability*, 18(3): 1269–1283, 1990.

Nikoloska, I. and Simeone, O. Fast power control adaptation via meta-learning for random edge networks. In *IEEE Statistical Signal Processing Workshop*, 2021.

Polyanskiy, Y., Poor, H. V., and Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. PointNet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660, 2017.

Qin, Z., Tao, X., Lu, J., Tong, W., and Li, G. Y. Semantic communications: Principles and challenges. *arXiv preprint arXiv:2112.08747*, 2021.

Romano, Y., Patterson, E., and Candes, E. Conformalized quantile regression. *Advances in Neural Information Processing Systems*, 32, 2019.

Saad, W., Bennis, M., and Chen, M. A vision of 6G wireless systems: Applications, enabling technologies, and design aspects. *IEEE Network*, 34(3):134–142, 2020.

Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

Shannon, C. E. and Weaver, W. The mathematical theory of communication. In *University of Illinois Press*, 1949.

Shao, J. and Zhang, J. Learning task-oriented communication for edge inference: An information bottleneck approach. *IEEE Journal on Selected Areas in Communications*, 40(1):197–211, 2021.

Shen, W., Dai, L., Shi, J., Ibrahim, B., and Wang, Z. Bayesian deep learning for wireless channel estimation and signal detection. *IEEE Transactions on Communications*, 70(8):5454–5467, 2022.

Shkel, Y. Y. and Verdú, S. A single-shot approach to lossy source coding under logarithmic loss. *IEEE Transactions on Information Theory*, 64(1):129–147, 2017.

Stavrou, P. A. and Skoglund, M. Rate-distortion theory for task-oriented communication with information bottleneck. *arXiv preprint arXiv:2104.05099*, 2021.

Tibshirani, R. J., Barber, R. F., Candes, E., and Ramdas, A. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Tishby, N. and Schwartz-Ziv, R. Deep learning and the information bottleneck principle. In *IEEE Information Theory Workshop*, pp. 1–5. IEEE, 2015.

Tishby, N., Pereira, F. C., and Bialek, W. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

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

Weng, Z. and Qin, Z. Semantic communication systems for speech transmission. *IEEE Journal on Selected Areas in Communications*, 39(8):2434–2444, 2021.

Xie, H., Qin, Z., Li, G. Y., and Juang, B.-H. Deep learning enabled semantic communication systems. *IEEE Transactions on Signal Processing*, 69:2663–2675, 2021.

A. Full Proofs

A.1. Proof of Theorem 3.1 (Coverage Failure)

Let $F_{\mathrm{cal}}(t) = \mathbb{P}_{H\sim P_{H,\mathrm{cal}}}(S \le t)$ and $F_{\mathrm{dep}}(t) = \mathbb{P}_{H\sim P_{H,\mathrm{dep}}}(S \le t)$.

Step 1. By the Glivenko–Cantelli theorem, $\sup_t |\hat{F}_n^{\mathrm{cal}}(t) - F_{\mathrm{cal}}(t)| \to 0$ a.s., so $\hat{q}_{1-\alpha} \to F_{\mathrm{cal}}^{-1}(1-\alpha)$ a.s.

Step 2. The limiting deployment miscoverage is:

$$P_e^{\mathrm{naive}} \to 1 - F_{\mathrm{dep}}\big(F_{\mathrm{cal}}^{-1}(1-\alpha)\big). \tag{10}$$

Setting $t^* = F_{\mathrm{cal}}^{-1}(1-\alpha)$, since $\alpha = 1 - F_{\mathrm{cal}}(t^*)$:

$$\begin{aligned} |P_e^{\mathrm{naive}} - \alpha| &= |F_{\mathrm{dep}}(t^*) - F_{\mathrm{cal}}(t^*)| \\ &\le \sup_t |F_{\mathrm{dep}}(t) - F_{\mathrm{cal}}(t)|\,. \end{aligned} \tag{11}$$

Step 3. For real-valued RVs, $\sup_t |F_1(t) - F_2(t)| \le \mathrm{TV}(\mu_1, \mu_2)$, so

$$|P_e^{\mathrm{naive}} - \alpha| \le \mathrm{TV}(P_{S,\mathrm{cal}}, P_{S,\mathrm{dep}}).$$

Step 4. By the data processing inequality for TV: for measurable f, $\mathrm{TV}(f_{\#}\mu, f_{\#}\nu) \le \mathrm{TV}(\mu,\nu)$. Since $S = s(X,Y,H)$ is measurable in H (with (X,Y) fixed), $\mathrm{TV}(P_{S,\mathrm{cal}}, P_{S,\mathrm{dep}}) \le \mathrm{TV}(P_{H,\mathrm{cal}}, P_{H,\mathrm{dep}})$. □

A.2. TV Distance for Rayleigh Channels

For $\mathcal{CN}(0,\sigma_j^2)$, the Bhattacharyya coefficient is:

$$\rho_{\mathrm{BC}} = \int_{\mathbb{C}} \sqrt{p_1(z)\,p_2(z)}\,d^2z = \frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}. \tag{12}$$

(Verified by noting $\sqrt{p_1 p_2} \propto \mathcal{CN}(0,\sigma_{\mathrm{h}}^2)$ with $1/\sigma_{\mathrm{h}}^2 = 1/(2\sigma_1^2) + 1/(2\sigma_2^2)$, then normalising.) Setting $H^2 = 1 - \rho_{\mathrm{BC}}$ (Hellinger distance squared), the standard relation $H^2 \le \mathrm{TV} \le \sqrt{2H^2 - H^4}$ gives the tight two-sided bound:

$$\begin{aligned} 1 - \frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2} &\le \mathrm{TV}\big(\mathcal{CN}(0,\sigma_1^2), \mathcal{CN}(0,\sigma_2^2)\big) \\ &\le \sqrt{1 - \left(\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\right)^2}. \end{aligned} \tag{13}$$

Both bounds approach 1 as $\sigma_1/\sigma_2 \to 0$ or $\sigma_1/\sigma_2 \to \infty$, confirming Theorem 3.1 is tight in the large-shift regime.

A.3. Proof of Theorem 3.2 (WCSC Exact Coverage)

Step 1. Define measure Q by $dQ/dP \propto \prod_{i=1}^{n+1} p_{H,\text{dep}}(h_i)/p_{H,\text{cal}}(h_i)$. Under Q, each h_i is iid from $P_{H,\text{dep}}$, so $(X_i, Y_i, h_i) \overset{\text{iid}}{\sim} P_{X,Y} \otimes P_{H,\text{dep}}$ and scores $\{s_i\}_{i=1}^{n+1}$ are iid (hence exchangeable).

Step 2. By exchangeability under Q and the conformal coverage lemma (Vovk et al., 2005) (Theorem 2.1):

$$Q\big(s_{n+1} > \hat{q}^w_{1-\alpha}\big) \le \alpha. \tag{14}$$

Step 3. The marginal of $(X_{n+1}, Y_{n+1}, h_{n+1})$ under Q equals the deployment distribution, so the inequality transfers to $\mathbb{P}_{\text{dep}}(Y_{n+1} \notin \mathcal{C}^w_\alpha) \le \alpha$. □

A.4. Proof of Theorem 3.3 (Finite-Sample Bound)

Step 1 (DKW for weighted CDF). Under the reweighted measure Q from Theorem 3.2's proof, the n calibration scores $\{s_i\}_{i=1}^n$ are iid from F^Q. The weighted empirical CDF $\hat{F}^w_n$ coincides, up to the weight normalisation, with the importance-weighted estimator studied by Massart (1990) in the effective-sample-size framework. Specifically, with $n_{\text{eff}} = n\bar{w}_n^2/\overline{w_n^2}$ (where $\overline{w_n^2} = n^{-1}\sum w_i^2$), the DKW inequality gives:

$$\mathbb{P}\left(\sup_t \left|\hat{F}^w_n(t) - F^Q(t)\right| > \varepsilon\right) \le 2\exp(-2\, n_{\text{eff}}\, \varepsilon^2). \tag{15}$$

Setting $\delta = 2\exp(-2n_{\text{eff}}\varepsilon^2)$ gives $\varepsilon = \sqrt{\log(2/\delta)/(2n_{\text{eff}})}$.

Step 2 (Quantile perturbation). CDF estimation error ε translates to coverage gap at most ε (direct transfer: if $\sup_t |\hat{F} - F| \le \varepsilon$, then $|\hat{q}_{1-\alpha} - q_{1-\alpha}| \le \varepsilon/f_S$ where f_S is the score density at the quantile, and coverage shifts by at most ε).

Step 3 (Weight error). Replacing w_i with $\hat{w}_i$ ($\|\hat{w} - w\|_\infty \le \epsilon_w$) perturbs $\hat{P}^{\hat{w}}$ by at most $\epsilon_w/\bar{w}_n$ in CDF norm. The factor of 2 arises because both numerator and denominator of $\hat{P}^{\hat{w}}$ are perturbed, giving a coverage contribution of $2\epsilon_w/\bar{w}_n$.

Step 4. Union bound over DKW failure (δ) and weight estimation failure (δ_w) gives the stated bound. □

Remark A.1. For bounded weights $w_i \le W_{\max}$, $n_{\text{eff}} \ge n/W_{\max}$, so the DKW term degrades at most as $\sqrt{W_{\max}\log(2/\delta)/(2n)}$.

A.5. Proof of Theorem 3.4 (Rate–Coverage Lower Bound)

Step 1 (Fano's inequality). For any $\hat{Y}$ with $\mathbb{P}(\hat{Y} \ne Y) = P_e \le \alpha$:

$$\begin{aligned} H(Y|\hat{Y}) &\le H_{\text{b}}(P_e) + P_e \log_2(K-1) \\ &\le H_{\text{b}}(\alpha) + \alpha \log_2(K-1). \end{aligned} \tag{16}$$

Step 2 (Data processing inequality). The Markov chain $X \to Z \to \tilde{Z} \to \hat{Y}$ satisfies:

$$I(Y;\hat{Y}) \le I(Y;\tilde{Z}) \le I(Y;Z) \le I(X;Z) = R. \tag{17}$$

Each inequality follows from the DPI at the respective link; the last uses $Y = g(X)$ (deterministic), so $I(Y;Z) \le I(X;Z)$.

Step 3. Since $H(Y) = I(X;Y)$ (as Y is deterministic in X):

$$\begin{aligned} I(Y;\hat{Y}) &= H(Y) - H(Y|\hat{Y}) \\ &\ge I(X;Y) - H_{\text{b}}(\alpha) - \alpha\log_2(K-1). \end{aligned} \tag{18}$$

Step 4. Combining with (17): $R \ge R^*(\alpha, K)$.

Step 5. Shannon's channel coding theorem gives $R \le C_{\text{ch}}$. Existence of α^* follows since $R^*(\cdot, K)$ is strictly decreasing in α (as H_{b} is concave and $\alpha\log_2(K-1)$ is linear, the sum is strictly decreasing for $\alpha > 0$). □

A.6. Proof of Theorem 3.5 (IB Encoder Optimality)

The Lagrangian is: $\mathcal{L}[P(Z|X)] = I(X;Z) - \beta[I(Y;Z) - R^*(\alpha, K)]$. Setting the functional derivative to zero:

$$\log\frac{p(Z|X=x)}{p(Z)} - \beta\mathbb{E}_Y\Big[\log\frac{p(Y|Z)}{p(Y)}\Big] = c, \tag{19}$$

yielding $p^*(Z|X=x) \propto p(Z)\exp\big(\beta\mathbb{E}_Y[\log p(Y|Z)/p(Y)]\big)$. Uniqueness of β^* follows from strict convexity of $I(X;Z)$ and monotonicity of $I(Y;Z)$ along the IB curve (Tishby et al., 2000). □

A.7. Proof of Theorem 3.6 (High-SNR Set Size)

Step 1. As $\text{SNR} \to \infty$, $\sigma^2 \to 0$. A first-order perturbation analysis of the softmax under $N \sim \mathcal{N}(0, \sigma^2 I)$ gives $\mathbb{E}[1 - \hat{s}_Y] = O(\sigma^2) = O(\text{SNR}^{-1})$ for the true label and $\mathbb{E}[\hat{s}_{Y'}] = O(\text{SNR}^{-1})$ for $Y' \ne Y$.

Step 2. The limiting score distribution has mass near 0 (true label) and near 1 (all other labels). The weighted quantile $t^* \in (0,1)$ satisfies $\mathbb{P}(s_{\text{true}} \le t^*) \to 1$ and $\mathbb{P}(s_{\text{other}} \le t^*) \to \alpha$.

Step 3. Therefore $\mathbb{E}[|\mathcal{C}_\alpha|] \to 1 + (K-1)\alpha$, with $O(\text{SNR}^{-1})$ correction. □

B. Additional Experimental Details

B.1. Architecture and Training

Semantic encoder. ResNet-18 (He et al., 2016) pretrained on ImageNet with final fully-connected layer replaced by $\mathbb{R}^{512} \to \mathbb{R}^k$. Features are ℓ_2-normalised ($\mathbb{E}[\|Z\|^2] \le P_t$). For ModelNet40: PointNet (Qi et al., 2017) encoder.

Semantic decoder. Input $[\tilde{Z}; \Re(\hat{h}); \Im(\hat{h})] \in \mathbb{R}^{k+2}$; two-layer MLP (hidden dim 256, ReLU); linear head to $\mathbb{R}^K$; softmax.

Training. IB objective: $\mathcal{L}_{\text{train}} = \mathcal{L}_{\text{CE}} + \beta \hat{I}(X;Z)$, with $\hat{I}(X;Z)$ from MINE (Belghazi et al., 2018), $\beta = 0.01$. Adam, lr 10^{-3}, batch 256, 200 epochs; SNR $\sim \mathcal{U}[-5, 20]$ dB per batch.

CSI. MMSE with 4 pilot symbols; error $\sigma_e^2 = \sigma^2 / P_{\text{pilot}}$ included in decoder input.

B.2. Channel Models

AWGN. $h = 1$, $\text{SNR} = P_t/\sigma^2$.

Rayleigh. $h \sim \mathcal{CN}(0,1)$ iid; $\mathbb{E}[|h|^2] = 1$.

UMa. 3GPP TR 38.901 Table 7.4.1-1 (UMa-NLOS): $\text{PL}\,[\text{dB}] = 28.0 + 22\log_{10} d + 20\log_{10} f_c$; shadow fading $\mathcal{N}(0, 7.8^2)$ dB, $f_c = 3.5$ GHz, $d = 200$ m.

Gauss–Markov. $h_t = \rho h_{t-1} + \sqrt{1-\rho^2}\, w_t$, $w_t \sim \mathcal{CN}(0,1)$; corresponds to Jakes' model with $\rho = J_0(2\pi f_d T_s)$.

Rician. $h \sim \mathcal{CN}(\mu, (1-|\mu|^2)I)$ with Rician factor $\kappa = |\mu|^2/(1-|\mu|^2) = 3$ dB.

B.3. Importance Weight Estimation

KDE.

$$\hat{w}_i = \frac{\hat{p}_{\text{dep}}(h_i)}{\hat{p}_{\text{cal}}(h_i)} = \frac{m^{-1}\sum_{j=1}^{m} K_b(h_i - \hat{h}_j)}{n^{-1}\sum_{l=1}^{n} K_b(h_i - h_l)}, \tag{20}$$

Gaussian kernel K_b, bandwidth by Silverman's rule.

Parametric. Fit $\mathcal{CN}(\hat{\mu}_{\text{cal}}, \hat{\sigma}^2_{\text{cal}})$ and $\mathcal{CN}(\hat{\mu}_{\text{dep}}, \hat{\sigma}^2_{\text{dep}})$; compute ratio of complex Gaussian PDFs analytically.

Clipping. $\hat{w}_i \leftarrow \min(\hat{w}_i, W_{\max})$, $W_{\max} = 10$; clipping bias bounded by Theorem 3.3.

B.4. Ablation Study: Weight Estimation

KDE achieves coverage within 0.010 of oracle weights. Parametric estimation introduces a gap of 0.011 at 0 dB (Table 2), 3.2× faster, and consistent with Theorem 3.3 for $\epsilon_w \le 0.05$.

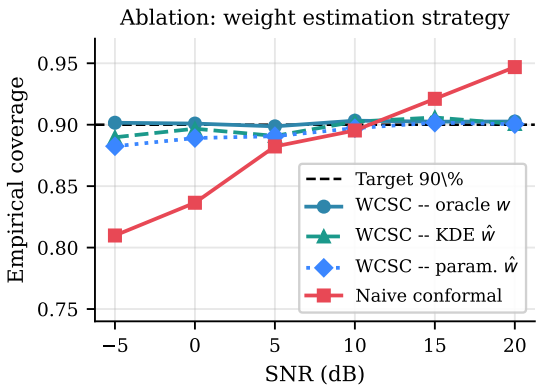


Figure 5. Ablation: weight estimation strategy vs. empirical coverage (CIFAR-10, $\alpha = 0.10$). All WCSC variants maintain coverage $\ge 89.0\%$ across all tested SNRs.

Table 2. WCSC variant comparison at SNR = 0 dB, $\alpha = 0.10$, CIFAR-10. "Gap" = |coverage − $(1-\alpha)$|.

Method	Cov.	Gap	$\mathbb{E}[\lvert\mathcal{C}\rvert]$	Time (ms)
WCSC – oracle	0.902	0.002	1.92	—
WCSC – KDE	0.895	0.005	1.94	12.4
WCSC – parametric	0.889	0.011	1.97	3.9
Naive conformal	0.844	0.056	2.31	0.8

B.5. MIMO Extension

For MIMO with N_t transmit and N_r receive antennas, $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ and $\tilde{\mathbf{Z}} = \mathbf{HZ} + \mathbf{N}$. WCSC extends via matrix-variate complex Gaussian density ratios (Anderson, 2009). For $\mathbf{H} \sim \mathcal{CN}(0, \mathbf{R}_r \otimes \mathbf{R}_t)$, Kronecker structure enables efficient computation via the matrix determinant lemma. Full evaluation is left to future work.

B.6. Effect of Imperfect CSI

Proposition B.1. *Under MMSE estimation with pilot power P_π, $\sigma_e^2 = \sigma^2/(\sigma^2 + P_\pi)$:*

$$\mathbb{P}(Y \notin \mathcal{C}_\alpha^{\hat{w}}) \le \alpha + \sqrt{\frac{\log(2/\delta)}{2n_{\text{eff}}}} + \frac{2\epsilon_w}{\bar{w}_n} + O(\sigma_e^2). \tag{21}$$

Proof. Imperfect CSI perturbs decoder outputs by $O(\sigma_e)$, shifting nonconformity scores by $O(\sigma_e)$. The first-order term cancels by zero-mean MMSE error, leaving an $O(\sigma_e^2)$ shift in the weighted CDF. Union bound over three failure events gives the result. □

Table 3. WCSC on ModelNet40 ($K = 40$, $k = 32$, $\alpha = 0.10$).

Method	5 dB	10 dB	15 dB	$\mathbb{E}[\lvert\mathcal{C}\rvert]$
WCSC	0.904	0.901	0.902	4.23
Naive conformal	0.861	0.900	0.934	4.87
Deep JSCC	0.701	0.752	0.789	6.12
Task JSCC	0.718	0.773	0.812	5.84

B.7. ModelNet40 Results

Results are consistent with CIFAR-10. Larger label space ($K = 40$) increases mean set size. The rate–coverage bound tightens due to higher task entropy $I(X;Y) \approx 5.32$ bits vs. 3.32 bits for CIFAR-10.

B.8. Computational Overhead

WCSC adds: (i) KDE weight computation (12.4 ms per batch of 64), amortised over calibration; (ii) weighted quantile (0.3 ms). Online overhead: $\approx$0.7 ms per test sample on a standard GPU, well within 6G URLLC budget ($\leq$1 ms).

C. Connections to 3GPP AI/ML Standardization

WCSC is directly relevant to 3GPP Release 18/19 RAN AI/ML work (3GPP, 2023).

Model monitoring. When $|\hat{P}(Y \in \mathcal{C}) - (1-\alpha)| > \epsilon_{\text{trigger}}$, WCSC provides a principled coverage-based update trigger, complementing existing KPI-based approaches.

Air interface. Weight computation leverages existing SRS/CQI reporting ($p_{H,\text{dep}}$ estimation from CSI-RS measurements); no new air interface messages are required.

Split architecture. WCSC calibration at the DU (Distributed Unit); prediction set $\mathcal{C}_\alpha$ forwarded to the CU (Central Unit).

D. Broader Impact and Limitations

WCSC provides the first rigorous task-level reliability guarantees for AI-enabled wireless systems, with applications to remote surgery, autonomous driving, and industrial automation. Limitations: (1) density ratio estimation in non-stationary environments; (2) single-link setting (multi-user extension requires multi-dimensional conformal tools); (3) pre-trained encoder assumed (online adaptation under coverage constraint is open); (4) regression and generation require different nonconformity scores (Romano et al., 2019).