Beyond Point Estimates: Likelihood-Based Full-Posterior Wireless Localization

Haozhe Lei*[1], Hao Guo[1,2], Tommy Svensson[2], and Sundeep Rangan[1]
[1]NYU WIRELESS, Tandon School of Engineering, New York University, Brooklyn, NY 11201, USA
[2]Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden
{hl4155, hg2891, srangan}@nyu.edu; tommy.svensson@chalmers.se

***Abstract*—Modern wireless systems require not only position estimates, but also quantified uncertainty to support planning, control, and radio resource management. We formulate localization as posterior inference of an unknown transmitter location from receiver measurements. We propose *Monte Carlo Candidate-Likelihood Estimation (MC-CLE)*, which trains a neural scoring network using Monte Carlo sampling to compare true and candidate transmitter locations. We show that in line-of-sight simulations with a multi-antenna receiver, MC-CLE learns critical properties including angular ambiguity and front-to-back antenna patterns. MC-CLE also achieves lower cross-entropy loss relative to a uniform baseline and Gaussian posteriors. alternatives under a uniform-loss metric.**

***Index Terms*—Wireless localization, Full-posterior inference, Monte Carlo likelihood, Neural density estimation**

I. Introduction

Wireless localization plays a central role in a wide range of applications and is now an integral component of the 5G standard [1], [2]. This paper considers the problem of estimating the location $\boldsymbol{x}^t$ of a wireless transmitter using measurements collected by one or more receivers. Conventional approaches typically yield a point estimate $\widehat{\boldsymbol{x}}^t$ based on some observations $\boldsymbol{y}$. However, in many practical scenarios, these estimates are accompanied by significant uncertainty, and downstream tasks increasingly demand a richer characterization of that uncertainty, or so-called uncertainty awareness [3]. Applications such as robust path planning, risk-aware control, and adaptive beamforming or handover decisions [4]–[9] rely not just on a single location guess, but on an understanding of where the transmitter could plausibly be located—and with what confidence.

To meet these demands, this work considers posterior inference instead of point estimation. Specifically, we aim to estimate the posterior distribution

$$p(\boldsymbol{x}^t|\boldsymbol{y}), \tag{1}$$

which represents the probability of the transmitter being at location $\boldsymbol{x}^t$ given the observations $\boldsymbol{y}$. For given observations $\boldsymbol{y}$, the posterior $p(\boldsymbol{x}^t|\boldsymbol{y})$ can be visualized as a spatial heat map, revealing regions of high and low likelihood across the search space. Figure 1 illustrates such a heat map generated by our proposed method, Monte Carlo Conditional Likelihood Estimation (MC-CLE), which we introduce in detail below.

* Corresponding author: Haozhe Lei (hl4155@nyu.edu).

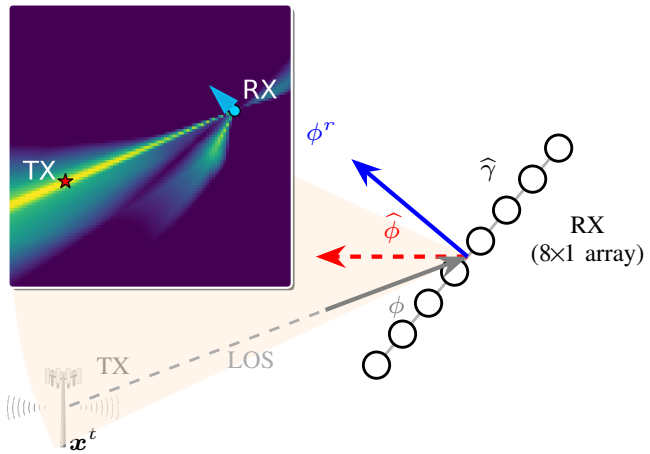

Fig. 1: Example learned posterior $p(\boldsymbol{x}^t \mid \boldsymbol{y})$ for an unknown transmitter location $\boldsymbol{x}^t$ from receiver measurements $\boldsymbol{y}$ using an 8×1 uniform linear array. The posterior is visualized as a heat map.

Challenges, even under line-of-sight: We consider the problem of estimating the posterior distribution (1) from an unknown transmitter location from receiver measurements in a line-of-sight (LOS) channel. The receiver is equipped with an array and performs matched filtering to detect the angle and SNR of a known transmitted signal, and the posterior is to be estimated from these measurements along with the receiver pose. Even in this simple LOS case, we show that the posterior distribution can be complex: (1) Most importantly, the posterior is frequently a mixture distribution due to false alarms—typically manifesting as a dominant peak corresponding to the true path and a secondary residual distribution arising from noise-induced detections. (2) Array geometries often introduce angular ambiguities; for instance, uniform linear arrays (ULAs) produce identical responses at angles ϕ and $\pi - \phi$. (3) Antenna elements, such as patch antennas, exhibit strong directionality, making the received signal highly sensitive to the receiver's orientation.

Contributions

Our contributions are as follows: (i) We formalize localization as posterior density estimation over a 2D scene (Section II) and discuss several challenges for this problem even

in the LOS case (Section II-B). (ii) We introduce *Monte Carlo Candidate-Likelihood Estimation (MC-CLE)*, which learns the likelihood by comparing the score of the true location to randomly selected candidate locations. (iii) MC-CLE is evaluated with realistic antenna models, including antenna patterns with significant front-to-back gain. It is shown that MC-CLE captures many of the key features of the conditional distribution, including distribution mixtures, multi-modalities from angular ambiguity, and directionality. The method outperforms parametric methods such as Gaussian conditional distributions (Section IV).

Prior work

The overwhelming majority of conventional RF localization methods provide point estimates. The characterization of the uncertainty of these estimates is much less understood. The Cramér-Rao bound is widely-used for bounding the error of unbiased estimates [10]. However, applying the CR bound requires a model of the measurements and is typically only achieved in the high SNR regime. Related work [4] learns the standard deviation of angular errors; [5] quantifies errors in triangulization; and [11] characterizes the effect of location uncertainty in the anchor nodes. None of these methods provide a full spatial characterization of the spatial posterior as considered in this work.

Bayesian approaches maintain posterior maps and fuse heterogeneous measurements via particle filtering [12] and factor-graph belief propagation [13], with von Mises–Fisher angle noise improving likelihood shape [14], yet they rely on well-specified likelihoods, fixed grids or tuned proposals, and significant computation, and can be biased under model mismatch.

The general problem of learning conditional distributions is a classic problem in statistics. Non-parametric techniques, such as kernel estimation, are widely-used – see the survey [15]. Closely related are so-called localization methods [16] that locally fit logistic models using kernel weights. As non-parametric methods generally suffer from the curse of dimensionality, this work attempts to learn a parametric approximation of the likelihood $g_\theta(\boldsymbol{x}^t, \boldsymbol{y}) \approx \ln p(\boldsymbol{y}|\boldsymbol{x}^t)$. The well-known challenge in the parametric method is that the normalization constant or partition function $Z_\theta(\boldsymbol{y})$ in (7) involves an integral that does not have a closed-form expression. One approach is to discretize the integral [17], [18]. As we discuss below, our method can be seen as Monte-Carlo sampling.

II. Problem Formulation

A. Posterior Density Estimation

Consider the problem of locating a transmitter (TX) at an unknown position, $\boldsymbol{x}^t \in \mathbb{R}^d$. For simplicity, we consider the 2D localization problem shown in Figure 1, so $d = 2$. We assume that TX localization has a known prior $p_0(\boldsymbol{x}^t)$. Typically, we take $p_0(\boldsymbol{x}^t)$ to be uniform in some compact set $\mathcal{A} \subset \mathbb{R}^d$, but any prior distribution may be used. A receiver (RX) is at a known 2D *pose*, $(\boldsymbol{x}^r, \phi^r)$, where $\boldsymbol{x}^r \in \mathbb{R}^d$ is the RX position and ϕ^r is its azimuth orientation.

To locate the TX, the TX broadcasts some periodic signals. For example, the TX could broadcast IEEE 802.11 Wi-Fi beacon frames every 100 ms [19] in a small room, IEEE 802.15.4z UWB blink pulses in large indoor warehouses [20], or GPS L1 C/A codes every 1 ms [21] or 5G NR SSBs every 20 ms [22] in outdoor spaces, enabling the RX the angle of arrival (AoA) and SNR for localization.

In this work, we focus on a single path LOS channel with a single measurement circumstance (we will discuss multiple measurements in future work). We assume the RX applies a matched filter to the TX signal. The resulting output is given by [23]:

$$\boldsymbol{z}[n] = gB(\phi - \phi^r)\boldsymbol{a}(\phi - \phi^r)\mathrm{sinc}((\tau - nT)/T) + \boldsymbol{w}[n], \quad (2)$$

where, $\boldsymbol{z}[n] \in \mathbb{C}^{N_r}$ is the filtered response on each of the N_r antennas at a sample n; g is a complex channel gain; ϕ is the true AoA in the global coordinate system; ϕ^r is the orientation of the RX; $B(\cdot)$ is the element response of the antenna; τ is time of arrival relative to the sampling window; and T is the sampling period. We assume that the angle of arrival is then estimated by maximizing the correlation

$$\widehat{\phi} = \arg\max_{\phi} \max_{n} |\boldsymbol{a}(\phi - \phi^r)^{\mathsf{H}} \boldsymbol{z}[n]|^2. \quad (3)$$

The SNR is estimated from the value of the maximum:

$$\widehat{\gamma} = \max_{n} |\boldsymbol{a}(\widehat{\phi} - \phi^r)^{\mathsf{H}} \boldsymbol{z}[n]|^2. \quad (4)$$

We let $\boldsymbol{y}$ denote the observations:

$$\boldsymbol{y} = (\boldsymbol{x}^r, \phi^r, \widehat{\phi}, \widehat{\gamma}), \quad (5)$$

which includes both the receiver's position and orientation $(\boldsymbol{x}^r, \phi^r)$ as well as the the estimated angle of arrival $\widehat{\phi}$ and estimated SNR $\widehat{\gamma}$. Our basic problem is to estimate the posterior density $p(\boldsymbol{x}^t|\boldsymbol{y})$, which represents the conditional distribution of the TX location from the observations $\boldsymbol{y}$.

B. Challenges

There are at least three complexities in the posterior from such angular estimates:

False alarms: The estimate (3) is computed from the maximum of a number of angular hypotheses. In low SNR settings, the strongest angular may correspond to noise, i.e., a false alarm. The resulting posterior distribution will thus, in general, be a mixture of the distribution corresponding to a correct detection and the case where the peak arose from noise. Each component of this distribution will have significantly different shapes.

Angle ambiguity. For uniform arrays, this array response $\boldsymbol{a}(\phi)$ typically has angular symmetries. For example, in a uniform linear array, $\boldsymbol{a}(\phi)$ is a function of $\sin(\phi)$, so an angle ϕ and $\pi - \phi$ have identical responses. This ambiguity creates a multi-modal distribution over two or more potential angles.

Antenna Gain–strength coupling. The antenna directional gain $B(\phi)$ of the RX enters the matched-filter statistics multiplicatively with the path gain α. In most antennas, particularly patch antennas, the antenna pattern is highly directional with

significant front-to-back gain. As a result, the measured signal strength $|z|$ conflates two factors: true path loss and angular attenuation due to the beam pattern.

III. Proposed Solution

In this section, we introduce Monte Carlo Candidate Likelihood Estimation (MC-CLE), a data-driven framework for approximating the posterior distribution $p(\boldsymbol{x}^t|\boldsymbol{y})$ over a finite set of candidate transmit positions.

A. Log Likelihood Estimation

We consider an estimate of the posterior probability distribution $p(\boldsymbol{x}^t|\boldsymbol{y})$, of the form:

$$\widehat{p}_\theta(\boldsymbol{x}^t|\boldsymbol{y}) = \frac{1}{Z_\theta(\boldsymbol{y})} e^{g_\theta(\boldsymbol{x}^t,\boldsymbol{y})} p_0(\boldsymbol{x}^t), \tag{6}$$

where $p_0(\boldsymbol{x}^t)$ is the known prior distribution on $\boldsymbol{x}^t$; $g_\theta(\cdot)$ is a learned function of the position $\boldsymbol{x}^t$, observations $\boldsymbol{y}$ and parameters θ; and $Z_\theta(\boldsymbol{y})$ is the partition function

$$Z_\theta(\boldsymbol{y}) = \int e^{g_\theta(\boldsymbol{x}^t,\boldsymbol{y})} p_0(\boldsymbol{x}^t)\, d\boldsymbol{x}^t. \tag{7}$$

Under the distribution hypothesis (6), the corresponding estimate for the likelihood of $\boldsymbol{x}^t$ given $\boldsymbol{y}$ can be found from Bayes' Rule:

$$\widehat{p}(\boldsymbol{y}|\boldsymbol{x}^t) = \frac{\widehat{p}(\boldsymbol{x}^t|\boldsymbol{y})p(y)}{Z_\theta(\boldsymbol{y}} = \frac{e^{g_\theta(\boldsymbol{x}^t,\boldsymbol{y})}}{Z_\theta(\boldsymbol{y})}, \tag{8}$$

and $g_\theta(\boldsymbol{x}^t, \boldsymbol{y})$ can be considered as an approximation of an un-normalized log likelihood:

$$g_\theta(\boldsymbol{x}^t, \boldsymbol{y}) \approx \log p(\boldsymbol{x}^t|\boldsymbol{y}) + \log \frac{Z_\theta(\boldsymbol{y})}{p(y)}. \tag{9}$$

Hence, in the sequel, we will call $g_\theta(\cdot)$ the *un-normalized log likelihood estimate*. As we will describe below, we parametrize $g_\theta(\cdot)$ as a neural network and learn its parameters θ directly from data.

Ideally, to train $g_\theta(\cdot)$, we would minimize a cost such as the cross-entropy [24]:

$$\begin{aligned} H(p, \widehat{p}_\theta) &:= -\mathbb{E}\left[\log \widehat{p}(\boldsymbol{x}|\boldsymbol{y})\right] \\ &= -\mathbb{E}\left[g_\theta(\boldsymbol{x}^t, \boldsymbol{y})\right] + H(p_0) + \mathbb{E}\left[\log Z_\theta(\boldsymbol{y})\right], \end{aligned} \tag{10}$$

where the expectation is with respect to the true joint density $p(\boldsymbol{x}^t, \boldsymbol{y})$ and $H(p_0) = -\mathbb{E} \log p_0(\boldsymbol{x}^t)$ is the entropy of the prior.

B. Sampled Cross Entropy Loss

The problem in the cross-entropy loss (10) is the expectation of the log partition function, $\mathbb{E}\left[\log Z_\theta(\boldsymbol{y})\right]$. This normalization constant typically does not have an analytic expression when $g_\theta(\cdot)$ is a neural network. To avoid this problem, we draw a *candidate set* $\mathcal{C} = \{\bar{\boldsymbol{x}}_k^t\}_{k=1}^K \subset \mathbb{R}^d$ and approximate the integral in (7) by the Monte-Carlo sum

$$\widehat{Z}_\theta(\boldsymbol{y}) = \frac{1}{K} \sum_{k=1}^{K} e^{g_\theta(\bar{\boldsymbol{x}}_k^t, \boldsymbol{y})}. \tag{11}$$

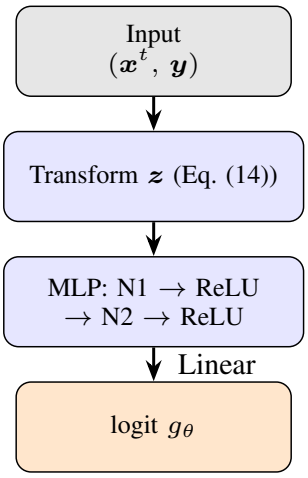

Fig. 2: Architecture of the MC-CLE model.

From the Law of Large Numbers, we have that

$$\widehat{Z}_\theta(\boldsymbol{y}) \to Z_\theta(\boldsymbol{y}). \tag{12}$$

as $K \to \infty$ for any fixed $\boldsymbol{y}$.

Now suppose, that for training, we have data samples $(\boldsymbol{x}_i^t, \boldsymbol{y}_i)$, $i = 1, \ldots, n$, with true TX locations $\boldsymbol{x}_i^t$ along with measurements $\boldsymbol{y}_i$. For each data sample i, we draw candidates $\overline{\boldsymbol{x}}_{ik}$, $k = 1, \ldots, K$. Replacing the expectation in (10), and substituting the true log partition $\log Z_\theta(\boldsymbol{y}_i)$ with the Monte-Carlo estimate $\widehat{Z}_\theta(\boldsymbol{y}_i)$, we obtain the *sampled cross-entropy loss (CEL)* as:

$$\mathcal{L}(\theta) := \frac{1}{n} \sum_{i=1}^{n} \left[-g_\theta(\boldsymbol{x}_i^t, \boldsymbol{y}_i) + \log \left[\frac{1}{K} \sum_{k=1}^{K} e^{g_\theta(\overline{\boldsymbol{x}}_{ik}^t, \boldsymbol{y}_i)} \right] \right]. \tag{13}$$

We can then learn the parameters θ by minimizing this loss.

C. Position Estimation with Neural Networks

Having introduced the sampled cross-entropy loss in Section III-B, we now specify the function class for $g_\theta(\boldsymbol{x}^t, \boldsymbol{y})$.

Feature transformation: In the first step of the network, we apply a fixed transform to the input

$$(\boldsymbol{x}^t, \boldsymbol{y}) = (\boldsymbol{x}^t, \boldsymbol{x}^r, \phi^r, \widehat{\phi}, \widehat{\gamma}),$$

to yield a feature vector $\boldsymbol{z}$ given by:

$$\boldsymbol{z} = [\boldsymbol{u}, d, s, \cos(\phi^r), \sin(\phi^r), \cos(\widehat{\phi}), \cos(\widehat{\phi})], \tag{14}$$

where $\boldsymbol{u}$ is a unit vector in the direction of $\boldsymbol{x}^t - \boldsymbol{x}^r$,

$$\boldsymbol{u} = \frac{\boldsymbol{x}^t - \boldsymbol{x}^r}{\|\boldsymbol{x}^t - \boldsymbol{x}^r\|}. \tag{15}$$

The feature d is a scaled logarithmic distance:

$$d = \log_{10}(\max\{1, \|\boldsymbol{x}^t - \boldsymbol{x}^r\|), \tag{16}$$

where the distance $\|\boldsymbol{x}^t - \boldsymbol{x}^r\|$ is in meters; and the feature s is a scaled SNR in dB:

$$s = \min\left\{ \max\left\{ 0, \frac{10}{s_{\max}} \log_{10}(\widehat{\gamma}), 1 \right\} \right\}, \tag{17}$$

where $s_{\max} = 60\,\mathrm{dB}$. This particular transformation is physically-motivated, i.e., the likelihood should only depend on the relative position $\boldsymbol{x}^t - \boldsymbol{x}^r$.

In free-space propagation, the distance effect is logarithmic [23] – hence the use of the logarithm in (16). The clipping in both (16) and (17) keep the variables in reasonable ranges for typical observed values. The use of the sines and cosines of the angles in (14) keeps the functions continuous modulo 2π.

Architecture: The transform feature vector $\boldsymbol{z}$ in (14) has dimension $N_z = 8$. We then realize $g_\theta(\boldsymbol{x}^t, \boldsymbol{y})$ as two hidden linear layers of widths $N_1 = 64$ and $N_2 = 16$, each followed by ReLU. The last linear layer outputs a scalar – see Figure 2.

Training objective: All parameters θ are trained end-to-end by minimizing the sampled cross-entropy loss (13).

Gaussian Baselines: To demonstrate the utility of the method, we compare against the case of a *Gaussian posterior*. Specifically,

$$g_\theta(\boldsymbol{x}^t, \boldsymbol{y}) := -\frac{1}{2}(\boldsymbol{x}^t - \boldsymbol{\mu})^\intercal \boldsymbol{Q}(\boldsymbol{x}^t - \boldsymbol{\mu}), \tag{18}$$

where the mean $\mu \in \mathbb{R}^d$ and covariance distribution $\boldsymbol{Q} \in \mathbb{R}^{d\times d}$ are neural network functions of the observations using the same structure except the output in Figure 2. We consider both the case where the Gaussian is in Cartesian and polar coordinates.

IV. Numerical Experiments

To evaluate MC-CLE, we use Sionna [25] as the simulation platform. All experiments operate at carrier frequency $12\,\mathrm{GHz}$ and bandwidth of $200\,\mathrm{MHz}$. This frequency is in the upper mid-band [26], which has attracted considerable interest for next generation cellular systems as well as localization [27]. The TX is modeled as a single isotropic antenna at a fixed, known height. The RX uses the 8×1 azimuth-plane uniform linear array, mounted at the same elevation as the TX to restrict the problem to two-dimensional localization. We assume a 3GPP antenna pattern [28] which has a $3\,\mathrm{dB}$ beamwidth of $65°$. Both TX and RX are generated uniformly in a $100 \times 100\,\mathrm{m}$ free-space environment, yielding a LOS scenario. Since the distribution is uniform in the region, the prior $p_0(\boldsymbol{x}^t)$ is constant. These settings are identical for all posterior models (MC-CLE, Gauss–Cart, and Gauss–Polar).

We adopt the network topology shown in Section III-C. A total of $100{,}000$ TX–RX pairs are generated, and a $75\%/25\%$ train–validation split is used to select the best epoch during training. All models use the same training configuration for $2{,}000$ epochs; MC-CLE uses AdamW with a learning rate 0.01, and the Gaussian models use a learning rate 0.002. The MC-CLE training used 1.5 GPU-hours, while each Gaussian model used 2.5 GPU-hours, on NVIDIA RTX A6000.

A. CEL Performance

Two evaluation sets of $25{,}000$ samples each are prepared:

- **Grid-based set:** candidate positions are placed on a regular grid with approximately $3.33\,\mathrm{m}$ spacing, resulting in 961 candidate positions within the $100 \times 100\,\mathrm{m}$ area.

TABLE I: Evaluation losses for each model on grid and random evaluation sets.

Dataset	MC-CLE	Gauss–Cart	Gauss–Polar
Eval Grid	-3.4703	-2.3441	-1.6788
Eval Random	-3.4259	2.3071	-1.6634

TABLE II: Evaluation results reported as geometric improvement $\mathcal{G}$ and gap-closure ratio $\mathcal{R}$ (higher is better). K=961 for the grid set and K=1000 for the random set.

Dataset	MC–CLE		Gauss–Cart		Gauss–Polar	
	$\mathcal{G}$	$\mathcal{R}$	$\mathcal{G}$	$\mathcal{R}$	$\mathcal{G}$	$\mathcal{R}$
Eval Grid	32.15	50.53%	10.42	34.13%	5.36	24.44%
Eval Random	30.75	49.59%	10.05	33.40%	5.28	24.08%

- **Random set:** $1{,}000$ candidate positions are selected uniformly at random within the same area.

Given the sampled CEL $\mathcal{L}$ from (13) and the number of candidates $K = \{961, 1000\}$ for the sets $\{$Grid-based, Random$\}$, we report the results in Table I.

To interpret the loss $\mathcal{L}$ in (13), first observe that a uniform likelihood estimate, $g_\theta(\boldsymbol{x}^t, \boldsymbol{y}) =$ constant, results in $\mathcal{L} = 0$. Also, recall that the candidates $\overline{\boldsymbol{x}}^t_{ik}$ are selected such that the true position $\boldsymbol{x}^t_i$ is one of the candidates. That is, $\boldsymbol{x}^t_i = \overline{\boldsymbol{x}}^t_{ik_0}$ for some $k_0 = \sigma(i)$. Therefore, the sampled CEL in (13) is bounded above by:

$$\begin{aligned}\mathcal{L}(\theta) &\le \frac{1}{n}\sum_{i=1}^{n}\left[-g_\theta(\boldsymbol{x}^t_i, \boldsymbol{y}_i) + \log\left[\frac{1}{K}e^{g_\theta(\boldsymbol{x}^t_i, \boldsymbol{y}_i)}\right]\right] \\ &= -\log(K),\end{aligned} \tag{19}$$

where the bound is achieved when the likelihood concentrates around the true location $\boldsymbol{x}^t = \boldsymbol{x}^t_i$. Hence, the loss is bounded by

$$-\log(K) \le \mathcal{L} \le 0. \tag{20}$$

This motivates two metrics:

$$\mathcal{G} \triangleq e^{-\mathcal{L}}, \quad \mathcal{R} \triangleq \frac{-\mathcal{L}}{\log K} \times 100\%. \tag{21}$$

Here $\mathcal{R} = 0\%$ corresponds to uniform prediction ($\mathcal{L}_{\text{adj}} = 0$), while $\mathcal{R} = 100\%$ corresponds to a perfect prediction $\mathcal{L}_{\text{adj}} = -\log K$. The results are shown in Table II, and shows that the proposed method MC-CLE significantly outperforms the Gaussian baselines.

B. Probability Mass Visualization

To visualize the ability of MC-CLE to capture the complexities of the posterior distribution, Figure 3 plots 30 panels (3 models × 10 geometries) of the estimated probability distributions. Each panel illustrates the estimated posterior distribution $\log p(\boldsymbol{x}^t, \boldsymbol{y})$ for one true TX location (red circle) and one RX location (blue circle) with the blue arrow representing the RX orientation. The label for each plot $\Delta x, \Delta y$ represents the relative TX-RX position. We observe that the learned distributions can capture significant features not possible under the assumption of a Gaussian posterior. For example, when the

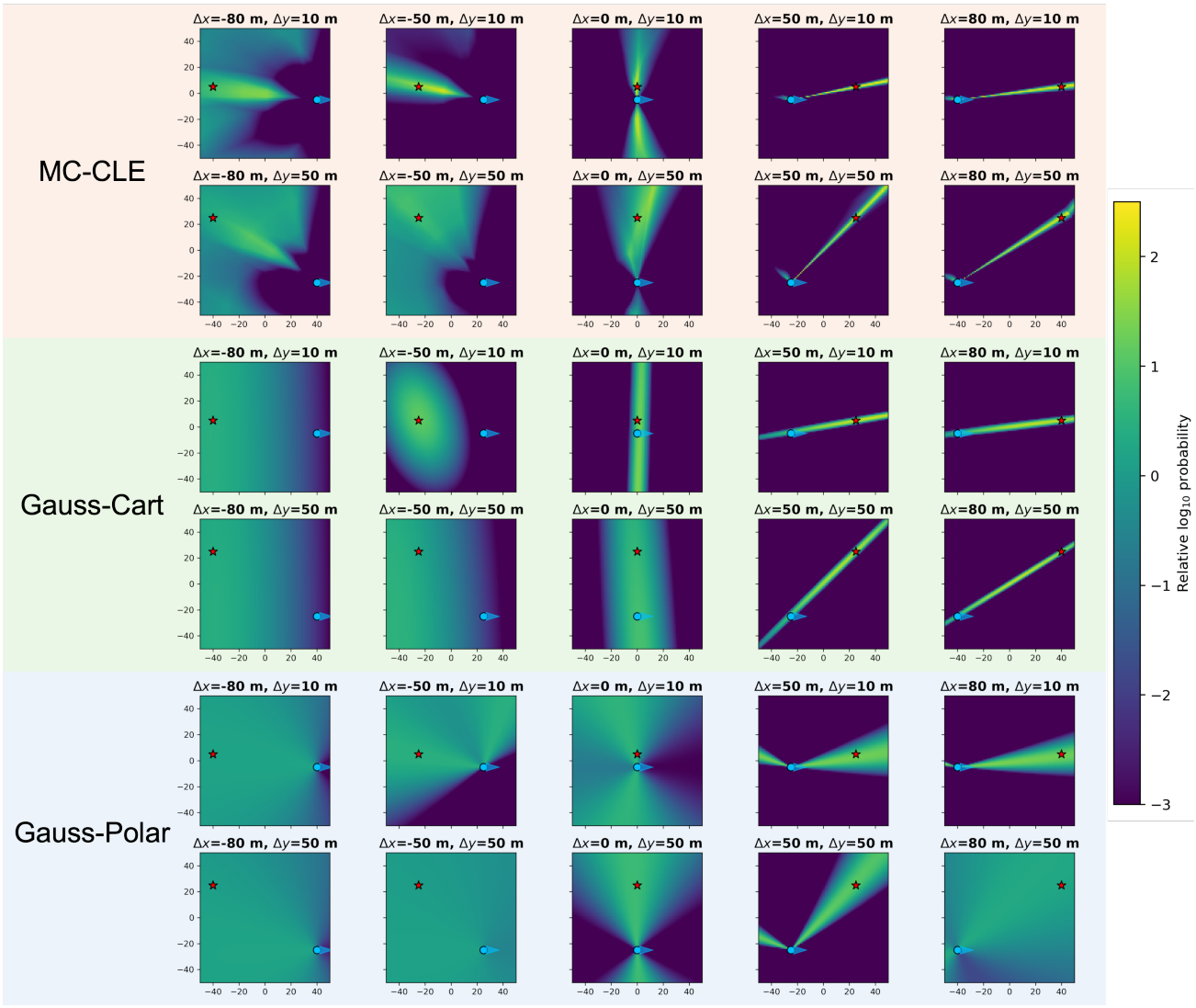

Fig. 3: Relative log-probability maps for two RX–TX geometries (see Table III). Columns sweep the horizontal offset $\Delta x \in \{-80, -50, 0, 50, 80\}$ m. Each panel spans a 100×100 m region discretized on a dense grid. The red star marks the TX, the blue circle the RX, and the cyan arrow its heading. Colour encodes the relative log-probability $L(\boldsymbol{x}_k) = \log\big(p_\theta(\boldsymbol{x}_k, \boldsymbol{y})\, K_{\text{grid}}\big)$; warmer colours indicate larger $L(\boldsymbol{x}_k)$. Each horizontal strip corresponds to one model: MC–CLE (top), Gauss–Cart (middle), and Gauss–Polar (bottom). MC–CLE produces sharper ridges aligned with the TX–RX geometry; Gauss–Cart yields smoother, elongated bands; Gauss–Polar concentrates mass in narrower angular sectors with less accurate peak localization. When the RX faces the TX, high-confidence ridges appear; side- or back-looking orientations spread the posterior into broader, lower-confidence regions.

TABLE III: Five reused RX offset/heading cases ($\boldsymbol{x}^t$ is the fixed TX position). Each row lists the RX offset $\boldsymbol{x}^r = (\Delta x, -\Delta y)$ and heading ϕ^r, using 0° = right$(+x)$ and 90° = up$(+y)$.

Row	Δy (m)	ϕ^r ($^\circ$)	Brief description
1	10	0	RX 10 m below TX, heading right
2	50	0	RX 50 m below TX, heading right

TX is behind the RX ($\Delta x = -80$ and -50 m), the signal is weak due to the front-to-back gain of the antenna. The MC-CLE captures the mixture distribution with one component in an angular cone corresponding to the estimated angle, and a second component corresponding to the case where the TX location is not known, but somewhere behind the RX. This mixture distribution cannot be captured in the Gaussian posteriors. Additionally, when $\Delta x = 50$ and 80 m, the MC-CLE method is able to concentrate the distribution to a cone around the true angle ϕ, but there is a second small cone for the angle $\pi - \phi$.

Furthermore, we evaluate overall concentration via Shannon

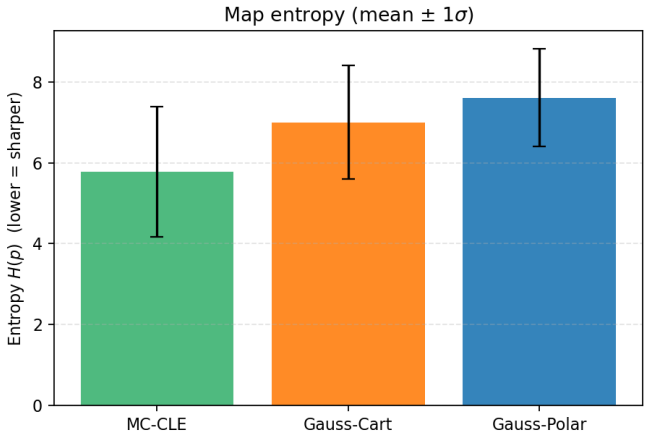

Fig. 4: Posterior entropy across candidates, averaged over episodes for each model with error bars indicating $\pm 1\sigma$. Lower entropy indicates more concentrated posteriors.

entropy in Figure 4:

$$H(p) \triangleq -\sum_{i=1}^{K} p_i \log p_i. \tag{22}$$

Lower entropy indicates that probability mass is focused on fewer regions. Empirically, MC-CLE yields lower entropy than the Gaussian baselines, indicating more concentrated posteriors.

V. Conclusion

This paper presents a simple method, *Monte Carlo Candidate-Likelihood Estimation (MC-CLE)*, for estimation of the posterior distribution of the TX from RX measurements. In a LOS setting, MC-CLE improves cross-entropy over Gaussian baselines and a uniform predictor, produces geometry-consistent ridges in relative log-probability maps, and yields lower posterior entropy. A calibrated spatial posterior provides a reliable bridge from signals to decisions, such as planning, control, and resource management. Future work will consider how to extend these methods to NLOS settings, with and without knowledge of the environment and obstacles.

References

[1] 3GPP, "Study on expanded and improved NR positioning," 3rd Generation Partnership Project (3GPP), 3GPP Technical Report 38.859 V18.1.0, Jul. 2024, release 18.

[2] H.-S. Cha, G. Lee, A. Ghosh, M. Baker, S. Kelley, and J. Hofmann, "5G NR positioning enhancements in 3gpp release-18," 2024. [Online]. Available: https://arxiv.org/abs/2401.17594

[3] S. Wang, W. Dai, J. Sun, Z. Xu, and G. Y. Li, "Uncertainty awareness in wireless communications and sensing," *IEEE Commun. Mag.*, 2025.

[4] T. Zhang, D. Zhang, G. Wang, Y. Li, Y. Hu, Q. Sun, and Y. Chen, "Rloc: Towards robust indoor localization by quantifying uncertainty," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, no. 4, pp. 1–28, 2024.

[5] F. Shamsfakhr, A. Antonucci, L. Palopoli, D. Macii, and D. Fontanelli, "Indoor localization uncertainty control based on wireless ranging for robots path planning," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[6] M. Yin, T. Li, H. Lei, Y. Hu, S. Rangan, and Q. Zhu, "Zero-shot wireless indoor navigation through physics-informed reinforcement learning," in *Proc. 2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 5111–5118.

[7] T. Li, H. Lei, H. Guo, M. Yin, Y. Hu, Q. Zhu, and S. Rangan, "Digital twin-enhanced wireless indoor navigation: Achieving efficient environment sensing with zero-shot reinforcement learning," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 2356–2372, 2025.

[8] T. Li, H. Lei, M. Yin, and Y. Hu, "Reinforcement learning with physics-informed symbolic program priors for zero-shot wireless indoor navigation," in *Proc. Reinforcement Learning Conf. (RLC)*, 2025.

[9] A. Shahmansoori, G. E. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," vol. 17, no. 3, pp. 1822–1835, Mar. 2018.

[10] H. Godrich, A. M. Haimovich, and R. S. Blum, "Cramér–rao bound on target localization estimation in MIMO radar systems," in *Proc. Conf. Inf. Sci. Syst. (CISS)*. IEEE, 2008, pp. 134–139.

[11] L. S. Muppirisetty, T. Svensson, and H. Wymeersch, "Spatial wireless channel prediction under location uncertainty," vol. 15, no. 2, pp. 1031–1044, 2015.

[12] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian bayesian tracking," vol. 50, no. 2, pp. 174–188, Feb. 2002.

[13] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," vol. 97, no. 2, pp. 427–450, Feb. 2009.

[14] H. Nurminen, L. Suomalainen, S. Ali-Löytty, and R. Piché, "3D angle-of-arrival positioning using von Mises–Fisher distribution," in *Proc. Int. Conf. Information Fusion (FUSION)*, 2018, pp. 2036–2041.

[15] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Routledge, 2018.

[16] P. Hall, R. C. L. Wolff, and Q. Yao, "Methods for estimating a conditional distribution function," *J. Amer. Stat. Assoc.*, vol. 94, no. 445, pp. 154–163, 1999.

[17] J. K. Lindsey, "Comparison of probability distributions," *J. R. Stat. Soc. B (Methodol.)*, vol. 36, no. 1, pp. 38–47, 1974.

[18] Z. Gao and T. Hastie, "LINCDE: Conditional density estimation via lindsey's method," *J. Mach. Learn. Res.*, vol. 23, no. 52, pp. 1–55, 2022.

[19] *IEEE Standard for Information Technology–Telecommunications and Information Exchange Between Systems–Local and Metropolitan Area Networks–Specific Requirements–Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Std. 802.11-2020, 2021, revision of IEEE Std 802.11-2016.

[20] *IEEE Standard for Low-Rate Wireless Networks—Amendment 1: Enhanced Ultra Wideband (UWB) Physical Layers (PHYs) and Associated Ranging Techniques*, IEEE Std. 802.15.4z-2020, 2020, amendment to IEEE Std 802.15.4-2020.

[21] "Interface control document—navstar GPS space segment/navigation user interfaces (icd-GPS-200)," U.S. Department of Defense, Technical Report ICD-GPS-200, Aug. 2022, describes GPS L1 C/A code properties including 1023-chip, 1 ms period, 1.023 Mcps.

[22] *3GPP TS 38.211: NR; Physical Channels and Modulation*, 3GPP Std. TS 38.211, 2024.

[23] R. W. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge University Press, 2018.

[24] T. M. Cover and J. A. Thomas, *Entropy, Relative Entropy, and Mutual Information*. John Wiley & Sons, Ltd, 2005, ch. 2, pp. 13–55. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch2

[25] J. Hoydis, S. Cammerer, F. Ait Aoudia, M. Nimier-David, L. Maggi, G. Marcus, A. Vem, and A. Keller. (2022) Sionna. Ver. 1.1.0, [Online]. Available: https://nvlabs.github.io/sionna/.

[26] S. Kang, M. Mezzavilla, S. Rangan, A. Madanayake, S. B. Venkatakrishnan, G. Hellbourg, M. Ghosh, H. Rahmani, and A. Dhananjay, "Cellular wireless networks in the upper mid-band," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 2058–2075, 2024.

[27] T. Raviv, S. Kang, M. Mezzavilla, S. Rangan, and N. Shlezinger, "Multi-frequency upper mid-band localization," in *Proc. IEEE 25th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*. IEEE, 2024, pp. 736–740.

[28] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (release 18)," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.901 V18.0.0, May 2024, release 18. [Online]. Available: https://www.3gpp.org/dynareport/38901.htm