

# Extreme Value Theory-based Predictive Interference Management for 6G Subnetworks with Transformer

Pramesh Gautam\*, Carsten Bockelmann\*, Armin Dekorsy\*

\*Department of Communications Engineering, University of Bremen, Germany,

Email: {gautam, bockelmann, dekorsy}@ant.uni-bremen.de

**Abstract**—In-X subnetworks (SNs) encounter significant challenges in achieving hyper-reliable low-latency communication (HRLLC), particularly in hyper-dense deployment scenarios. These challenges stem from rapid and dynamic variations in interference caused by mobility, dynamic channel statistics, and varying traffic patterns. Predictive interference management plays a crucial role in meeting these extreme requirements, by enabling proactive and reliable resource allocation to prevent performance degradation. To address these challenges, we propose a probabilistic interference prediction technique using an inverted quantile transformer (iQTransformer) to learn interference dynamics effectively and capture tail statistics. It predicts interference for multiple sensor-actuator (SA) pairs in the SN with minimal disparity by leveraging their own interference dynamics alongside those of other active SA pairs. We integrate extreme value theory (EVT) with the iQTransformer to handle rare and extreme interference events. Results demonstrate that our proposed method outperforms baseline approaches in terms of prediction accuracy, achieving near-optimal target reliability.

**Index Terms**—6G, Interference Prediction, 3GPP, Interference Management, Subnetworks, HRLLC.

## I. INTRODUCTION

6G is an emerging paradigm that expands the ability to connect a broader range of use cases across various verticals, enabling more intelligent services, enhanced control, and new user experiences [1]. By introducing hyper-reliable, low-latency communication (HRLLC) as a sixth pillar envisioned in [2], 6G aims to drive technical breakthroughs and meet the demands of use cases with extreme requirements, such as 5-7 nines reliability and 0.1-1 ms latency. One prominent category requiring HRLLC is the In-X sub-networks (SNs) use-case family [3], which imposes similarly strict conditions to support short-range communication. Examples of In-X SN such as a typical scenario involving multiple robots operating in a factory hall, where coordinated motion control, precise positioning, and proximity management are designed to optimize production line efficiency, effectively replacing traditional wired network infrastructures. Additional discussions on industrial SNs, along with variations like In-Vehicle and In-House SNs, can be found in [3]. This work mainly focuses on In-Factory/industrial SN.

Industrial SNs with closed-loop control communication demand HRLLC for intra-SN communication. High band-

width is needed within the sub-band for intra-SN communication to meet these extreme requirements [3]. However, the combination of limited channel bandwidth and hyper-dense deployment of SNs leads to significant inter-SN interference, commonly called co-subband interference. Interference management poses a major challenge in SN deployment due to dynamic interference, which is characterized by wireless channel fading, mobility of SNs, and varying traffic patterns, such as the transmission of packets with finite blocklength. To meet HRLLC demands, a proactive prediction is necessary, enabling SNs to allocate resources in advance and prevent performance degradation. We employ a predictive interference management framework [4] that leverages explicit predicted interference knowledge using a probabilistic interference prediction incorporating uncertainties and heavy-tail statistics.

Interference prediction has been extensively explored in existing literature. Schmidt et al. [5] proposed an ARIMA-based Kalman filter for interference prediction, which incorporates channel dynamics and user equipment (UE) mobility. [6] proposed a kernel-based probability density estimation for interference prediction. Furthermore, [4] exploited the tail of the distribution to predict interference using a discrete state-space, discrete-time Markov chain. Their method leverages the interference probability density function (pdf) characterized by the minimum and maximum Interference-to-noise-ratio (INR) with a Gaussian distribution, while interference is often more accurately represented by a heavy-tailed distribution. [7] proposed a nonlinear autoregressive neural network method for interference prediction. Compared to [4], authors found that ML-based approaches offer more resource-efficient predictions. Following this, [8] and [9] used empirical vector decomposition to break down interference data into intrinsic mode functions (IMF) and residual components. This method requires a separate ML model for each decomposed series, which can lead to computationally complex architectures for In-X SNs. In our previous work, we proposed a long short-term memory (LSTM) with/without federated learning-based interference prediction technique with noise-free estimation assumptions with deterministic traffic in [10], and [11] for In-X SN architectures. Meanwhile, interference prediction that accounts for uncertainties, particularly by incorporating tail statistics, holds great significance due to the presence of random traffic, deep fading, and noisy observations resulting from measurement errors. The importance of moving beyond

This work is supported by the German Ministry of Education and Research (BMBF) under the grants of 16KISK109 (6G-ANNA) and 16KISK016 (Open6GHub).

average or central-limit theorems with the inclusion of extreme and rare events has been highlighted for Ultrareliable and low-latency wireless communication (URLLC) in [12]. To the best of our knowledge, an Extreme Value Theory (EVT)-based approach with quantile transformer for interference prediction in SN or HURLLC setting, considering multiple sensor-actuator (SA) pair has not been explored in literature before. The key contributions of this research are outlined as follows:

**Contribution 1:** We propose a probabilistic interference prediction method using the *iQTransformer*, which integrates transformer-based models with quantile regression to capture interference dynamics incorporating uncertainties. This approach captures its own interference dynamics while leveraging attention mechanisms to capture cross-correlations between SA pairs.

**Contribution 2:** We present *EVT*-based technique, using the Peak-over-threshold (POT) mechanism to model tail statistics efficiently. The *iQTransformer* estimates probabilistic thresholds for *EVT*, enabling the model to capture and address extreme interference events.

**Contribution 3:** The effectiveness of our methods evaluated with third generation partnership project (3GPP) channel model demonstrating the accuracy and robustness of interference prediction in ensuring target reliability.

## II. SYSTEM MODEL

Consider a set of hyper-densely deployed  $N$  number of SNs denoted as  $\mathcal{N} = \{n_1, \dots, n_N\}$  uniformly distributed in a factory area  $A \subset \mathbb{R}^2$ . Each SN contains  $M$  number of collocated SA pairs,  $\mathcal{M}_{n_q} = \{m_1, \dots, m_M\}$ , for all  $n_q \in \mathcal{N}$  distributed as a binomial point process over a disc with radius  $r$  with a uniform probability density function (pdf) of  $\frac{1}{\pi r^2}$ . The SNs, together with their SA pairs, move at a velocity of  $v$  m/s, following a specified mobility model. We consider SN controller of SN communicates with respective SA pairs.

The total available bandwidth  $B$  is partitioned into  $K$  sub-bands, where  $K \ll N$ , and a SN is assumed to operate in time division duplexing (TDD) mode. Furthermore, hyper-dense deployment of SN with limited subband results in an inter-SN interference due to the allocation of the same sub-band for multiple SN within a confined area. We assume that the time required for data transmission from a sensor to the SN controller (uplink) and from the SN controller to an actuator (downlink) is symmetric. The radio frame is divided into  $2 \cdot N_{\text{sl}}$  equal slots indexed as  $\boldsymbol{\varrho} = [1, \dots, 2 \cdot N_{\text{sl}}]$ . Each slot's duration is considered to be equal to that of an Orthogonal Frequency Division Multiplexing (OFDM) symbol. Among these slots,  $\boldsymbol{\varrho}_{\text{UL}} = [1, \dots, N_{\text{sl}}]$  are allocated for uplink, while the remaining  $\boldsymbol{\varrho}_{\text{DL}} = [N_{\text{sl}} + 1, \dots, 2 \cdot N_{\text{sl}}]$  are allocated for downlink. As the SNs are non-synchronized, any SA pair  $m_a \in \mathcal{M}_{n_q}$  can be active in time slots  $\boldsymbol{\varrho}$  within the SN. We consider that the SN controller schedules  $\mathcal{M}_{n_q} = [m_{1n_q}, \dots, m_{N_{\text{sl}}n_q}]$  SA pairs at time slot  $\boldsymbol{\varrho}_{\text{UL}}$  in a near-orthogonal fashion. Based on this assumption, we consider intra-SN interference to be negligible or minimal compared to inter-SN interference [13]. One specific SA pair  $m_{\tau n_q}$  of the  $n_q$ th SN interferes with each

SA pair among the set  $\mathcal{B}$  that uses the same sub-band, where  $\mathcal{B} \subset \mathcal{N}$  is scheduled in the  $\tau$ th slot. The set of interfering SA pairs for the  $n_q$ th SN is defined as  $\mathcal{C}_{m_{\tau n_q}} = \bigcup_{n_j \in \mathcal{B} \setminus \{n_q\}} m_{\tau n_j}$  during the  $\tau$ th time slot. We consider a spatially consistent 3GPP channel model [14] with unit transmit power, the observed interference power value can be expressed as:

$$I_{m_{\tau n_q}}[k] := \sum_{c \in \mathcal{C}_{m_{\tau n_q}}} \left( \beta \cdot |h_{\text{LOS},c}[k]|^2 \cdot l_{\text{LOS},c}[k] \cdot \zeta_{\text{LOS},c}[k] + \sqrt{(1 - \beta^2)} \cdot |h_{\text{NLOS},c}[k]|^2 \cdot l_{\text{NLOS},c}[k] \cdot \zeta_{\text{NLOS},c}[k] \right) \cdot \chi_c[k], \quad (1)$$

where  $h_{\text{LOS},c}$  is the small-scale fading coefficient, correlated across transmission time intervals (TTIs) and following a Rician pdf [15];  $l_{\text{LOS},c}$  is the path loss component [14]; and  $\zeta_{\text{LOS},c}$  represents the spatially correlated shadowing effect, which follows a log-normal pdf [16]. Similarly, for non-line-of-sight (NLOS) interference, we consider small-scale fading with a Rayleigh distribution, denoted as  $h_{\text{NLOS},c}$ , alongside path loss  $l_{\text{NLOS},c}$  and shadowing  $\zeta_{\text{NLOS},c}$ , with corresponding NLOS parameters. The parameter  $\beta \in (0, 1)$  is a smoothing factor that prevents abrupt fluctuations in the channel response caused by transitions between line-of-sight (LOS) and NLOS conditions [14]. Additionally, the traffic is denoted as  $\chi_c[k]$ . We consider two extreme traffic scenarios: full buffer traffic for a specific slot, where  $\chi_c[k] := 1$ , and random traffic, defined by a Bernoulli random variable with probability in the range  $\chi_c[k] \sim \text{Bern}(0, 1)$  for specified slot. In practice, the observed interference power estimate, which follows heavy tailed distribution due to a finite number of interferers and log normally distributed shadowing, is imperfect due to various noises such as estimation error or process noise. The observed interference power estimate can be expressed as  $\tilde{I}_{m_{\tau n_q}}[k] := I_{m_{\tau n_q}}[k] + \varsigma[k]$ , where  $\varsigma[k]$  is the noise modeled as  $\varsigma \sim \mathcal{N}(0, \rho^2)$ .

## III. RESOURCE ALLOCATION PROBLEM

We consider Resource Allocation (RA) problem for uplink communication between the  $m_{\tau}$ -th SA and the SN controller for the  $n_q^{\text{th}}$  SN at a given TTI to determine the optimal number of required resource elements,  $R_{m_{\tau n_q}}[k]$ , by incorporating the target block error rate (BLER) as a reliability constraint. According to finite block length theory [17], the total number of channel usage that can be transmitted with a decoding error probability of  $\varepsilon_{\text{target}}$  for transmission of  $D$  information bits in an additive white Gaussian noise (AWGN) channel with signal to interference plus noise ratio (SINR)  $\hat{\gamma}_{m_{\tau n_q}}[k]$  is given [17]

$$R_{m_{\tau n_q}}[k] \approx \frac{D}{C(\hat{\gamma}_{m_{\tau n_q}}[k])} + \frac{Q^{-1}(\varepsilon_{\text{target}})^2 V(\hat{\gamma}_{m_{\tau n_q}}[k])}{2C(\hat{\gamma}_{m_{\tau n_q}}[k])^2} \cdot \left[ 1 + \sqrt{1 + \frac{4DC(\hat{\gamma}_{m_{\tau n_q}}[k])}{Q^{-1}(\varepsilon_{\text{target}})^2 V(\hat{\gamma}_{m_{\tau n_q}}[k])}} \right], \quad (2)$$

where  $C(\hat{\gamma}_{m_{\tau n_q}}) = \log_2(1 + \hat{\gamma}_{m_{\tau n_q}}[k])$  represents the Shannon capacity of AWGN channels under the infinite blocklength regime,  $Q^{-1}$  is the inverse Q-function, and  $V$  is channel

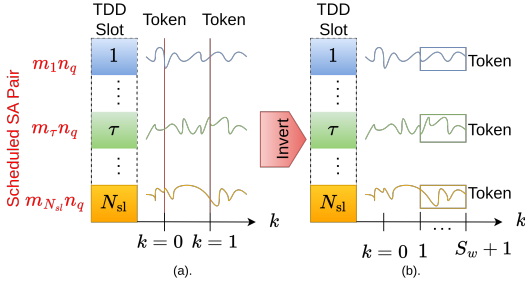


Fig. 1: Proposed inverted token representation of interference power values in (b) for  $N_{sl} := M$  SA pairs for training.

dispersion. The selection of number of resource elements  $R_{m_\tau n_q}[k]$  depends solely on  $\hat{\mathbf{I}}_{n_q}[k]$ , i.e., interference dynamics, as the variation in signal power is insignificant over the TTI due to the short-range deployment and the fixed positions of the SN controller and the SA pair. In order to proactively select  $R_{m_\tau n_q}[k]$ , we predict interference for the  $n_q$ -th SN beginning of  $k$ -th TTI for multi-SA pair as follows:

$$\hat{\mathbf{I}}_{n_q}[k] = \mathbf{G}(f_\alpha([\tilde{\mathbf{I}}_{n_q}[k-1], \dots, \tilde{\mathbf{I}}_{n_q}[k-S_w+1]], \boldsymbol{\theta})), \quad (3)$$

where  $f_\alpha(\cdot)$  is a probabilistic function modeling the non-linear interference dynamics by leveraging multiple SA pairs within the same SN, using  $S_w$  historical noisy interference estimates  $[\tilde{\mathbf{I}}_{n_q}[k-1], \dots, \tilde{\mathbf{I}}_{n_q}[k-S_w+1]] \in \mathbb{R}^{S_w \times M}$  as input to predict  $\hat{\mathbf{I}}_{n_q}[k] := [\hat{I}_{m_1 n_q}[k], \dots, \hat{I}_{m_{N_{sl}} n_q}[k]]^T \in \mathbb{R}^{1 \times M}$ . The symbols  $\alpha$  and  $\boldsymbol{\theta}$  correspond to the quantile and the model parameters, respectively. The symbol  $\mathbf{G}(\cdot) := G_{m_1 n_q}(\tilde{I}_{m_1 n_q}[k]), \dots, G_{m_{N_{sl}} n_q}(\tilde{I}_{m_{N_{sl}} n_q}[k])$  is independent tail p.d.f.s with individual  $G_{m_a n_q}(\cdot)$  p.d.f. for each SA pair using predicted threshold  $\tilde{\mathbf{I}}_{n_q}[k]$  with  $f_\alpha(\cdot)$  function.

#### IV. PROPOSED INTERFERENCE PREDICTOR

We have proposed a hybrid approach that combines the strengths of the Transformer network architecture [18] with a well-established statistical method, Extreme Value Theory (EVT), to model and predict interference. Interference exhibits a heavy-tailed distribution, resulting from the summation of interference from a finite number of interferers, influenced by a mixture of distributions—such as Rice or Rayleigh distributions, log-normally distributed shadowing, traffic variations, and rare events due to fading or traffic misalignment. We utilize EVT to assess the stochastic properties of interference. EVT-based approaches typically involve deriving block maxima or extracting peak values above (or below) a certain threshold from continuous interference estimates [12], [19]. EVT is applied to model the distribution of extreme events, using the Generalized Pareto Distribution (GPD) or Generalized Extreme Value (GEV) distribution to capture the tail statistics arising from rare, stochastic interference behaviors. Further details on EVT can be found in [12], [19].

##### A. Extreme Value Theory for interference prediction

We utilize the Peak-over-Threshold (POT) mechanism, which models exceedances of interference over a dynamically

defined threshold. To achieve this, we apply EVT through the POT method, where exceedances over the predicted  $\alpha$ th quantile threshold  $\tilde{I}_{m_\tau n_q}[k]$  are modeled using the *Generalized Pareto Distribution*, characterized by the shape parameter  $\xi$  and scale parameter  $\sigma$ , with the cumulative distribution function as follows [19]:

$$G_{m_\tau n_q}(\tilde{I}_{m_\tau n_q}[k]) = 1 - \left(1 + \xi \frac{\tilde{I}_{m_\tau n_q}[k] - \check{I}_{m_\tau n_q}[k]}{\sigma}\right)^{-\frac{1}{\xi}}, \quad (4)$$

where  $\sigma > 0$ , exceedances  $\tilde{I}_{m_\tau n_q}[k] > \check{I}_{m_\tau n_q}[k]$ , and  $1 + \xi \frac{\tilde{I}_{m_\tau n_q}[k] - \check{I}_{m_\tau n_q}[k]}{\sigma} > 0$ . We estimate  $\xi$  and  $\sigma$  using log-likelihood distribution fitting. The parameter  $\xi$  controls the tail heaviness, while  $\sigma$  governs the scale of the exceedances for each SA pair. A key challenge in EVT-based approaches is determining the probabilistic threshold ( $\alpha$ th quantile)  $\check{I}_{m_\tau n_q}[k]$ , which can vary over time accounting the *non-stationary nature of the SN* and the corresponding interference power values (IPVs) across successive TTIs for individual SA pairs.

##### B. Probabilistic threshold prediction using iQTransformer

The Transformer network can be utilized in prediction to model global dependencies over temporal tokens, with each token capturing slot-level interference estimates of the SA pair over the same TTI, as shown in Fig. 2 (a), see [20] for further details. However, distinct interference dynamics resulting from the spatial positioning of SA pairs often lead to user-centric statistics with their respective values [11]. This can result in a meaningless attention map, producing unreliable predictions that may be unsuitable for HRLLC use cases. Inspired by the inverted Transformer (iTransformer) proposed by [20], we propose an inverted Transformer-like architecture that is more suitable for multi-SA pair scenarios for multivariate probabilistic inference due to its ability to enrich learning exploiting cross-correlation as well. Since our focus is on learning probabilistic thresholds for each TTI in a multi-SA pair setting, we introduce an inverted quantile transformer (iQTransformer) to predict dynamic probabilistic thresholds that adapt to the variability of interference across TTIs, while retaining potential uncertainty. In this framework, we replace the feedforward network in the Transformer with an LSTM to better capture temporal dependencies in interference dynamics from historical data [11]. We use the name iQTransformer to reflect its core transformer-based architecture. This method uses successive IPVs of each SA pair as tokens, as shown in Fig. 2 (b). The attention mechanism captures cross-correlations between SA pairs, while the LSTM captures the non-linear dynamics of each SA pair within the same SN over time. The proposed iQTransformer includes several key components—an embedding layer, a multivariate attention layer, layer normalization, and an LSTM layer.

The embedding layer transforms the raw interference data  $\tilde{I}_{m_\tau n_q}[k-1]$  into a lower-dimensional latent space, which captures essential features of the interference while reducing dimensionality. This transformation is represented as

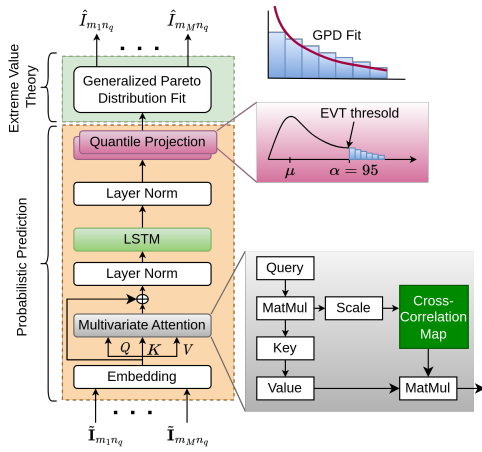


Fig. 2: Proposed multi-SA pair interference predictor with EVT-based iQTransformer incorporating tail statistics.

$\mathbf{h}_{m_\tau n_q} = \text{Embedding}(\tilde{\mathbf{I}}_{m_\tau n_q}[k-1])$ , which captures key temporal features with  $M$  embedded tokens denoted as  $\mathbf{H} = \{\mathbf{h}_{m_\tau n_q}, \dots, \mathbf{h}_{m_M n_q}\}$ . Following the embedding, the multivariate attention layer captures dependencies between different SA pair's interference. The attention mechanism [18] computes attention scores  $\alpha_{ij}$  for each SA pair based on the query  $\mathbf{Q}$ , key  $\mathbf{K}$ , and value  $\mathbf{V}$  matrices with extracted information  $\mathbf{H}$  as  $\alpha_{ij} = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_j^\top}{\sqrt{d_k}}\right)$ , where  $d_k$  is the dimensionality of the key vectors. The output of the attention layer is then given as  $\mathbf{A}_{\text{attn}} = \sum_j \alpha_{ij} \mathbf{V}_j$ . This mechanism enables the model to focus on relevant contributions of different SA pairs, ensuring an accurate representation across multi-SA pairs with enhanced interpretability revealing multivariate correlations (can be a weak or strong condition). To stabilize the training process and improve model efficiency, *layer normalization* is applied. Layer normalization adjusts the activations in each layer by normalizing the mean and variance with  $\text{LN}(\mathbf{H}) = \left\{ \frac{\mathbf{h}_{m_\tau n_q} - \mu_{\mathbf{h}_{m_\tau n_q}}}{\sigma_{\mathbf{h}_{m_\tau n_q}}} \mid \tau = 1, \dots, N_{\text{sl}} \right\}$ , where  $\mu_{\mathbf{h}_{m_\tau n_q}}$  and  $\sigma_{\mathbf{h}_{m_\tau n_q}}$  represent the mean and standard deviation of the activations in the current layer. This ensures that each layer operates on inputs with a consistent scale, making the model compact and robust during training. The *LSTM layer* plays a key role in capturing long-term dependencies in the interference data [11]. It employs input, forget, and output gates to retain relevant information from previous TTIs and discard irrelevant data. The LSTM's hidden state at time  $k$  is updated using the gated mechanism  $\Psi_k = o_k \odot \tanh(c_k)$ , where  $o_k$  is the output gate,  $c_k$  is the cell state, and  $\odot$  denotes element-wise multiplication. This enables the LSTM to learn such states to capture short-term fluctuations and long-term patterns in interference, improving prediction accuracy.

Interference power values often exhibit uncertainty due to channel statistics, traffic patterns, mobility, and impairments like TDD misalignment. Accurately quantifying and accounting for such uncertainty is crucial for reliable communication. Therefore, we aim to predict interference at an  $\alpha$  quantile such

as the 95th quantile, beyond the mean for reliable communication [12]. Thus, *quantile projection* is performed by utilizing the *pinball loss function* to estimate dynamic probabilistic thresholds for interference prediction without assuming the distribution of the target IPV. The pinball loss focuses on predicting a specific quantile  $\alpha$  to train the transformer model, which can be defined as follows:

$$\mathcal{L}_{\alpha, i} = \begin{cases} (1 - \alpha)(\check{\mathbf{I}}_{n_q}[k] - \tilde{\mathbf{I}}_{n_q}[k]), & \check{\mathbf{I}}_{n_q}[k] \geq \tilde{\mathbf{I}}_{n_q}[k]; \\ \alpha(\tilde{\mathbf{I}}_{n_q}[k] - \check{\mathbf{I}}_{n_q}[k]), & \check{\mathbf{I}}_{n_q}[k] < \tilde{\mathbf{I}}_{n_q}[k], \end{cases} \quad (5)$$

In this equation,  $\alpha$  represents the quantile level,  $\check{\mathbf{I}}_{n_q}[k]$  is the predicted interference threshold at the quantile level  $\alpha$ , and  $\tilde{\mathbf{I}}_{n_q}[k]$  is the actual observed interference value. Following this, we fit the GPD using a threshold based on exceedances during the training phase, providing a robust method for handling exceedances and extreme events. The fitted GPD is then used to characterize the tail statistics of the predicted interference.

Overall, the EVT-based-iQTransformer enables the model to effectively capture the tail behavior of the interference distribution and predict interference, offering a comprehensive solution for both typical and sporadic interference scenarios in highly dynamic environments. The overall algorithm EVT-based-iQTransformer-In-X for Predictive Interference Management (PIM) is summarized in algorithm 1. This approach uses  $\beta$  quantile of the fitted tail p.d.f denoted as  $\mathbf{G}_\beta(\cdot)$  for resource allocation.

---

#### Algorithm 1 PIM with EVT-based-iQTransformer-In-X

---

- 1: **Input:** Target BLER  $\varepsilon$ , Signal Power  $S$ ,  $\mathbf{I}_{n_q}$
  - 2: **Input:** Model  $f(\cdot)$ , target quantile  $\alpha$
  - 3: **Output:** Channel Usage  $\mathbf{R}$  and achieved BLER  $\varepsilon$
  - 4: Train iQTransformer to learn  $f_\alpha(\cdot)$  until convergence
  - 5: Exceedances  $\psi$  from training data and fit GPD  $\mathbf{G}(\cdot)$
  - 6: **for**  $k = S_w + 1, \dots, \Omega_{test}$  **do**
  - 7:     **Interference Prediction Phase:**
  - 8:      $\check{\mathbf{I}}_{n_q}[k] = f_\alpha(\mathbf{I}_{n_q}[k - S_w - 1 : k - 1])$
  - 9:      $\hat{\mathbf{I}}_{n_q}[k] := \check{\mathbf{I}}_{n_q}[k] + \mathbf{G}_\beta(\psi)$
  - 10:    **Resource Allocation Phase:**
  - 11:     $\mathbf{R}_{n_q}[k] := [R_{m_1 n_q}[k], \dots, R_{m_{N_d} n_q}[k]]$  based on (2)
  - 12:    Calculate Achievable BLER [4]
  - 13:    Update  $\mathbf{R} \leftarrow \mathbf{R}_{n_q}[k]$  and  $\varepsilon \leftarrow \varepsilon_{n_q}$
  - 14: **end for**
- 

## V. NUMERICAL RESULTS

### A. Baseline Methods

The following baseline methods are considered for comparison in this work:

1) *Genie RA*: This baseline performs a number of resource element selections based on the target BLER under the assumption that a SA pair's SINR is perfectly predicted at the SN controller [4]. It is not a practical approach, but it has been used to compare optimal BLER.

2) *Moving-average (MA) Predictor*: The IPV at every  $k$  is obtained as a weighted-sum of IPV's estimated at TTIs  $k-1$  and  $k-2$ . This method has been used in link adaptation for traditional enhanced mobile broadband (eMBB) services [4].

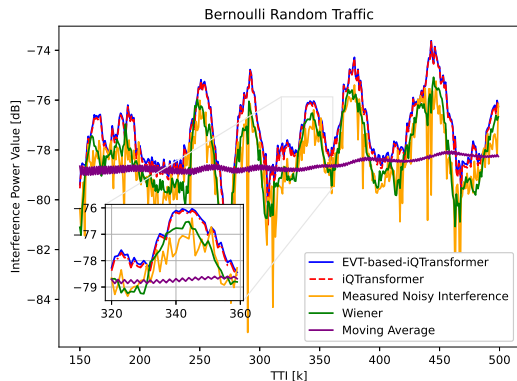


Fig. 3: Dynamics of observed and predicted interference.

3) *Wiener Predictor*: We have predicted interference using order  $S_w$  Wiener filter [11]. For one-step prediction, the resulting equation is identical to the Yule-Walker equations which are used to solve the autoregressive ( $S_w$ ) process [21]. This method utilizes second-order statistics of interference.

### B. Performance Metrics

1) *Coverage Probability*: Coverage probability, often defined as predictor reliability, is the probability that the predicted interference exceeds the observed interference as [22]:

$$\frac{\sum_{k=1}^{\Omega_{test}} M_i}{\Omega_{test}}, M_k = \begin{cases} 1 & \tilde{I}_{m_\tau n_q}[k] \leq \hat{I}_{m_\tau n_q}[k] \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

2) *Average Coverage Width*: : The average coverage width is given by [22]:

$$\frac{1}{\Omega_{test}} \sum_{k=1}^{\Omega_{test}} |\hat{I}_{m_\tau n_q}[k] - \tilde{I}_{m_\tau n_q}[k]|. \quad (7)$$

The ideal outcome is to minimize the coverage width while maintaining a high coverage probability to achieve the target BLER.

### C. Performance Analysis

We have assumed deployment density and mobility parameters together with detailed 3GPP channel parameters mentioned in table I. The channel parameters are considered based on the technical report produced by 3GPP [14]. In Fig. 3, it can be observed that the conventional moving approach is unable to track the local variability of interference for a specific SA pair, despite predicting the overall mean interference. On the other hand, the Wiener predictor can follow the non-linear dynamics by exploiting second-order statistics with a minimal coverage width for both isochronous and Bernoulli random traffic, though this compromises coverage probability, which directly impacts the achieved BLER, as shown in table II. Moreover, table II shows that the iQTransformer learns its own interference dynamics effectively while leveraging multivariate attention to improve prediction accuracy without affecting others' performance. The average normalized coverage width is higher for Bernoulli random traffic due to additional uncertainty. The proposed technique outperforms other models

TABLE I: Simulation Parameters

Parameter	Value
<b>Deployment Parameter</b>	
Number of subnetworks, $ \mathcal{N} $	16
Number of sensor-actuator pair, $ \mathcal{M} $	{4,8,12,16}
Interfering subnetworks, $ \mathcal{K} $	5
Deployment density (subnetwork/ $km^2$ )	40000
Mobility Model	RDMM
Cell Radius, $r$	2 [m]
Velocity, $v$	2 [m/s]
Minimum distance, $d$	3 [m]
<b>Channel Parameter</b>	
Carrier frequency	6 GHz
Number of sub-bands, $K$	4
Frequency reuse	1/4
Pathloss	3GPP InF-DL [23]
Shadow fading std. deviation	4 dB(LOS)   7.2 dB(NLOS)
Decorrelation distance,	10 m
Doppler frequency, $f_{d_n}$	80 Hz
Transmit Power	0 dBW
TTI duration	1 ms
Packet size	200 bits
<b>iQTransformer Parameters</b>	
Activation Function	tanh
Loss Function	Pinball Loss function
Learning Rate and Training epoch	0.001 and 300
Encoder Layer	2
Embedding hidden units and dropout	256 and 0.1
number of heads for attention, hidden units, and dropout	8, 256 and 0.1
Number of LSTM layers and hidden units per layer	1 and 256
Optimizer	Adaptive moments (Adam)
Weight Initialization	Xavier
Threshold quantile, $\alpha$	95th
EVT-quantile, $\beta$	50th
Train and test data ( $\Omega_{test}$ ) size	8000, 2000

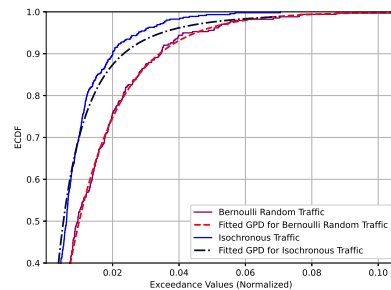


Fig. 4: GPD fitting utilizing threshold from iQTransformer.

even during the scaling of SA pairs, thanks to its ability to effectively capture the interference dynamics of other SA pairs, as highlighted in table II.

We utilize the tail statistics from the exceedances in the training data to characterize the tail distribution in the POT approach, as achieved through the iQTransformer, as shown in Fig. 4. Moreover, Bernoulli random traffic results in a heavier tail than isochronous traffic due to the stochastic variations introduced by Bernoulli randomness. By combining advancements in machine learning with the iQTransformer and leveraging well-established statistical methods for tail statistics using the POT method within the EVT framework, we achieve at least 94% coverage probability across both traffic conditions, albeit with a compromise in coverage width.

Our findings indicate that the proposed predictor outperforms baseline methods and achieves near-optimal RA performance, utilizing an additional 14% to 22% of channel resources as compared to Genie RA, as shown in Fig. 5. Such predictive interference management also provides an alternative approach instead of naive packet retransmission and blind repetition. Similarly, in the presence of random traffic, Fig. 5 highlights the robustness of the proposed EVT-based technique, enhanced by the hybrid ML and statistical approach, which effectively captures the non-linear dynamics of interference in the presence of uncertainties.

TABLE II: Coverage width and probability of baseline and proposed predictor for fixed/scaling SA pair

Fixed number of SA Pair	Isochronous Traffic								Bernoulli Random Traffic							
	Coverage Probability				Average Coverage Width				Coverage Probability				Average Coverage Width			
	SA 1	SA 2	SA 3	SA 4	SA 1	SA 2	SA 3	SA 4	SA 1	SA 2	SA 3	SA 4	SA 1	SA 2	SA 3	SA 4
Moving Average	0.561	0.580	0.597	0.460	0.095	0.088	0.081	0.107	0.565	0.483	0.546	0.550	0.103	0.870	0.127	0.0936
Wiener	0.461	0.487	0.503	0.441	<b>0.018</b>	<b>0.015</b>	<b>0.016</b>	<b>0.021</b>	0.551	0.735	0.718	0.793	<b>0.036</b>	<b>0.043</b>	<b>0.049</b>	<b>0.049</b>
iQTransformer	0.927	0.943	0.920	0.936	0.054	0.050	0.042	0.068	0.927	0.928	0.924	0.918	0.095	0.080	0.096	0.087
EVT-based iQTransformer	<b>0.956</b>	<b>0.964</b>	<b>0.948</b>	<b>0.955</b>	0.061	0.054	0.046	0.074	<b>0.954</b>	<b>0.952</b>	<b>0.952</b>	<b>0.946</b>	0.103	0.089	0.104	0.094
Scaling of SA Pairs	4	8	12	16	4	8	12	16	4	8	12	16	4	8	12	16
Moving Average	0.588	0.588	0.588	0.588	0.103	0.103	0.103	0.103	0.547	0.547	0.547	0.547	0.098	0.098	0.098	0.098
Wiener	0.457	0.457	0.457	0.457	<b>0.018</b>	<b>0.018</b>	<b>0.018</b>	<b>0.018</b>	0.512	0.512	0.512	0.512	<b>0.025</b>	<b>0.025</b>	<b>0.025</b>	<b>0.025</b>
iQTransformer	0.939	0.938	0.949	0.937	0.050	0.050	0.053	0.051	0.925	0.933	0.934	0.931	0.056	0.056	0.056	0.054
EVT-based iQTransformer	<b>0.965</b>	<b>0.966</b>	<b>0.973</b>	<b>0.969</b>	0.057	0.057	0.061	0.059	<b>0.951</b>	<b>0.963</b>	<b>0.966</b>	<b>0.963</b>	0.060	0.060	0.061	0.058

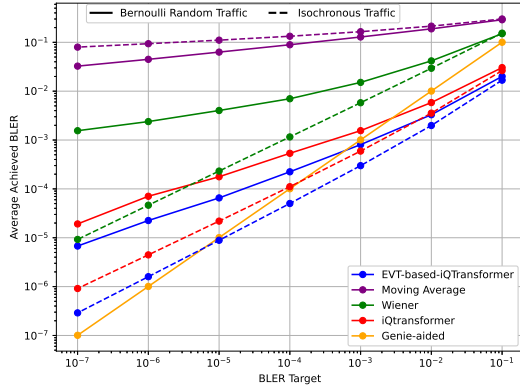


Fig. 5: Achieved BLER with different traffic model

## VI. CONCLUSIONS

In this research, we proposed an EVT-based iQTransformer for interference prediction—a hybrid probabilistic model that accounts for stochastic statistics related to extreme and rare events caused by traffic, noise, or heavy-tail distributions of interference. The proposed technique achieves HRLLC in deterministic traffic scenarios, delivering superior performance compared to baseline methods and approaching the performance of a Genie-aided predictor in sporadic traffic conditions. This work demonstrates that such a hybrid approach, combining advanced machine learning with statistical methods, can foster the design of proactive and robust predictors, which are crucial for predictive interference management frameworks. Moreover, adopting the proposed method in resource-constrained SN remains an open topic for future research.

## REFERENCES

- [1] M. Hoffmann, G. Kunzmann, T. Dudda *et al.*, “A Secure and Resilient 6G Architecture Vision of the German Flagship Project 6G-ANNA,” *IEEE Access*, vol. 11, pp. 102 643–102 660, 2023.
- [2] ITU-R, “Framework and overall objectives of the future development of IMT for 2030 and beyond,” *International Telecommunication Union (ITU) Recommendation (ITU-R)*, 2023.
- [3] G. Berardinelli, P. Baracca, R. O. Adeogun *et al.*, “Extreme Communication in 6G: Vision and Challenges for ‘in-X’ Subnetworks,” *IEEE Open Journal of the Communications Society*, vol. 2, pp. 2516–2535, 2021.
- [4] N. H. Mahmood, O. A. López, H. Alves, and M. Latva-Aho, “A Predictive Interference Management Algorithm for URLLC in Beyond 5G Networks,” *IEEE Communications Letters*, vol. 25, no. 3, pp. 995–999, 2021.
- [5] J. F. Schmidt, U. Schilcher, M. K. Atiq, and C. Bettstetter, “Interference Prediction in Wireless Networks: Stochastic Geometry Meets Recursive Filtering,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2783–2793, 2021.

- [6] A. Brighente, J. Mohammadi, and P. Baracca, “Interference distribution prediction for link adaptation in ultra-reliable low-latency communications,” in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–7.
- [7] C. Padilla, R. Hashemi, N. H. Mahmood, and M. Latva-Aho, “A nonlinear autoregressive neural network for interference prediction and resource allocation in urllc scenarios,” in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2021, pp. 184–189.
- [8] C. Jayawardhana, T. Sivalingham, N. H. Mahmood *et al.*, “Predictive resource allocation for URLLC using empirical mode decomposition,” in *2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2023, pp. 174–179.
- [9] S. Gunarathne, T. Sivalingham, N. H. Mahmood *et al.*, “Decomposition Based Interference Management Framework for Local 6G Networks,” in *2023 IEEE Globecom Workshops*. IEEE, 2023, pp. 1633–1638.
- [10] P. Gautam, M. Vakilifard, C. Bockelmann, and A. Dekorsy, “Cooperative Interference Estimation Using LSTM-Based Federated Learning for In-X Subnetworks,” in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 1338–1344.
- [11] P. Gautam, C. Bockelmann, and A. Dekorsy, “Interference Prediction in Unconnected In-X Mobile 6G Subnetworks Using a Data-Driven Approach,” in *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2024, pp. 2046–2052.
- [12] M. Bennis, M. Debbah, and H. V. Poor, “Ultrareliable and low-latency wireless communication: Tail, risk, and scale,” *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.
- [13] R. Adeogun, G. Berardinelli, P. E. Mogensen, I. Rodriguez, and M. Razzaghpour, “Towards 6G in-X Subnetworks With Sub-Millisecond Communication Cycles and Extreme Reliability,” *IEEE Access*, vol. 8, pp. 110 172–110 188, 2020.
- [14] 3GPP, “5G: Study on channel model for frequencies from 0.5 to 100 GHz (Release 16),” 3rd Generation Partnership Project, Tech. Rep. 38.901 v16.1.0, 2020.
- [15] C. Xiao, Y. R. Zheng, and N. C. Beaulieu, “Novel Sum-of-Sinusoids Simulation Models for Rayleigh and Rician Fading Channels,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3667–3679, 2006.
- [16] S. Lu, J. May, and R. J. Haines, “Effects of correlated shadowing modeling on performance evaluation of wireless sensor networks,” in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, 2015, pp. 1–5.
- [17] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [18] A. Vaswani *et al.*, “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017.
- [19] L. Haan and A. Ferreira, *Extreme value theory: an introduction*. Springer, 2006, vol. 3.
- [20] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, “itransformer: Inverted transformers are effective for time series forecasting,” *arXiv preprint arXiv:2310.06625*, 2023.
- [21] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
- [22] V. Jensen, F. M. Bianchi, and S. N. Anfinson, “Ensemble conformalized quantile regression for probabilistic time series forecasting,” *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [23] 3GPP, “Study on Communication for Automation in Vertical Domains (Release 16),” 3rd Generation Partnership Project, Tech. Rep. 22.804 v1.2.0, 2018.