Model-Consistency-Based PRACH Peak Validation Under Large Carrier Frequency Offsets

Hamidreza Khaleghi * and Thierry Lucidarme

Posted Date: 4 May 2026

doi: 10.20944/preprints202605.0050.v1

Keywords: PRACH; carrier frequency offset (CFO); peak disambiguation; GLRT; residual-based detection; Zadoff-Chu sequences; 3GPP channels; non-terrestrial networks (NTN)

 doi:10.20944/preprints202605.0050.v1

Article

Model-Consistency-Based PRACH Peak Validation Under Large Carrier Frequency Offsets

Hamidreza Khaleghi and Thierry Lucidarme *

IRT b-com, 35510 Cesson-Sévigné, France
* Correspondence: hamidreza.khaleghi@b-com.com

Abstract

Large carrier frequency offsets (CFOs) can severely distort the correlation response of the Physical Random Access Channel (PRACH), generating multiple significant peaks even for a single transmitting user equipment (UE), such that CFO-induced pseudo-peaks may exceed the detection threshold and be erroneously identified as valid peaks. This work addresses the problem of peak disambiguation under such conditions by formulating peak selection as a model-consistency validation problem under mismatch. A generalized likelihood ratio test (GLRT) is first formulated to provide a principled statistical validation of each detected candidate peak based on the estimated timing advance (TA) and CFO parameters. While theoretically grounded, this approach is shown to be insufficient under realistic large-CFO conditions, where CFO-induced peak ambiguity is further complicated by multipath-induced model mismatch. To address this limitation, a complementary residual-energy-based criterion is introduced, along with a weighted combination of both metrics, interpreted as a penalized consistency criterion for robust peak selection under model mismatch. The proposed framework enables the selection of a single reliable TA/CFO pair among multiple candidates, improving receiver robustness and reducing spurious updates. Performance is evaluated using precision, recall, and F1-score for both short and long PRACH formats under 3GPP-aligned channel models, including high-CFO and high-Doppler scenarios. Results demonstrate that the proposed weighted strategy generally provides a more robust trade-off than the individual GLRT-only and residual-only criteria.

Keywords: PRACH; carrier frequency offset (CFO); peak disambiguation; GLRT; residual-based detection; Zadoff-Chu sequences; 3GPP channels; non-terrestrial networks (NTN)

1. Introduction

In cellular systems, the Physical Random Access Channel (PRACH) enables uplink time synchronization by estimating the propagation delay of a user equipment (UE) from the location of the detected correlation peak in the PRACH receiver output [1]. PRACH preambles are constructed from Zadoff–Chu (ZC) sequences, which exhibit constant-amplitude zero-autocorrelation (CAZAC) properties and are well suited for timing estimation [2]. After correlation with the Zadoff–Chu sequence, the receiver forms a power delay profile (PDP) whose samples correspond to delay bins representing candidate timing hypotheses. Under ideal conditions, the strongest PDP peak coincides with the true propagation delay of the PRACH preamble.

In practice, carrier frequency offset (CFO) alters the autocorrelation of ZC sequences and distorts the resulting PDP, introducing a coupling between timing and frequency offset such that the PDP peak may no longer correspond to the true propagation delay [3]. This effect becomes particularly pronounced under large Doppler shifts, residual frequency offsets, or high-mobility conditions, as encountered in non-terrestrial networks (NTN) [4]. In such regimes, multiple significant peaks may appear in the correlation output, leading to ambiguity between true and CFO-induced pseudo-peaks, which can be erroneously identified as valid peaks.

 doi:10.20944/preprints202605.0050.v1

Although the standard provides configuration options intended to improve PRACH robustness under high-mobility or high-Doppler conditions (e.g., restricted preamble sets), these mechanisms do not prevent the appearance of multiple correlation peaks under large residual CFO, and a receiver-side peak validation stage remains necessary.

Several works have investigated PRACH detection and timing estimation robustness in the presence of frequency mismatch and channel impairments. Modified correlation metrics and likelihood-based detection schemes have been proposed to improve detection reliability and mitigate CFO-induced degradation under multipath fading [5,6]. In addition, multiuser and contention-aware random access detection under CFO has been studied, highlighting the emergence of multiple competing timing hypotheses in the presence of frequency offsets [7]. These approaches primarily focus on enhancing detection performance or reducing timing estimation error, and generally assume that a dominant correlation peak can be reliably identified once detection is achieved. However, the problem of validating and selecting the correct peak among multiple candidates generated by a single transmission remains largely unaddressed.

Beyond classical signal processing techniques, learning-based approaches have recently been explored to address severe CFO conditions, particularly in high-mobility and NTN scenarios. Neural-network and clustering-based methods have been proposed for CFO estimation using random access preambles, demonstrating improved estimation accuracy under large Doppler uncertainty [8]. While effective for parameter estimation, these approaches do not explicitly address the ambiguity arising from multiple correlation peaks.

Unlike prior approaches that primarily focus on detection or parameter estimation accuracy, this work addresses the complementary problem of *peak validation under model mismatch*. In particular, peak selection is interpreted as a consistency test between the observed local correlation structure and a reduced-form analytical model, where multipath propagation and other channel effects introduce deviations not captured by the nominal model.

As illustrated in Figure 1, a single PRACH transmission may produce multiple candidate peaks under large CFO. In practical receivers, the selection of the correct timing hypothesis is often based on simple amplitude criteria, which become unreliable in such conditions. More sophisticated statistical validation methods can be considered; however, as shown later in this work, purely model-based criteria may themselves be insufficient under realistic channel conditions. A robust mechanism to validate and disambiguate detected PRACH peaks based on model-consistency considerations is therefore still lacking.

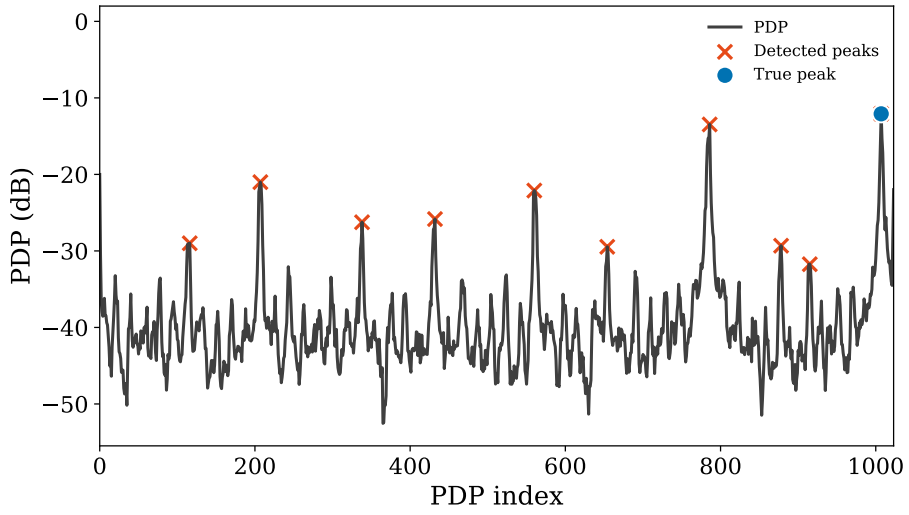


Figure 1. PRACH delay-domain power profile under a large carrier frequency offset (approximately half the subcarrier spacing). Multiple peaks exceed the detection threshold, although only one corresponds to the true propagation delay.

 doi:10.20944/preprints202605.0050.v1

1.1. Contribution and Positioning

This work addresses the problem of *PRACH peak disambiguation under large CFO*, focusing on reliable peak validation in scenarios where multiple candidates arise from a single transmission.

The contributions of this paper are fourfold. First, we formulate candidate-peak validation as a model-consistency problem under mismatch. Second, we introduce a reduced-dimensional feature representation tailored to the dominant three-lobe structure of the PRACH correlation response. Third, we propose a robust validation rule combining a GLRT-based score and a residual-based consistency score. Finally, a geometric manifold interpretation is provided to explain the complementary behavior of these criteria.

1.2. Notation

Scalars are denoted by italic letters (e.g., x), and vectors by bold lowercase letters (e.g., $\mathbf{x}$). The Euclidean norm is denoted by $\|\cdot\|$. The inner product between two complex vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as $\langle \mathbf{x} \mid \mathbf{y} \rangle = \mathbf{x}^H \mathbf{y}$. The notation $\mathbb{C}$ denotes the set of complex numbers. The identity matrix of size L is denoted by $\mathbf{I}_L$. The notation $(\cdot)_N$ denotes the modulo-N operation.

2. Receiver Flow and Candidate Validation Problem

We consider a standard PRACH receiver architecture consistent with current cellular systems and with our previous work on high-precision TA/CFO estimation [9]. Only the elements required for the peak-disambiguation framework are recalled here for completeness.

The considered PRACH receiver follows a standard correlation-based processing chain, summarized in Figure 2. After cyclic prefix removal, the received PRACH signal is processed in the frequency domain and transformed back to the delay domain via an inverse discrete Fourier transform, yielding the correlation output $c(l)$ for $l \in [0, N-1]$. A constant false-alarm rate (CFAR) detector is then applied to $|c(l)|^2$ to identify significant local maxima, which may result in multiple candidate delay-bin peaks $\{l_p^{(1)}, \ldots, l_p^{(P)}\}$ for a single PRACH transmission.

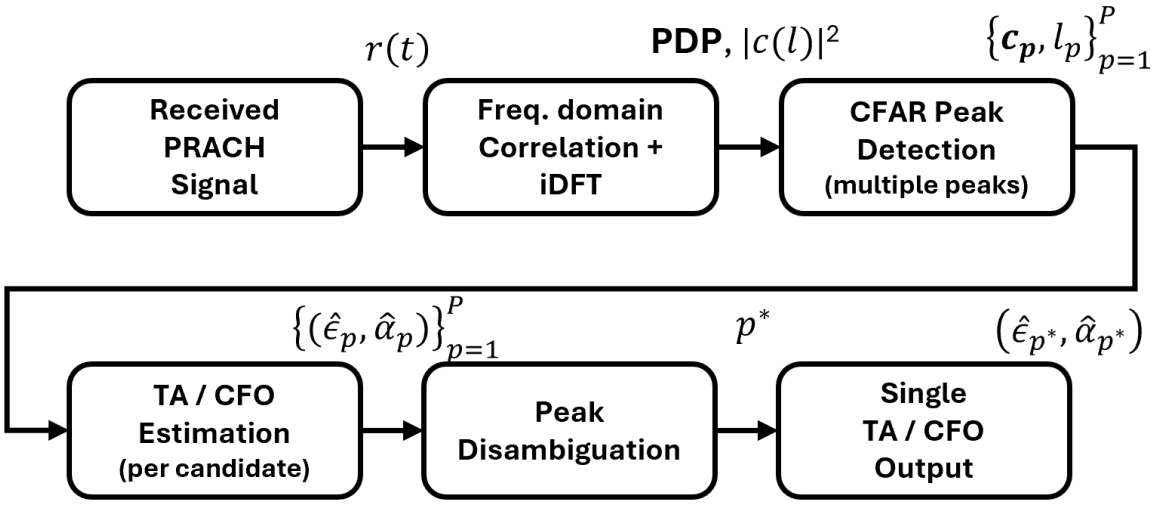


Figure 2. Conceptual PRACH receiver flow for correlation processing, candidate peak detection, TA/CFO estimation, and peak disambiguation.

For each detected peak candidate p, a reduced feature representation is constructed by collecting $L = 21$ complex samples around the detected peak and its two expected neighboring lobes. This three-lobe structure captures the dominant shape of the PRACH correlation response predicted by the analytical model while significantly reducing the observation dimensionality. Specifically, 7 samples are retained around each of the three dominant lobes, resulting in a compact yet informative feature vector $\mathbf{c}_p \in \mathbb{C}^{21}$. The choice of $L = 21$ results from a trade-off between representation accuracy and dimensionality reduction. Retaining 7 samples per dominant lobe captures the main structure of the

 doi:10.20944/preprints202605.0050.v1

correlation response while limiting sensitivity to noise and model mismatch. Increasing L provides only marginal improvement while increasing computational cost and sensitivity to estimation errors.

Using this feature vector, the normalized delay α (timing advance) and normalized CFO ϵ are estimated independently for each candidate peak according to the maximum-likelihood criterion introduced in the next section. Since several candidate peaks may be associated with a single PRACH occasion, a final validation stage is required to select the candidate whose local correlation structure is most consistent with the analytical PRACH model and to forward a single reliable timing advance and CFO estimate to higher protocol layers. The next section introduces the reduced-form analytical model used both for parameter estimation and for the subsequent candidate-validation criteria.

3. Analytical PRACH Model and Reduced-Form Representation

We adopt the PRACH analytical framework introduced in our prior work on high-precision TA and CFO estimation [9], and recall here only the key expressions required for the proposed peak-disambiguation method. After cyclic prefix removal, frequency-domain processing, and inverse discrete Fourier transform (iDFT), the analytical expression of the correlation output can be written as:

$$c(l) = A_{00} e^{j\frac{\pi}{N}(N-1)l} \sum_{d=-(N-1)}^{N-1} e^{-j\pi\varphi(d,l)} S_L(d+\epsilon)\, S_N(\alpha - C_v - \tilde{u}d - l), \quad l \in [0, N-1], \quad (1)$$

where A_{00} is a complex gain, u is the ZC root, $\tilde{u}$ its modular inverse, and C_v the cyclic shift. The functions $S_L(\cdot)$ and $S_N(\cdot)$ represent the deterministic shaping effects introduced by the PRACH signal structure: $S_L(\cdot)$ corresponds to the Dirichlet kernel arising from the finite PRACH subcarrier allocation, while $S_N(\cdot)$ denotes the periodic sinc function induced by the N-point iDFT used to obtain the delay-domain correlation output, as detailed in [9]. The phase term $\varphi(d,l)$ collects the phase contributions.

3.1. Dominant Peaks and Reduced Feature Set

The dominant energy of $c(l)$ is concentrated in the terms $d = 0, \pm 1$, leading to three principal correlation peaks located at

- the main peak $l_p = q(\alpha - C_v, \epsilon)$,
- the positive neighboring peak $l_p^+ = (\tilde{u} + l_p)_N$,
- the negative neighboring peak $l_p^- = (-\tilde{u} + l_p)_N$.

Here, $q(\cdot,\cdot)$ denotes the dominant delay-bin index as a function of the normalized delay and CFO. Around each of these three indices, 7 complex samples are retained, yielding a total of $L = 21$ samples. Denoting by S_0 the set of selected indices with $\mathrm{card}(S_0) = 21$, the corresponding noisy feature model is

$$c(l') = \hat{c}(l') + w(l'), \qquad l' \in S_0, \quad (2)$$

where $w(l')$ are independent and identically distributed complex Gaussian noise samples. Here, $c(l')$ denotes the observed correlation output, while $\hat{c}(l')$ denotes its noiseless analytical counterpart.

3.2. Reduced-Form Model and Estimation Criterion

Using the analytical model, the noiseless correlation samples at the selected indices admit the reduced-form approximation

$$\hat{c}(l') \approx A_{00}\, f(l', \epsilon, \alpha), \qquad l' \in S_0, \quad (3)$$

where $f(l', \epsilon, \alpha)$ is a deterministic function of the normalized CFO ϵ and normalized delay α induced by the three-dominant-peak structure.

 doi:10.20944/preprints202605.0050.v1

Stacking the samples $\{c(l_i')\}_{i=1}^{L}$ into a vector,

$$\begin{aligned} \mathbf{c} &= \left[c(l_1'), \ldots, c(l_L')\right]^T, \\ \mathbf{f}(\epsilon, \alpha) &= \left[f(l_1', \epsilon, \alpha), \ldots, f(l_L', \epsilon, \alpha)\right]^T, \end{aligned} \tag{4}$$

where $\mathbf{c}$ represents the observed feature samples and $\mathbf{f}(\epsilon, \alpha)$ denotes the corresponding deterministic model template, the reduced noisy model can be written compactly as

$$\mathbf{c} = A_{00}\, \mathbf{f}(\epsilon, \alpha) + \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_L). \tag{5}$$

The normalized maximum-likelihood (ML) criterion for estimating (ϵ, α) is given by

$$(\hat{\epsilon}, \hat{\alpha}) = \arg\min_{\epsilon, \alpha} \left(-\frac{|\langle \mathbf{c} \mid \mathbf{f}(\epsilon, \alpha) \rangle|^2}{\|\mathbf{f}(\epsilon, \alpha)\|^2} \right). \tag{6}$$

In [9], a compact neural-network-based estimator compatible with real-time implementation was trained using this criterion and shown to achieve high estimation accuracy; however, the peak-disambiguation framework developed next relies only on the resulting parameter estimates and is independent of the specific estimation method. This reduced-form model constitutes the basis of the statistical validation criteria introduced in the next section.

4. Proposed Peak Disambiguation Method

This section presents the statistical validation framework used to select a single physically consistent PRACH peak among multiple detected candidates.

4.1. Hypotheses for a Given Detected Peak

For each candidate $p \in \{1, \ldots, P\}$, a reduced feature set $S_0^{(p)}$ of $L = 21$ indices is formed around the detected peak and its two expected neighboring lobes (7 samples per lobe). The corresponding observation vector $\mathbf{c}_p \in \mathbb{C}^L$ is obtained by collecting the correlation samples $\{c(l_i')\}_{i=1}^{L}$ with $l_i' \in S_0^{(p)}$, where $c(l')$ denotes the observed correlation output.

Using the estimation criterion in (6), a pair of parameter estimates $(\hat{\epsilon}_p, \hat{\alpha}_p)$ is obtained for each candidate, from which the associated deterministic model vector $\mathbf{f}_p = \mathbf{f}(\hat{\epsilon}_p, \hat{\alpha}_p)$ is constructed, where $\mathbf{f}_p$ represents the model corresponding to the estimated parameters.

For each candidate peak, the following hypotheses are tested:

$$\begin{cases} H_0^{(p)} : \mathbf{c}_p = \mathbf{w}_p, \\ H_1^{(p)} : \mathbf{c}_p = A_{00}\mathbf{f}_p + \mathbf{w}_p, \end{cases}, \tag{7}$$

where $\mathbf{w}_p \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_L)$. Under $H_0^{(p)}$, the candidate peak is not consistent with the model template $\mathbf{f}_p$, and the projection of $\mathbf{c}_p$ onto $\mathbf{f}_p$ is attributable to noise-only fluctuations.

4.2. ML Estimation of the Complex Amplitude

Under $H_1^{(p)}$, with (ϵ, α) fixed to the estimated values $(\hat{\epsilon}_p, \hat{\alpha}_p)$, the only unknown parameter is the complex amplitude A_{00}. Assuming complex Gaussian noise, the log-likelihood function is given by

$$\ln p(\mathbf{c}_p \mid A_{00}, H_1^{(p)}) = -\frac{1}{\sigma^2} \left\| \mathbf{c}_p - A_{00}\mathbf{f}_p \right\|^2 + \text{const.} \tag{8}$$

 doi:10.20944/preprints202605.0050.v1

Maximizing (8) with respect to A_{00} is equivalent to minimizing the squared Euclidean distance $\|\mathbf{c}_p - A_{00}\mathbf{f}_p\|^2$, which yields the maximum-likelihood estimate

$$\hat{A}_{00,p} = \frac{\mathbf{f}_p^H \mathbf{c}_p}{\mathbf{f}_p^H \mathbf{f}_p} = \frac{\langle \mathbf{f}_p \mid \mathbf{c}_p \rangle}{\|\mathbf{f}_p\|^2}. \tag{9}$$

The corresponding reconstructed model vector is

$$\hat{\mathbf{c}}_p = \hat{A}_{00,p}\, \mathbf{f}_p. \tag{10}$$

4.3. Peak Disambiguation Criteria

The proposed validation framework relies on two complementary principles. The GLRT-based criterion evaluates the statistical consistency of each candidate with the analytical model through a projection-based likelihood test, while the residual-energy criterion quantifies the mismatch between the observed feature vector and the model prediction. Their combination enables robust validation in the presence of model imperfections and CFO-induced ambiguities, where the analytical model captures CFO-induced structure but does not explicitly account for multipath effects.

4.3.1. GLRT-Based Criterion

The generalized likelihood ratio for candidate p is defined as

$$\Lambda_p(\mathbf{c}_p) = \frac{p(\mathbf{c}_p \mid \hat{A}_{00,p}, H_1^{(p)})}{p(\mathbf{c}_p \mid H_0^{(p)})}. \tag{11}$$

This leads to the GLRT statistic

$$G_p = \frac{|\mathbf{f}_p^H \mathbf{c}_p|^2}{\sigma^2\, \mathbf{f}_p^H \mathbf{f}_p} = \frac{\left|\langle \mathbf{f}_p \mid \mathbf{c}_p \rangle\right|^2}{\sigma^2 \left\|\mathbf{f}_p\right\|^2}, \tag{12}$$

which measures the alignment between the observed vector and the model template. While the GLRT provides a statistically optimal test under perfect model matching, its effectiveness degrades under realistic channel conditions, particularly in the presence of multipath. In such regimes, the analytical model may not fully capture the observed correlation structure due to multipath propagation and other channel effects not explicitly included in the reduced-form model. As a result, pseudo-peaks may still yield large GLRT values, motivating the use of an additional consistency criterion.

4.3.2. Residual-Energy-Based Criterion

The residual energy associated with candidate p is

$$R_p = \left\|\mathbf{c}_p - \hat{A}_{00,p}\mathbf{f}_p\right\|^2. \tag{13}$$

For a true peak, $\mathbf{c}_p$ is well explained by the reduced-form model and R_p remains close to its noise-floor value, whereas mismatched candidates exhibit significantly larger residuals. A normalized residual score is defined as

$$S_{\text{res},p} = 1 - \frac{R_p}{\left\|\mathbf{c}_p\right\|^2}. \tag{14}$$

4.3.3. Weighted Combination Criterion

The combined score

$$S_p = \lambda \tilde{G}_p + (1-\lambda) S_{\text{res},p} \tag{15}$$

provides a unified selection rule that jointly exploits the GLRT statistic and the residual-based consistency measure.

 doi:10.20944/preprints202605.0050.v1

This criterion can be interpreted as a *penalized consistency measure*, where the GLRT term captures alignment with the nominal analytical model, while the residual term penalizes deviations induced by model mismatch, primarily due to multipath and unmodeled channel effects. From this perspective, the proposed formulation can be viewed as a practical surrogate for likelihood-based validation under a perturbed model, where deviations from the nominal structure are explicitly penalized. This formulation does not claim optimality, but provides a practical and robust compromise for peak validation under model mismatch. A geometric interpretation of this behavior is provided in Section 5.

4.4. Threshold Selection and Practical Parameterization

Under $H_0^{(p)}$, the observation vector satisfies $c_p = w_p$. The scalar projection $z_p = f_p^H c_p$ is therefore complex Gaussian with variance $\sigma^2 f_p^H f_p$. With the normalization used in G_p, it follows that G_p/σ^2 is exponentially distributed with unit mean. For a target false-alarm probability P_{FA}, the corresponding GLRT threshold is

$$G_{\mathrm{th}} = \sigma^2(-\ln P_{\mathrm{FA}}). \tag{16}$$

In contrast, the residual threshold and the weighting parameter λ do not admit closed-form optimal values under model mismatch. Instead of performing fine per-SNR tuning, we adopt a coarse two-regime parameterization.

Empirically, the performance of the proposed framework is relatively insensitive to small variations of λ (e.g., ± 0.1), while the residual threshold primarily governs false rejection at low SNR and ambiguity resolution at high SNR. In general, higher values of λ are preferable when the analytical model accurately represents the observed structure, whereas lower values improve robustness under multipath and model mismatch.

Based on these observations, a simple two-regime parameterization is adopted. In the low-SNR regime, a relaxed residual threshold and a smaller λ are used to avoid rejecting true peaks. In the high-SNR regime, a stricter residual threshold and a larger λ are employed to better resolve competing candidates.

This coarse parameterization avoids per-SNR tuning while maintaining stable performance across a wide range of operating conditions. In practice, the noise variance σ^2 can be estimated from noise-only regions of the correlation output, and the operating regime can be inferred from coarse SNR estimates or receiver-side quality indicators. The complete peak disambiguation procedure is summarized in Algorithm 1.

Algorithm 1 Weighted GLRT/Residual Peak Disambiguation

Input: P detected candidate peaks; target false-alarm probability P_{FA}; weight $\lambda \in [0,1]$; residual threshold $S_{\mathrm{res,th}}$
for $p = 1$ to P **do**
    Build $\mathbf{c}_p$ from the 21 samples in $S_0^{(p)}$.
    Estimate $(\hat{e}_p, \hat{\alpha}_p)$ and construct $\mathbf{f}_p$.
    Compute $\hat{A}_{00,p}$, G_p, and $S_{\mathrm{res},p}$.
    Compute S_p.
end for
Compute G_{th}.
$\mathcal{P} \leftarrow \{p : \ G_p \geq G_{\mathrm{th}}, \ S_{\mathrm{res},p} \geq S_{\mathrm{res,th}}\}$
if $\mathcal{P} \neq \emptyset$ **then**
    $p^\star = \arg\max_{p \in \mathcal{P}} S_p$
    Output: $(\hat{\alpha}_{p^\star}, \hat{e}_{p^\star})$
else
    Output: no reliable PRACH peak
end if

 doi:10.20944/preprints202605.0050.v1

5. Manifold Interpretation

The validation framework introduced in Section 4 also admits a geometric interpretation, which helps explain the complementary behavior of the projection-based and residual-based criteria. This interpretation is provided for qualitative insight and does not directly affect the validation criterion.

The reduced feature vector $c_p \in \mathbb{C}^L$ has $2L$ real degrees of freedom; for $L = 21$, this corresponds to a 42-dimensional real observation space. In contrast, the noiseless reduced-form model in (5) depends on only four real parameters, $\Re(A_{00})$, $\Im(A_{00})$, α, ϵ. Accordingly, the set of all noiseless feasible feature vectors

$$\mathcal{M} = \{A_{00}\,\mathbf{f}(\epsilon,\alpha) : A_{00} \in \mathbb{C},\ \epsilon \in \mathbb{R},\ \alpha \in \mathbb{R}\} \subset \mathbb{C}^L \tag{17}$$

forms a four-dimensional manifold embedded in the $2L$-dimensional real space of correlation observations.

From this perspective, a true PRACH peak produces a feature vector that lies close to the manifold $\mathcal{M}$, as illustrated in Figure 3. The GLRT statistic G_p measures the projection of $\mathbf{c}_p$ onto the model subspace, while the residual energy R_p quantifies the distance of $\mathbf{c}_p$ from this manifold. This deviation primarily reflects channel-induced effects, such as multipath propagation, which are not explicitly captured by the analytical model.

For a valid peak, both criteria are consistent: the feature vector is well aligned with $\mathcal{M}$ and the residual energy remains close to the noise floor. In contrast, CFO-induced pseudo-peaks may partially align with the model subspace due to local similarity with the analytical correlation structure, leading to relatively large projection values and potentially high GLRT statistics. However, these candidates typically exhibit a larger distance to the manifold, resulting in higher residual energy.

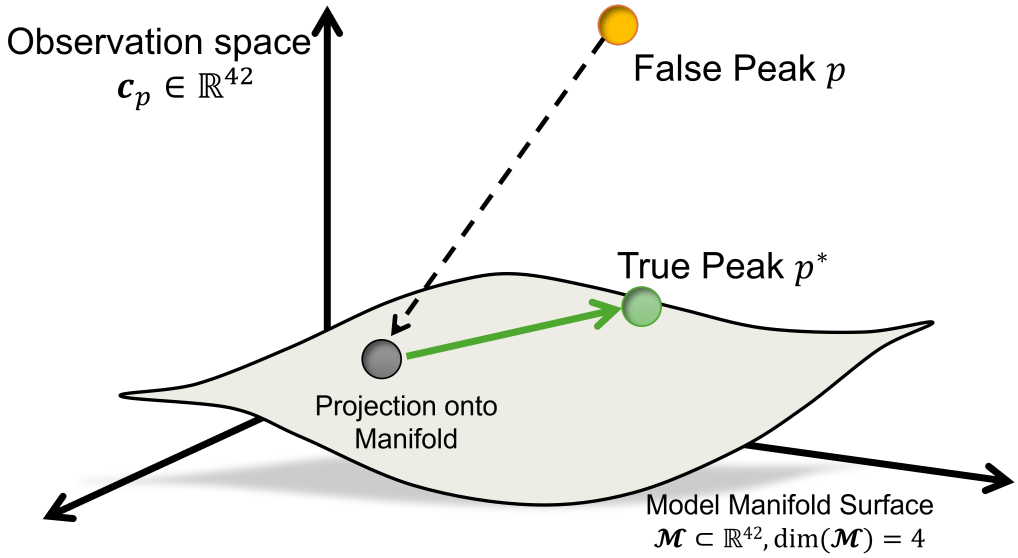


Figure 3. Manifold interpretation of PRACH features: valid peaks lie close to the model manifold, while pseudo-peaks exhibit larger deviation.

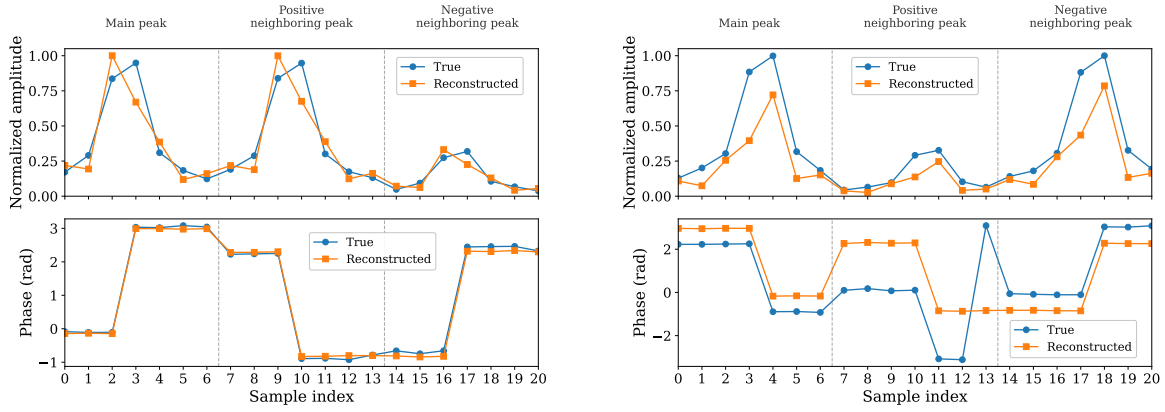
5.1. Illustrative Example Supporting the Manifold Interpretation

To further illustrate the geometric interpretation introduced above, we consider a representative PRACH realization under large CFO using the reduced 21-sample feature representation, as shown in Figure 4. Two detected candidate peaks are examined: the true peak and a falsely detected peak. For each candidate, the observed feature samples are compared with the reconstructed samples obtained from the estimated parameters $(\hat{\alpha}_p, \hat{\epsilon}_p)$ and the estimated complex amplitude $\hat{A}_{00,p}$.

For the true peak, the reconstructed signal closely matches the observed local feature in both amplitude and phase, indicating good consistency with the analytical model. In contrast, for the false peak, the reconstructed signal exhibits a visibly larger mismatch, especially in the relative lobe

 doi:10.20944/preprints202605.0050.v1

structure and phase evolution. This difference is reflected by the validation scores reported in Table 1, where the true peak yields substantially higher GLRT, residual, and weighted scores than the false peak.



(a) (b)

Figure 4. Illustrative example of 21-sample feature reconstruction under large CFO. (a) True detected peak: the reconstructed model closely matches the observed feature samples in both amplitude and phase. (b) False detected peak: the reconstructed model exhibits significant mismatch, particularly in phase and relative lobe structure.

Table 1. Validation scores for the illustrative example.

Candidate	$\hat{\alpha}$	$\hat{\epsilon}$	G_p	$S_{\mathrm{res},p}$	S_p
True peak	-0.4416	0.5003	1.000	0.929	0.979
False peak	0.4746	-0.5217	0.454	0.423	0.444

6. Computational Complexity and Practical Considerations

6.1. Computational Complexity

The proposed peak disambiguation framework operates on a reduced feature representation of size $L = 21$ for each detected candidate peak. For a given candidate p, the main computational steps include the construction of the feature vector $\mathbf{c}_p$ from L samples, the evaluation of the model vector $\mathbf{f}_p$, the computation of the projection $\mathbf{f}_p^H \mathbf{c}_p$, the estimation of $\hat{A}_{00,p}$, and the computation of the GLRT statistic G_p and residual energy R_p.

All these operations are dominated by inner products and vector norms of length L, resulting in a per-candidate complexity of $\mathcal{O}(L)$. Since L is fixed and small ($L = 21$), this complexity is effectively constant.

Let P denote the number of detected candidate peaks for a given PRACH occasion. The total complexity of the proposed method is therefore $\mathcal{O}(P\,L)$, i.e., linear in the number of candidates.

In comparison, a simple amplitude-based selection rule requires only a search over the P candidates and has complexity $\mathcal{O}(P)$. The proposed method introduces only a small additional constant factor due to the evaluation of the statistical criteria, while providing improved robustness under large CFO conditions.

Importantly, the feature extraction and parameter estimation steps are already required for TA/CFO estimation and are shared across all candidate-validation strategies. The additional computational cost of the proposed framework is therefore limited to a small number of vector operations per candidate.

Given the small feature dimension and the typically limited number of candidate peaks produced by the CFAR detector, the overall complexity remains compatible with real-time PRACH receiver implementations.

 doi:10.20944/preprints202605.0050.v1

6.2. Practical Considerations and Limitations

This subsection discusses the practical applicability and limitations of the proposed framework in realistic receiver implementations.

The method operates at the candidate level and is therefore naturally compatible with receiver pipelines where multiple detected peaks are processed independently, including scenarios where different peaks may originate from different users. However, resolving overlapping or strongly interfering PRACH signatures would require joint multi-user modeling and is beyond the scope of this work.

The oracle results presented in Section 7 isolate the impact of the selection rule from parameter estimation errors. The observed performance gap indicates that, in challenging conditions, estimation accuracy becomes the dominant limiting factor rather than the validation criterion itself. For example, at 10 dB in the long-PRACH cases, the oracle weighted score reaches approximately 0.70 in AWGN and 0.66 in TDL-C, compared with 0.60 and 0.50 using estimated parameters, respectively.

These observations highlight that the proposed framework addresses peak validation under model mismatch, while complementary improvements in parameter estimation and multi-user processing remain important directions for further performance gains.

7. Simulation Results and Performance Evaluation

This section evaluates the proposed PRACH peak disambiguation framework under large CFO conditions using Monte Carlo simulations aligned with 3GPP channel modeling assumptions.

7.1. Simulation Setup

All evaluated methods share the same receiver pipeline, including PRACH correlation, CFAR-based peak detection, reduced feature extraction ($L = 21$ samples), and TA/CFO estimation. Peak selection is then performed using four strategies: (i) amplitude-based candidate selection, (ii) GLRT-only validation, (iii) residual-energy-based validation, and (iv) the proposed weighted GLRT–residual criterion.

The evaluated scenarios are summarized in Table 2. In all cases, a non-zero CFO is applied to reproduce the large-CFO regime targeted in this work. The normalized CFO ϵ is defined with respect to the PRACH subcarrier spacing (SCS).

Table 2. Evaluated PRACH Scenarios (3GPP-Aligned).

Scenario	SCS	Channel	CFO (Hz)
Short (B4)	30 kHz	AWGN	3334
Short (B4)	30 kHz	TDL-C (6 taps)	3334
Long (F0)	1.25 kHz	AWGN	625
Long (F0)	1.25 kHz	TDL-C (6 taps)	625
NTN (A2)	15 kHz	TDL-C (4 taps)	1740
NTN (A2)	15 kHz	TDL-A (8 taps)	1740

Detailed simulation parameters are reported in Table 3. For each configuration, 1000 independent realizations are generated. The normalized delay α is uniformly drawn from $[0, 3]$, and the SNR ranges from -10 to 20 dB.

 doi:10.20944/preprints202605.0050.v1

Table 3. Simulation Parameters.

Parameter	Value
Number of realizations	1000 per scenario
SNR range	-10 to 20 dB
Normalized delay α	Uniform in $[0, 3]$
Normalized CFO ϵ	Scenario-dependent
Feature size L	21 samples (3 lobes $\times$ 7)
CFAR detector	Fixed configuration
P_{FA} (GLRT)	10^{-3}
Residual threshold	Two-stage tuning (low/high SNR)
Weight λ	Two-stage tuning (low/high SNR)
Noise variance	Estimated from PDP
Random seed	Fixed

Multipath channels follow simplified 3GPP TDL profiles derived from standardized channel models [10], while NTN scenarios are based on high-Doppler configurations consistent with non-terrestrial deployments described in [11]. PRACH configurations follow the specifications in [12].

7.2. Parameter Selection Strategy

Section 4 introduced the general two-regime parameterization principle. In the simulations, this principle is instantiated through a two-stage sweep procedure used to select the residual threshold and the weighting parameter λ for each scenario.

To ensure practical applicability while limiting parameter tuning, two operating regimes are defined: a low-SNR regime (SNR $\leq$ 5 dB), where performance is noise-limited, and a high-SNR regime (SNR $>$ 5 dB), where ambiguity between competing peaks dominates. For each scenario, a pair of parameters (residual threshold and λ) is selected for each regime, resulting in two parameter sets per scenario.

This coarse two-regime parameterization provides a good trade-off between robustness and implementation simplicity, as it avoids fine per-SNR tuning while capturing the main transition between noise-limited and ambiguity-limited operating conditions.

7.3. Results

For reference, a simple amplitude-based selection rule is also included, where the peak with the maximum correlation magnitude is selected. While this approach is commonly used as a baseline, it is not suitable in practical PRACH receivers under multi-user conditions, where multiple candidate peaks must be detected and processed rather than selecting a single global maximum. It is therefore used here only as a reference to highlight the impact of CFO-induced ambiguity. Across all evaluated scenarios, the proposed weighted criterion consistently provides a strong and robust trade-off between precision and recall.

For short PRACH under AWGN (Figure 5a), all methods perform near-optimally at high SNR. At 10 dB, GLRT and amplitude-based selection both reach an F1-score of 0.98, while the weighted method achieves 0.97 and the residual-only approach remains lower at 0.94. Under multipath conditions (Figure 5b), the impact of model mismatch becomes visible: GLRT drops to 0.89, residual-only to 0.68, while amplitude-based and weighted methods both reach 0.94. This indicates that, although ambiguity remains limited in these scenarios, combining criteria improves robustness.

For long PRACH under AWGN (Figure 6a), the situation becomes significantly more challenging due to the very large normalized CFO ($\epsilon \approx 0.5$). At 10 dB, GLRT achieves 0.49, amplitude-based selection 0.52, and residual-only drops to 0.06, while the weighted method improves performance to 0.60. Under multipath conditions (Figure 6b), the degradation is even more pronounced: GLRT falls to 0.2, residual-only to 0.13, while amplitude-based selection reaches 0.53 and the weighted method 0.50. These results confirm that the most critical cases are driven by large CFO combined with channel conditions, rather than the PRACH format itself.

 doi:10.20944/preprints202605.0050.v1

In NTN scenarios (Figure 7a,b), the normalized CFO remains moderate ($\epsilon \approx 0.116$), while the main impairment is high Doppler. As a result, ambiguity is less severe than in the high-CFO cases. At 10 dB, in the TDL-A scenario (Figure 7a), GLRT achieves 0.88, residual-only 0.69, while amplitude-based and weighted methods both reach 0.94. In the TDL-C scenario (Figure 7b), all methods perform strongly, with GLRT at 0.92, residual-only at 0.68, and both amplitude-based and weighted methods reaching 0.94. These results confirm that NTN scenarios are primarily Doppler-limited rather than CFO-limited.

The oracle F1-score, computed using true parameter values, remains consistently higher (e.g., 0.70 in long AWGN and 0.66 in long TDL-C at 10 dB), indicating that the remaining performance gap is primarily due to estimation errors in $(\hat{\alpha}, \hat{\epsilon})$ rather than limitations of the proposed validation framework.

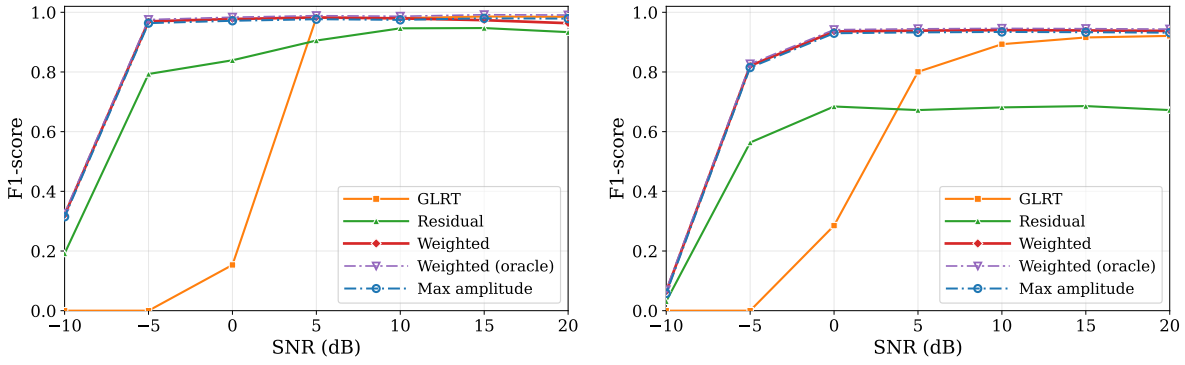


Figure 5. Performance versus SNR for short PRACH scenarios. (a) AWGN with large CFO; residual threshold 0.01 (low SNR) and 0.05 (high SNR), $\lambda = 0.30$ (low SNR) and 0.90 (high SNR). (b) TDL-C multipath channel; residual threshold 0.01 (low SNR) and 0.05 (high SNR), $\lambda = 0.30$ (both regimes).

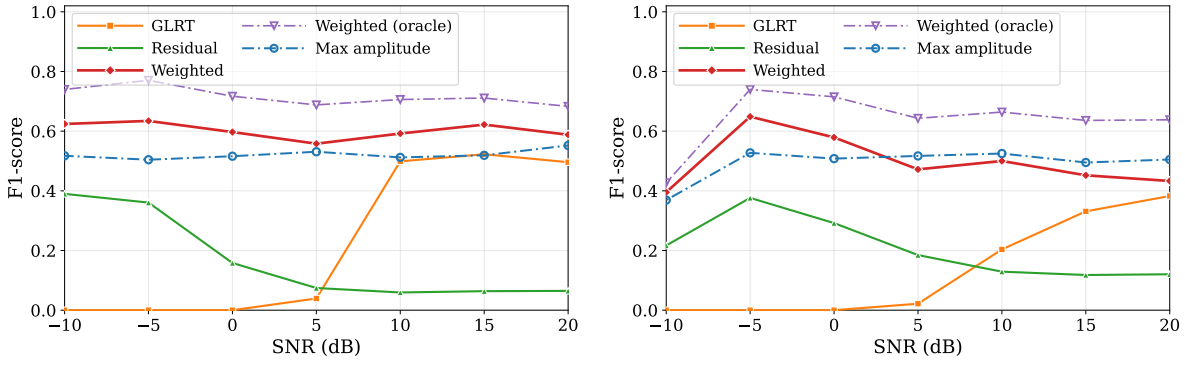


Figure 6. Performance versus SNR for long PRACH scenarios. (a) AWGN with large CFO; residual threshold 0.01 (both regimes), $\lambda = 0.90$ (both regimes). (b) TDL-C multipath channel; residual threshold 0.05 (both regimes), $\lambda = 0.50$ (low SNR) and 0.90 (high SNR).

 doi:10.20944/preprints202605.0050.v1

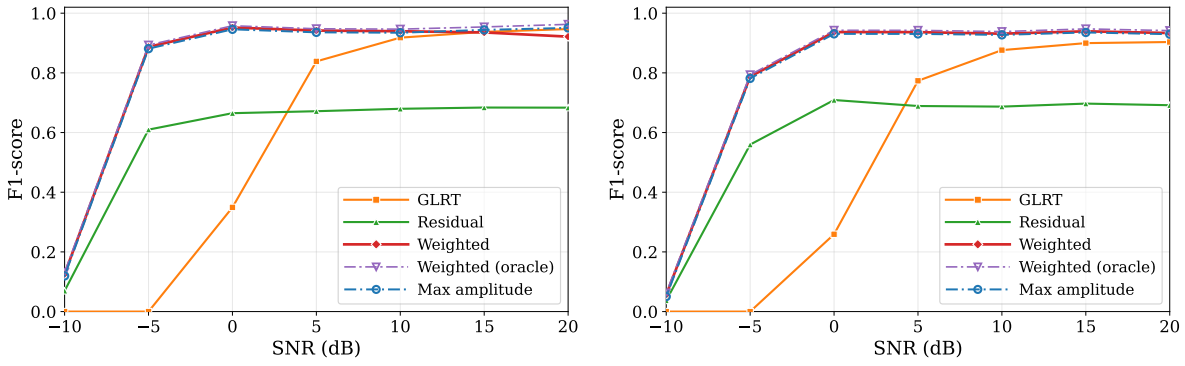


Figure 7. Performance versus SNR for NTN scenarios with moderate CFO and high Doppler. (a) TDL-A channel; residual threshold 0.01 (low SNR) and 0.10 (high SNR), $\lambda = 0.50$ (low SNR) and 0.90 (high SNR). (b) TDL-C channel; residual threshold 0.01 (both regimes), $\lambda = 0.30$ (low SNR) and 0.70 (high SNR).

Precision and recall values at 10 dB are summarized in Table 4, providing additional insight into the trade-off captured by the F1-score curves. These results suggest that performance degradation is primarily driven by the magnitude of the normalized CFO rather than the PRACH format itself.

Table 4. Precision and Recall at 10 dB.

	Precision				Recall			
Scenario	Max	Wght.	GLRT	Res.	Max	Wght.	GLRT	Res.
Short AWGN	1.00	1.00	1.00	1.00	0.96	0.96	0.96	0.90
Short TDL-C	0.97	0.97	0.98	0.96	0.91	0.91	0.82	0.53
Long AWGN	0.52	0.59	0.54	0.06	0.52	0.59	0.47	0.06
Long TDL-C	0.53	0.50	0.55	0.17	0.53	0.50	0.13	0.10
NTN TDL-A	0.97	0.97	0.98	0.97	0.90	0.89	0.79	0.53
NTN TDL-C	0.97	0.97	0.97	0.96	0.91	0.91	0.87	0.53

7.4. Discussion

The comparative results can be interpreted as an ablation study of the proposed criteria. GLRT-only provides strong selectivity when the analytical model accurately represents the observed correlation structure, but its sensitivity to mismatch limits its robustness in challenging conditions. Residual-only validation is more tolerant to mismatch, yet lacks sufficient discrimination capability when used alone. The weighted combination balances these two behaviors, which explains its more consistent performance across all evaluated scenarios.

An important observation is that the most critical cases are governed by the combined effect of large normalized CFO (which induces ambiguity) and channel conditions such as multipath (which induce model mismatch) rather than by the PRACH format alone. Scenarios with very large normalized CFO exhibit strong ambiguity in the correlation structure, especially when combined with multipath propagation. In contrast, scenarios with moderate CFO, such as the considered NTN cases, remain less ambiguous despite the presence of high Doppler.

The oracle results further clarify the origin of the remaining performance gap. Since the oracle relies on the true (α, ϵ) parameters, it isolates the impact of the selection rule from that of the estimator. The relatively small gap observed in mild conditions indicates that the proposed validation framework is near-optimal when the parameter estimates are accurate. In more challenging scenarios, the gap increases, showing that estimation errors in $(\hat{\alpha}, \hat{\epsilon})$ become the dominant limitation rather than the selection strategy itself.

Overall, these results confirm that the proposed weighted validation provides a robust and practical solution for PRACH peak disambiguation under large CFO, while maintaining resilience to channel-induced distortions. Across all evaluated scenarios, the proposed method maintains

 doi:10.20944/preprints202605.0050.v1

consistent performance without requiring fine parameter tuning, confirming its robustness under practical operating conditions.

8. Conclusions

This paper addressed PRACH peak disambiguation under large CFO, where a single transmission may generate multiple candidate peaks and obscure the true timing hypothesis. Starting from a reduced-form analytical model of the local PRACH correlation structure, we first formulated a GLRT-based validation criterion for candidate peaks. We then showed that projection-based validation alone is insufficient when CFO-induced peak ambiguity is combined with multipath-induced model mismatch, since pseudo-peaks may still partially align with the analytical model.

To address this limitation, a residual-energy-based consistency criterion was introduced and combined with the GLRT through a weighted selection rule. A geometric manifold interpretation was further provided to explain the complementary roles of projection-based and distance-based validation.

Simulation results under 3GPP-aligned scenarios, including multipath and high-Doppler NTN conditions, demonstrated that the proposed weighted framework provides robust peak disambiguation across a wide range of operating conditions. The approach maintains low computational complexity and requires only coarse parameter adaptation through two operating regimes, making it suitable for practical PRACH receiver implementations.

Author Contributions: Conceptualization, H.K. and T.L.; methodology, H.K. and T.L.; software, H.K.; validation, H.K. and T.L.; formal analysis, H.K. and T.L.; investigation, H.K. and T.L.; resources, H.K.; data curation, H.K.; writing—original draft preparation, H.K.; writing—review and editing, H.K. and T.L.; visualization, H.K. and T.L.; supervision, H.K. and T.L.; project administration, H.K. and T.L.; funding acquisition, H.K. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are generated synthetically for simulation purposes.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Linsalata, F.; Magarini, M.; Ferrari, R. Characterization of error events and design of a robust receiver for PRACH detection. *Phys. Commun.* **2020**, *41*, 101092.
2. Chu, D.C. Polyphase codes with good periodic correlation properties. *IEEE Trans. Inf. Theory* **1972**, *18*, 531–532.
3. Hua, Y.; Li, C.; Wu, J. Analysis of the frequency offset effect on Zadoff–Chu sequences. *IEEE Trans. Commun.* **2014**, *62*, 4024–4039.
4. Chougrani, H.; Kisseleff, S.; Martins, W.A.; Chatzinotas, S. NB-IoT random access for nonterrestrial networks: Preamble detection and uplink synchronization. *IEEE Internet Things J.* **2022**, *9*, 14913–14927.
5. Tao, J.; Yang, J. Improved Zadoff–Chu sequence detection in the presence of multipath and frequency offset. *IEEE Commun. Lett.* **2018**, *22*, 922–925.
6. Yang, X.; Fapojuwo, A.O. Enhanced preamble detection for PRACH in LTE. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), 2013; pp. 3306–3311.
7. Wang, Q.; Ren, G.; Wu, J. A multiuser detection algorithm for random access procedure with the presence of carrier frequency offsets in LTE systems. *IEEE Trans. Commun.* **2015**, *63*, 3299–3312.
8. Zhen, L.; Cheng, L.; Chu, Z.; Yu, K.; Xiao, P.; Guizani, M. Clustering-NN based CFO estimation using random access preambles for 5G non-terrestrial networks. *IEEE Wireless Commun. Lett.* **2023**, *13*, 587–591.
9. Khaleghi, H.; Paquelet, S. Unleashing timing advance precision: CRLB limit and neural implementation. In Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2024.

 doi:10.20944/preprints202605.0050.v1

10. 3GPP. NR; Base station (BS) conformance testing; Part 1: Conducted conformance testing. *3GPP TS 38.141-2*; 2023.
11. 3GPP. NR; Satellite Access Node radio transmission and reception. *3GPP TS 38.108*; 2024.
12. 3GPP. NR; Physical channels and modulation. *3GPP TS 38.212*; 2022.