

Matched-Filter Based Frame-Start Detector Resilient to Frequency Offset

Jozef LUKAC, Michael KIMMER, Jan SYKORA

Dept. of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University, Technicka 2, 166 27 Prague 6

lukacjo1@fel.cvut.cz, michaelkimmer11@gmail.com, jan.sykora@fel.cvut.cz

Submitted July 7, 2025 / Accepted November 4, 2025 / Online first November 14, 2025

Abstract. *The article proposes a (phase-marginalized) frame-start detector resilient to frequency offset between the transmitter and the receiver. The detector is a slight modification of the standard detector that looks for a known sequence/preamble using the matched filter, respectively correlation of the received signal with the known sequence. The modification is that the original long sequence is divided into shorter ones; the magnitude or magnitude-squared of the output of matched-filters is taken and then appropriately added to get the final detector metric. The proposed detector is a low-cost alternative to the standard approach (joint estimation of frame timing and frequency offset), where we need as many matched-filters as the number of testing frequency offsets. We derive probability characteristics of the proposed detector that reflect its resilience to the frequency offset. The frequency resilience of the detector is proportional to the number of segments into which the preamble is divided.*

Keywords

Frame synchronization, matched filtering, crosscorrelation, frequency offset

1. Introduction

Estimation of the beginning of a frame is one of the initial procedures the receiver needs to do. Even the most simplified channel models of a real over-the-air channel include the frequency offset between the transmitter and the receiver. Maximum likelihood (ML) joint-estimation of the frame-timing and the carrier-frequency-offset in the ordinary channel – with just additive white Gaussian noise, and carrier-frequency and carrier-phase offset – requires as many correlators as is the number of testing frequency-offsets. Hence, some suboptimal simplified algorithms emerged that do not require such extensive computational power, but instead decouple the timing and carrier-frequency-offset estimation. The article focuses on the frame-start detector that searches for a known pilot-signal/preamble without first, or simultaneous, estimation of the frequency offset. The pilot signal is often followed by a payload signal that carries data, but we will assume no data here.

The detector algorithms for estimation of frame-start – that are resilient to frequency offset – are often designed for a specific pilot signal or modulation; e.g., [1] for M-ary Phase Shift Keying (M-PSK) alphabet, and [2] for Gaussian Minimum-Shift Keying (GMSK) modulated data. The famous Schmidl-Cox algorithm [3] assumes periodicity of the pilot signal, and uses the autocorrelation (AC) of the received signal with its shifted replica. Many modifications of the algorithm have been published [4],[5] (cyclic prefix as the periodic pilot-signal), [6]. The algorithms based on the Schmidl-Cox algorithm are referred to as autocorrelation algorithms. On the other hand, there are algorithms based on crosscorrelation (XC) of the received signal with the known shifted pilot signal. These are effectively realized using matched filter, e.g. [1, 7, 8]. The [9] compares AC and XC detection algorithms. We will follow a similar approach here.

There are also frame-start detectors based on the increase of power on antenna(s) [10] – that enable very cost-effective hardware (HW) implementation for the price of lower detector's performance.

Our detector is built on the idea that the optimal carrier-phase-marginalized ML frame-start detector for zero frequency offset is the simple XC-detector. The detector's metric is the absolute or absolute-squared value of the crosscorrelation of the received signal with the known pilot signal. The key observation is that using the XC-detector (optimal for zero freq. offset) with short pilot-signals, even in cases with non-zero frequency offset, does not degrade the detector's performance much. Our proposed detector leverages the observation, divides the preamble signal into short sequences, and computes the XC-metric "per partes" – that makes it resilient to frequency offset. The main idea originated in the master thesis [11] (pg. 32). Our simulation scripts and results in .mat files have been published on GitHub and archived using Zenodo [12].

During the review process, we have found an article [13] that proposes the same method: dividing the preamble into M parts. In our article, we analyze both variants of the detector's metric – the sum of squared absolute values of partial metrics and the sum of absolute values (without squaring)

of partial metrics. Additionally, we derive very accurate analytical expressions for probability of misdetection of both detectors' variants.

The outline of the article is as follows. Section 2 states the system model, Section 3 shows the derivation of probability characteristics, Section 4 presents numerical results, and documents two practical over-the-air examples; and Section 5 concludes the paper.

2. System Model

The section introduces a system model that the detector assumes. The detector chooses between the two hypotheses, H_0 and H_1 . The model is given in discrete time and also in continuous time. The discrete model is derived from the continuous by sampling using the sampling rate f_{sa} , respectively the sampling period $T_{sa} = 1/f_{sa}$. Further, we proceed with just the discrete time model. The obtained results are then interpreted also for the continuous time case.

The two detector hypotheses are

$$H_0 : x[n] = w[n], \quad (1)$$

$$H_1 : x[n] = \eta e^{j\varphi} e^{j2\pi F_\varepsilon n} h[n] + w[n]. \quad (2)$$

Their continuous-time equivalent hypotheses are as follows.

$$H_0 : x(t) = w(t), \quad (3)$$

$$H_1 : x(t) = \eta e^{j\varphi} e^{j2\pi f_\varepsilon t} h(t) + w(t). \quad (4)$$

$w[n]$ and $w(t)$ are complex-valued white Gaussian noise vectors, respectively, samples of noise vectors. $w[n]$'s variance is σ_w^2 , and $w(t)$'s noise density is N_0 . $\eta > 0$ is the channel attenuation, $\varphi \in (0, 2\pi)$ is a random carrier-phase, f_ε is the carrier-frequency offset in Hz. $F_\varepsilon = f_\varepsilon \cdot T_{sa} = f_\varepsilon / f_{sa}$ is the normalized carrier-frequency offset. $x[n]$ and $x(t)$ are the vectors of the received signal. $n \in \mathbb{Z}$ is a discrete-time instant, an index of a sample; $t \in \mathbb{R}$ is a continuous-time instant. $h(t)$ is the known preamble, and $h[n] = h(nT_{sa})$ are samples of the preamble.

The preamble $h[n]$ is divided into $N_{seq} \in \mathbb{N}$ parts, each part is of L_{seq} samples long. If necessary, a few zero samples are added to the last part to have the length of L_{seq} samples. The preamble can be expressed as a sum of its delayed segments, as follows.

$$h[n] = \sum_{\ell=0}^{N_{seq}-1} h_\ell[n - \ell L_{seq}], \quad (5)$$

$$= \begin{cases} h[n], & n \in \{0, \dots, N_h - 1\} \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$$h_\ell[n] = \begin{cases} h[n], & n \in \{0, \dots, L_{seq} - 1\} \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

$$\ell \in \{0, \dots, N_{seq} - 1\}.$$

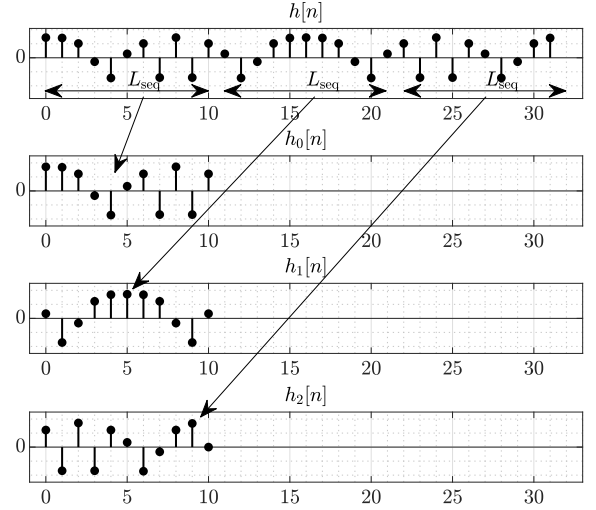


Fig. 1. An example of the partitioning of a preamble sequence $h[n]$ to $N_{seq} = 3$ subsequences.

N_h is the length of the $h[n]$ sequence. An example of the partitioning of the sequence $h[n]$ to $N_{seq} = 3$ subsequences is shown in Fig. 1 ($N_h = 32$, $N_{seq} = 3$, $L_{seq} = 11$). For illustrative purposes, we show an example of a real-valued $h[n]$.

Further, we define auxiliary sequences – matched-filter's (mf) impulse responses.

$$h_{mf}[n] = h^*[-n], \quad (8)$$

$$h_{\ell, mf}[n] = h_\ell^*[-n], \quad \ell \in \{0, \dots, N_{seq} - 1\}. \quad (9)$$

Upper index $*$ is the complex conjugation.

Optimal metric – the sufficient statistics – for the carrier-phase (φ) marginalized frame-start detector with zero frequency offset ($f_\varepsilon = 0$) is

$$|\mathcal{R}_{x,h}[n]| = |\langle x[k+n], h[k] \rangle| = \left| \sum_{k \in \mathbb{Z}} x[k+n] h^*[k] \right|,$$

$$= \left| \sum_{k \in \mathbb{Z}} x[n+k] h_{mf}[-k] \right| = |(x * h_{mf})[n]|. \quad (10)$$

$\mathcal{R}_{x,h}[n]$ is the crosscorrelation function/sequence of the sequences $x[n]$, $h[n]$. $|\cdot|$ represents the absolute value (abs) of a complex number. $\langle \cdot, \cdot \rangle$ is the inner product of sequences, and $*$ stands for the convolution operation. Using (5) in the metric (10), we get

$$|\mathcal{R}_{x,h}[n]| = |\langle x[k+n], h[k] \rangle|,$$

$$= \left| \sum_{\ell=0}^{N_{seq}-1} \langle x[k+n], h_\ell[k - \ell L_{seq}] \rangle \right|,$$

$$= \left| \sum_{\ell=0}^{N_{seq}-1} \mathcal{R}_{x,h_\ell}[n + \ell L_{seq}] \right|. \quad (11)$$

If necessary, the equation (11) can be squared. We propose to approximate the absolute, resp. absolute-squared sum by the sum of absolute/absolute-squared values as follows ($|\cdot|$ and $|\cdot|^2$ moves on summands).

$$|\mathcal{R}_{x,h}[n]| \stackrel{1)}{\approx} \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{R}_{x,h_\ell}[n + \ell L_{\text{seq}}]|, \quad (12)$$

$$|\mathcal{R}_{x,h}[n]|^2 \stackrel{2)}{\approx} \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{R}_{x,h_\ell}[n + \ell L_{\text{seq}}]|^2. \quad (13)$$

We will refer to (12) as approximation 1), and to (13) as approximation 2). The approximations 1) and 2) are the detectors' metric of our proposed detector(s).

If the preamble $h[n]$ is periodic with period L_{seq} , the equations (12), and (13) became

$$\begin{aligned} |\mathcal{R}_{x,h}[n]| &\stackrel{1)}{\approx} \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{R}_{x,h_0}[n + \ell L_{\text{seq}}]| \\ &= |\mathcal{R}_{x,h_0}[n]| * \Delta[n], \end{aligned} \quad (14)$$

$$\begin{aligned} |\mathcal{R}_{x,h}[n]|^2 &\stackrel{2)}{\approx} \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{R}_{x,h_0}[n + \ell L_{\text{seq}}]|^2 \\ &= |\mathcal{R}_{x,h_0}[n]|^2 * \Delta[n], \end{aligned} \quad (15)$$

$$\begin{aligned} \Delta[n] &:= \sum_{\ell=0}^{N_{\text{seq}}-1} \delta[n + \ell L_{\text{seq}}], \\ &= \begin{cases} 1, & n \in \{-(N_{\text{seq}}-1)L_{\text{seq}}, \dots, -L_{\text{seq}}, 0\} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (16)$$

$(f * g)[n]$ denotes convolution of sequences $f[n]$ and $g[n]$. Sometimes the convolution is also denoted as $f[n] * g[n]$. $\delta[n]$ is Kronecker delta, and $\Delta[n]$ is an auxiliary sequence – a part of Kronecker comb $\Delta_{L_{\text{seq}}}[n]$. Kronecker comb, $\Delta_N[n]$, is a periodic sequence of Kronecker-delta pulses. The period is N .

$$\Delta_N[n] = \sum_{k \in \mathbb{Z}} \delta[n - kN]. \quad (17)$$

Next, we will name the metric in approximation 1) as $y'_1[n]$, and the metric in approximation 2) as $y'_2[n]$. Additionally, we will name also intermediate signals.

$$\ell \in \{0, \dots, N_{\text{seq}}-1\}, \quad (18)$$

$$x_\ell[n] := (x * h_{\ell,\text{mf}})[n], \quad (19)$$

$$y_{\ell,1}[n] := |x_\ell[n]|, \quad (20)$$

$$y_{\ell,2}[n] := |x_\ell[n]|^2, \quad (21)$$

$$y'_1[n] := \sum_{\ell=0}^{N_{\text{seq}}-1} y_{\ell,1}[n + \ell L_{\text{seq}}], \quad (22)$$

$$y'_2[n] := \sum_{\ell=0}^{N_{\text{seq}}-1} y_{\ell,2}[n + \ell L_{\text{seq}}]. \quad (23)$$

The detector's metric $y'_1[n]$, and $y'_2[n]$ are then compared to their corresponding threshold values $r_{1,\text{thr}}$, $r_{2,\text{thr}}$. The exceeding of the threshold indicates the presence of the preamble in the near vicinity of the time index. The search for arg-max in the vicinity of the given time-index then determines the exact position of the frame-start. Block schemes of the two detectors are in Fig. 2. The auxiliary/intermediate signals are shown there.

We note that to make a filter physically realizable, it needs to be causal. Causality is achieved by delaying the filter's impulse response so that it is zero in negative time instants. Matlab implementation of the detectors' metrics is in the function `MF_correlate`, see [12]. During the review process, authors realized that the finer tuning of the detector can be achieved not by setting N_{seq} , but by setting L_{seq} . The last sequence should be filled with zeros to be L_{seq} long.

In the Sec. 3, we will show that optimal threshold values $r_{1,\text{thr}}$ and $r_{2,\text{thr}}$ depend on the noise variance, σ_w^2 . The dependence is expressed in the form

$$r_{1,\text{thr}} = \sigma_w f_1(p_{\text{fa}}), \quad (24)$$

$$r_{2,\text{thr}} = \sigma_w^2 f_2(p_{\text{fa}}). \quad (25)$$

f_1 and f_2 are functions of the false-alarm probability. Relations (24) and (25) are derived in the Sec. 3.1, see (56), and (57). The noise variance is usually not known, so it needs to be estimated. The detector's implementation then looks as in Fig. 3. There are two equivalent schemes; the upper is suitable for statistical derivations, and the lower for HW implementation.

It might be challenging to estimate the noise variance, as we are unsure if there is no transmission at that time. Looking at the scheme in Fig. 3, we propose to use another reference level – a plateau-level of the detector's own metric. We illustrate the plateau-level and the peak of the metric $y'_1[n]$ in Fig. 4. As the plateau-level estimators, we suggest moving average (mAv) or moving median (mMed).

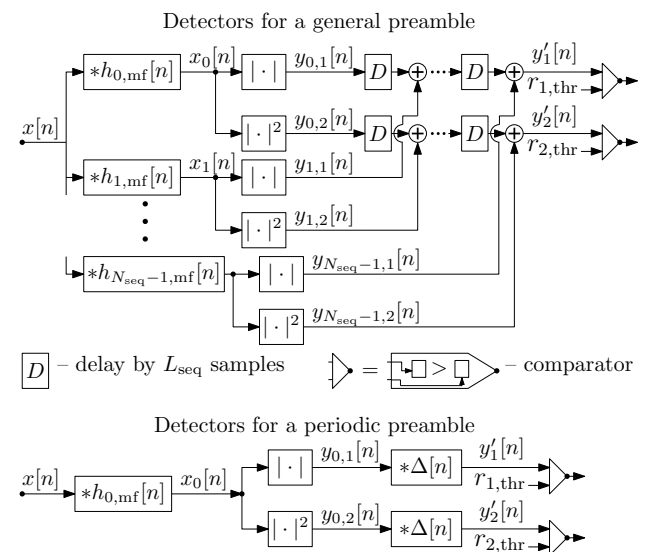


Fig. 2. The scheme of the proposed frame-start detectors.

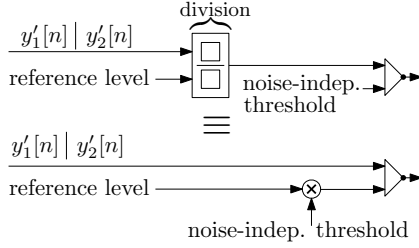


Fig. 3. Two equivalent schemes of the detector with a reference level and a noise-independent threshold.

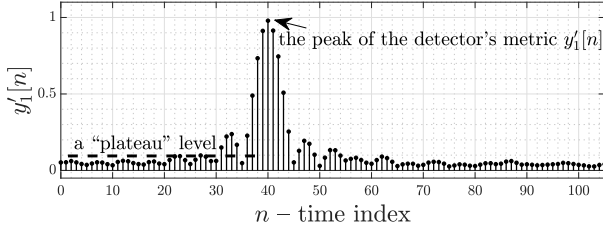


Fig. 4. An example of the $y'_1[n]$ detector metric with the peak and a plateau-level shown.

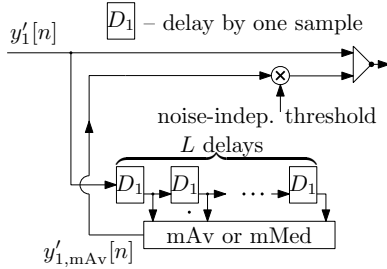


Fig. 5. A block diagram of the detector with a mAv reference-level estimator, for input metric $y'_1[n]$.

The mAv and mMed plateau-level estimators are defined as follows.

$$y'_{1,\text{mAv}}[n] := \frac{1}{L} \sum_{k=1}^L y'_1[n-k], \quad (26)$$

$$y'_{1,\text{mMed}}[n] := \text{med} \left(\{y'_1[n-k]\}_{k=1}^L \right), \quad (27)$$

$$y'_{2,\text{mAv}}[n] := \frac{1}{L} \sum_{k=1}^L y'_2[n-k], \quad (28)$$

$$y'_{2,\text{mMed}}[n] := \text{med} \left(\{y'_2[n-k]\}_{k=1}^L \right). \quad (29)$$

L is the averaging length.

Figure 5 shows the moving-average plateau-level estimator $y'_{1,\text{mAv}}[n]$ applied in the detector's block diagram. We get a detector with the constant false-alarm rate (CFAR) – independent of noise variance – similar to [2].

3. Probability Characteristics

The section shows derivations of the probabilistic characteristics of the detector – the probability of mis-detection p_{md} (probability of not detecting a frame when it is present) given a fixed false-alarm probability p_{fa} . First, we derive

p_{fa} , then p_{md} . In both cases, we proceed step-by-step by deriving probability distributions of all the intermediate auxiliary signals (19)–(23). In the derivation of p_{fa} , we assume the hypothesis H_0 . H_1 is assumed for the derivation of p_{md} . We approximate all the random variables (RVs) to be independent, even though they are not exactly. The assumption significantly simplifies derivations. The accuracy of the derived formulas is compared to the numerical simulation in Sec. 4. The numerical results match the derived formulas well, supporting our assumptions and approximations.

In derivations that aim for p_{fa} , the following probability distributions occur: Complex Normal, Rayleigh, Exponential, and Erlang. Further, the approximation of the sum of Rayleigh RVs is taken from [14], and we propose a parameter adjustment in it. The distribution is denoted as RaylS in our paper. In derivations aiming for p_{md} , the following probability distributions are used: Complex Normal, Rice, Non-central χ -squared, and (Real) Normal. The Normal distribution approximates the sum of Rice RVs by matching the mean and variance. Formulas for the mean and variance of the mentioned probability distributions are applied in some partial steps; some formulas are derived using Wolfram Mathematica.

3.1 Probability of False Alarm

The probability that the detector's metric is greater than a threshold given hypothesis H_0 , i.e., no frame present, is called the probability of false alarm, p_{fa} . The distributions of intermediate signals follow.

$$x[n] = w[n] \sim \mathcal{CN}(0, \sigma_w^2), \quad (30)$$

$$x_\ell[n] = (x * h_{\ell,\text{mf}})[n] = \sum_{k \in \mathbb{Z}} x[n-k] h_{\ell,\text{mf}}[k], \quad (31)$$

$$x_\ell[n] \sim \mathcal{CN}(0, \|h_\ell\|^2 \sigma_w^2), \quad (32)$$

$$E_{h_\ell} := \|h_\ell\|^2 = \sum_{k \in \mathbb{Z}} |h_\ell[k]|^2 = \sum_{k \in \mathbb{Z}} |h_{\ell,\text{mf}}[k]|^2, \quad (33)$$

$$y_{\ell,1}[n] = |x_\ell[n]| \sim \text{Rayl} \left(\sigma_\ell = \sqrt{\frac{\|h_\ell\|^2 \sigma_w^2}{2}} \right), \quad (34)$$

$$y_{\ell,2}[n] = |x_\ell[n]|^2 \sim \text{Exp} \left(\lambda_\ell = \frac{1}{\|h_\ell\|^2 \sigma_w^2} \right). \quad (35)$$

$\mathcal{CN}(\mu, \sigma^2)$ stands for complex-normal distribution with mean μ , and variance σ^2 ; E_{h_ℓ} is the signal-energy of the sequence $h_\ell[n]$; $\text{Rayl}(\sigma)$ denotes Rayleigh-distributed RV with parameter σ , and $\text{Exp}(\lambda)$ indicates exponentially distributed RV with parameter λ .

In order to proceed, we need to assume $\{E_{h_\ell}\}_\ell$ to be equal or at least approximately equal to each other: $E_{h_\ell} = E_{h_0}$. The equivalent assumptions are $\lambda_\ell = \lambda_0$, and $\sigma_\ell = \sigma_0$.

$$y'_1[n] = \sum_{\ell=0}^{N_{\text{seq}}-1} y_{\ell,1}[n + \ell L_{\text{seq}}], \quad (36)$$

$$y'_1[n] \sim \text{RaylS}(N_{\text{seq}}, \sigma_0), \quad (37)$$

$$y'_2[n] = \sum_{\ell=0}^{N_{\text{seq}}-1} y_{\ell,2}[n + \ell L_{\text{seq}}], \quad (38)$$

$$y'_2[n] \sim \text{Erl}(N_{\text{seq}}, \lambda_0). \quad (39)$$

$\text{Erl}(k, \lambda)$ marks Erlang distribution with k degrees of freedom and parameter λ . $\text{Erl}(k, \lambda)$ represents the distribution of the sum of k independent RVs, each distributed according to $\text{Exp}(\lambda)$. $\text{RaylS}(k, \sigma)$ denotes distribution of the sum of k independent RVs, each distributed according to $\text{Rayl}(\sigma)$.

The $\text{RaylS}(k, \sigma)$ distribution has no simple closed form. We take the approximation from [14]. Authors approximate $\text{RaylS}(k, \sigma)$ by the square root of $\text{Erl}(k, \lambda_{\text{aux}})$ and define an auxiliary variable b as $b = 1/(2\lambda_{\text{aux}})$. $b/2$ approximates the variance of $\text{RaylS}(k, \sigma)$. We propose to set b differently, namely, we set b to match the mean:

$$A \sim \text{RaylS}(k, \sigma), \quad (40)$$

$$\text{pdf}_A(\zeta) = \frac{\zeta^{2k-1} \exp\left(-\frac{\zeta^2}{2b}\right)}{2^{k-1} b^k (k-1)!} \mathbb{1}(\zeta), \quad (41)$$

$$\text{cdf}_A(\zeta) = 1 - Q_\Gamma\left(k, \frac{\zeta^2}{2b}\right), \quad (42)$$

$$\mu_A = \mathbb{E}[A] = \sqrt{2b} \frac{\Gamma(k + \frac{1}{2})}{(k-1)!}, \quad (43)$$

$$\mu_A \stackrel{!}{=} k \sqrt{\frac{\pi}{2}} \sigma, \quad (44)$$

$$b := \left(\frac{\Gamma(k+1)}{\Gamma(k + \frac{1}{2})} \right)^2 \frac{\pi}{4} \sigma^2 \approx k \frac{\pi}{4} \sigma^2. \quad (45)$$

pdf_A is the probability density function of the RV A ; cdf_A is the cumulative distribution function of the RV A ; $\mathbb{1}(\zeta)$ is the unit-step function; μ_A is the mean value of the RV A ; $\mathbb{E}[\cdot]$ is the mean-value operator; $\Gamma(x)$ is the gamma function; $Q_\Gamma(s, x)$ is the upper regularized incomplete gamma function. pdf_A in (41), cdf_A in equation (42), and μ_A in (43) are valid for any $b > 0$. We propose to match the mean of RV A , equation (44) – left side is valid for all $b > 0$, right side is the mean of the sum of k independent identically distributed Rayleigh RVs, each with parameter σ . The resulting b is given in (45). In notation, we will use the parameter σ or b interchangeably, $\text{RaylS}(k, \sigma) \equiv \text{RaylS}(k, b)$, but take into consideration that b depends on k .

Now, we recall the Erlang distribution.

$$A \sim \text{Erl}(k, \lambda), \quad (46)$$

$$\text{pdf}_A(\zeta) = \frac{\lambda^k \zeta^{k-1}}{(k-1)!} e^{-\lambda \zeta} \mathbb{1}(\zeta), \quad (47)$$

$$\text{cdf}_A(\zeta) = 1 - Q_\Gamma(k, \lambda \zeta) \quad (48)$$

$$= \left[1 - e^{-\lambda \zeta} \sum_{\ell=0}^{k-1} \frac{1}{\ell!} (\lambda \zeta)^\ell \right] \mathbb{1}(\zeta), \quad (49)$$

$$\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt, \quad (50)$$

$$Q_\Gamma(s, x) = \frac{\Gamma(s, x)}{\Gamma(s)}. \quad (51)$$

Finally, the probability of a false alarm is

$$\begin{aligned} p_{\text{fa},1} &:= \Pr\{y'_1[n] > r_{1,\text{thr}}\} = 1 - \Pr\{y'_1[n] \leq r_{1,\text{thr}}\} \\ &= 1 - \text{cdf}_{y'_1[n]}(r_{1,\text{thr}}) = Q_\Gamma\left(N_{\text{seq}}, \frac{r_{1,\text{thr}}^2}{2b}\right), \end{aligned} \quad (52)$$

$$b := \left(\frac{\Gamma(N_{\text{seq}} + 1)}{\Gamma(N_{\text{seq}} + \frac{1}{2})} \right)^2 \frac{\pi}{4} \frac{\|h_0\|^2 \sigma_w^2}{2}. \quad (53)$$

$$\begin{aligned} p_{\text{fa},2} &:= \Pr\{y'_2[n] > r_{2,\text{thr}}\} = 1 - \Pr\{y'_2[n] \leq r_{2,\text{thr}}\} \\ &= 1 - \text{cdf}_{y'_2[n]}(r_{2,\text{thr}}) = Q_\Gamma(N_{\text{seq}}, \lambda_0 r_{2,\text{thr}}), \end{aligned} \quad (54)$$

$$\lambda_0 := \frac{1}{\|h_0\|^2 \sigma_w^2}. \quad (55)$$

From (52) and (54), threshold values given p_{fa} are derived.

$$r_{1,\text{thr}} = \sqrt{2b Q_\Gamma^{-1}(N_{\text{seq}}, p_{\text{fa},1})}, \quad (56)$$

$$r_{2,\text{thr}} = \|h_0\|^2 \sigma_w^2 \Gamma^{-1}(N_{\text{seq}}, p_{\text{fa},2}), \quad (57)$$

$$Q_\Gamma^{-1}(s, y) = x \Leftrightarrow Q_\Gamma(s, x) = y. \quad (58)$$

For $N_{\text{seq}} = 1$, relations (56), and (57) reduce to

$$r_{1,\text{thr}} = \sqrt{E_h \sigma_w^2 (-1) \ln(p_{\text{fa},1})}, \quad (59)$$

$$r_{2,\text{thr}} = E_h \sigma_w^2 (-1) \ln(p_{\text{fa},2}). \quad (60)$$

$E_h = \sum_n |h[n]|^2$ is the energy of the sequence $h[n]$.

Both, $Q_\Gamma(s, x)$ and $Q_\Gamma^{-1}(s, x)$, are implemented in computational software languages; $Q_\Gamma(s, x)$ is implemented as `gammainc(x,s,'upper')` in Matlab, and as `GammaRegularized[s,x]` in Mathematica; $Q_\Gamma^{-1}(s, x)$ is implemented as `gammaincinv(x,s,'upper')` in Matlab, and as `InverseGammaRegularized[s,x]` in Mathematica. Note the order of x and s .

3.2 Probability of Misdetction

The probability that the frame is not detected, given hypothesis H_1 , i.e., frame is present, is called the probability of misdetction, p_{md} . Probability of detection p_{d} is complementary to it: $p_{\text{d}} = 1 - p_{\text{md}}$. For the event of detection, we should correctly assume the following events,

$$\mathcal{E}_1 := \exists n \in \{-N_h + 1, \dots, N_h - 1\} : y'_1[n] > r_{1,\text{thr}}, \quad (61)$$

$$\mathcal{E}_2 := \exists n \in \{-N_h + 1, \dots, N_h - 1\} : y'_2[n] > r_{2,\text{thr}}. \quad (62)$$

However, for simplicity of derivation, we evaluate the probability of the approximate events,

$$\mathcal{E}_{1,\text{approx}} := y'_1[0] > r_{1,\text{thr}}, \quad (63)$$

$$\mathcal{E}_{2,\text{approx}} := y'_2[0] > r_{2,\text{thr}}. \quad (64)$$

So, we derive an analytical expression for these probabilities, $p_{d,1} := \Pr\{\mathcal{E}_{1,\text{approx}}\}$, and $p_{d,2} := \Pr\{\mathcal{E}_{2,\text{approx}}\}$.

Again, we derive the probability distribution of auxiliary signals. First, we assign a shortcut to the frequency-modulated preamble and its parts,

$$h^\varepsilon[n] := h[n] e^{j2\pi F_\varepsilon n}, \quad (65)$$

$$h_\ell^\varepsilon[n] := h_\ell[n] e^{j2\pi F_\varepsilon n}. \quad (66)$$

$$x[n] = \eta e^{j\varphi} h^\varepsilon[n] + w[n] \sim \mathcal{CN}(\eta e^{j\varphi} h^\varepsilon[n], \sigma_w^2), \quad (67)$$

$$\begin{aligned} x_\ell[n] &= (x * h_{\ell,\text{mf}})[n] = \sum_k x[n-k] h_{\ell,\text{mf}}[k] \\ &= \eta e^{j\varphi} \sum_k h^\varepsilon[n-k] h_\ell^*[k] + \sum_k w[n-k] h_{\ell,\text{mf}}[k] \\ &= \eta e^{j\varphi} \mathcal{R}_{h^\varepsilon, h_\ell}[n] + \sum_k w[n-k] h_{\ell,\text{mf}}[k], \end{aligned} \quad (68)$$

$$x_\ell[n] \sim \mathcal{CN}(\eta e^{j\varphi} \mathcal{R}_{h^\varepsilon, h_\ell}[n], \|h_\ell\|^2 \sigma_w^2), \quad (69)$$

$$\begin{aligned} \mathcal{R}_{h^\varepsilon, h_\ell}[n] &= \sum_m h^\varepsilon[n+m] h_\ell^*[m] \\ &= \sum_m h[n+m] e^{j2\pi F_\varepsilon(n+m)} h_\ell^*[m] \\ &= \sum_{k=0}^{N_{\text{seq}}-1} e^{j2\pi F_\varepsilon k L_{\text{seq}}} \sum_m \left[h_k[n+m-k L_{\text{seq}}] \right. \\ &\quad \cdot \left. e^{j2\pi F_\varepsilon(n+m-k L_{\text{seq}})} h_\ell^*[m] \right] \\ &= \sum_{k=0}^{N_{\text{seq}}-1} e^{j2\pi F_\varepsilon L_{\text{seq}} k} \mathcal{R}_{h_k^\varepsilon, h_\ell}[n-k L_{\text{seq}}]. \end{aligned} \quad (70)$$

$$y_{\ell,1} = |x_\ell[n]|, \quad (72)$$

$$y_{\ell,1} \sim \text{Rice}\left(v[n] = \eta |\mathcal{R}_{h^\varepsilon, h_\ell}[n]|, \sigma = \sqrt{\frac{\|h_\ell\|^2 \sigma_w^2}{2}}\right), \quad (73)$$

$$y_{\ell,2} = |x_\ell[n]|^2, \quad (74)$$

$$y_{\ell,2} = \chi^2_2\left(v^2[n] = \eta^2 |\mathcal{R}_{h^\varepsilon, h_\ell}[n]|^2, \lambda_\ell = \frac{1}{\|h_\ell\|^2 \sigma_w^2}\right). \quad (75)$$

$\text{Rice}(v, \sigma)$ is the Rice distribution with parameters v and σ . $\chi^2_k(v^2, \lambda_0)$ is the non-central chi-squared distribution with k degrees of freedom and parameters v^2 and λ_0 . Traditional notation uses just one additional parameter λ : $\chi^2_k(\lambda)$. The relation between them is explained next.

$$A \sim \chi^2_k(\lambda), \quad \sigma^2 = \text{const} > 0, \quad (76)$$

$$B := \sigma^2 \cdot A \sim \chi^2_k(v^2, \lambda_0) \equiv \chi^2_k(v^2, \sigma^2), \quad (77)$$

$$\lambda_0 := \frac{1}{2\sigma^2}, \quad v^2 := \lambda \sigma^2. \quad (78)$$

The notation $\chi^2_k(v^2, \lambda_0)$ better reflects the process of generation of the RV. For more details, see Appendix 1.

Again, we assume identical energy in each subsequence h_ℓ : $E_{h_\ell} = E_{h_0}$.

$$y'_1[n] = \sum_{\ell=0}^{N_{\text{seq}}-1} y_{\ell,1}[n + \ell L_{\text{seq}}] \sim \text{Normal approx.} \quad (79)$$

$$\begin{aligned} y'_2[n] &= \sum_{\ell=0}^{N_{\text{seq}}-1} y_{\ell,2}[n + \ell L_{\text{seq}}], \\ &\sim \chi^2_{2N_{\text{seq}}}\left(v^2[n], \lambda_0 = \frac{1}{\|h_0\|^2 \sigma_w^2}\right), \end{aligned} \quad (80)$$

$$v^2[n] = \eta^2 \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{R}_{h^\varepsilon, h_\ell}[n + \ell L_{\text{seq}}]|^2, \quad \text{for } y'_2[n]. \quad (81)$$

$y'_1[n]$ will be approximated by a (real-valued) normal distribution (with matched mean and variance).

Assume $y'_1[0] \sim \mathcal{N}(\mu_{y'_1}, \sigma_{y'_1}^2)$. The probability of detection and misdetection is as follows.

$$\begin{aligned} p_{d,1} &= \Pr\{\mathcal{E}_{1,\text{approx}}\} = 1 - \Pr\{y'_1[0] \leq r_{1,\text{thr}}\} \\ &= 1 - \text{cdf}_{y'_1[0]}(r_{1,\text{thr}}), \end{aligned} \quad (82)$$

$$\begin{aligned} p_{\text{md},1} &= 1 - p_{d,1} = \text{cdf}_{y'_1[0]}(r_{1,\text{thr}}) \\ &= \Phi\left(\frac{r_{1,\text{thr}} - \mu_{y'_1}}{\sigma_{y'_1}}\right). \end{aligned} \quad (83)$$

$\Phi(x) = [1 + \text{erf}(x/\sqrt{2})]/2$ is the cdf of the Standard Normal distribution $\mathcal{N}(0, 1)$, and $\text{erf}(x) = 2/\sqrt{\pi} \int_0^x \exp(-t^2) dt$ is the error function.

$$\begin{aligned} p_{d,2} &= \Pr\{\mathcal{E}_{2,\text{approx}}\} = 1 - \Pr\{y'_2[0] \leq r_{2,\text{thr}}\} \\ &= 1 - \text{cdf}_{y'_2[0]}(r_{2,\text{thr}}), \end{aligned} \quad (84)$$

$$\begin{aligned} p_{\text{md},2} &= 1 - p_{d,2} = \text{cdf}_{y'_2[0]}(r_{2,\text{thr}}) \\ &= 1 - Q_{M, N_{\text{seq}}}(\sqrt{2\lambda_0 v^2}, \sqrt{2\lambda_0 r_{2,\text{thr}}}), \end{aligned} \quad (85)$$

$$\lambda_0 = \frac{1}{\|h_0\|^2 \sigma_w^2}, \quad (86)$$

$$v^2 := v^2[0] = \eta^2 \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{R}_{h^\varepsilon, h_\ell}[\ell L_{\text{seq}}]|^2. \quad (87)$$

$Q_{M,k}(a, b)$ is the Marcum Q-function of order k .

Now, we evaluate some auxiliary expressions involved in (83) and (85); namely $\mu_{y'_1}$, $\sigma_{y'_1}$, $v^2[0]$, and others that are involved in formulas for them.

Recall the range of non-zero values of $\mathcal{R}_{h_k^\varepsilon, h_\ell}[n]$.

$$\mathcal{R}_{h_k^\varepsilon, h_\ell}[n] = \begin{cases} \mathcal{R}_{h_k^\varepsilon, h_\ell}[n], & n \in \{-L_{\text{seq}} + 1, \dots, L_{\text{seq}} - 1\} \\ 0, & \text{otherwise.} \end{cases} \quad (88)$$

Define an auxiliary function $\mathcal{F}_{h_\ell}(F_\varepsilon)$, a frequency-dependent energy of $h_\ell[n]$.

$$\mathcal{F}_{h_\ell}(F_\varepsilon) := \mathcal{R}_{h_\ell^\varepsilon, h_\ell}[0] = \sum_k h_\ell[k] e^{j2\pi F_\varepsilon k} h_\ell^*[k]$$

$$= \sum_k |h_\ell[k]|^2 e^{j2\pi F_\varepsilon k} = \text{DtFT} \left[|h_\ell[k]|^2 \right] (-2\pi F_\varepsilon). \quad (89)$$

$\text{DtFT}[f[k]](\Omega)$ denotes Discrete-time Fourier transform of the sequence/signal $f[k]$:

$$\text{DtFT}[f[k]](\Omega) = \sum_{k \in \mathbb{Z}} f[k] e^{-j\Omega k}. \quad (90)$$

For continuous-time case, $\mathcal{F}_{h_\ell}(F_\varepsilon)$ becomes

$$\begin{aligned} \mathcal{F}_{h_\ell}(f_\varepsilon) &:= \mathcal{R}_{h_\ell^\varepsilon, h_\ell}(0) = \int_{\mathbb{R}} h_\ell(t) e^{j2\pi f_\varepsilon t} h_\ell^*(t) dt \\ &= \int_{\mathbb{R}} |h_\ell(t)|^2 e^{j2\pi f_\varepsilon t} dt = \text{FT} \left[|h_\ell(t)|^2 \right] (-2\pi f_\varepsilon). \end{aligned} \quad (91)$$

$\text{FT}[f(t)](\omega)$ denotes Fourier transform of the function $f(t)$:

$$\text{FT}[f(t)](\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt. \quad (92)$$

Now, we evaluate $\mathcal{R}_{h_k^\varepsilon, h_\ell}[n]$ in some multiples of L_{seq} . Assume $\ell \in \{0, \dots, N_{\text{seq}} - 1\}$.

$$\begin{aligned} \mathcal{R}_{h_k^\varepsilon, h_\ell}[\ell L_{\text{seq}}] &\stackrel{(71)}{=} \sum_{k=0}^{N_{\text{seq}}-1} e^{j2\pi F_\varepsilon L_{\text{seq}} k} \mathcal{R}_{h_k^\varepsilon, h_\ell}[(\ell - k)L_{\text{seq}}] \\ &= |\ell \stackrel{!}{=} k| = e^{j2\pi F_\varepsilon L_{\text{seq}} \ell} \mathcal{R}_{h_k^\varepsilon, h_\ell}[0] = e^{j2\pi F_\varepsilon L_{\text{seq}} \ell} \mathcal{F}_{h_\ell}(F_\varepsilon). \end{aligned} \quad (93)$$

Using formulas for the mean of a Rice RV $y_1[\ell L_{\text{seq}}]$, we get

$$\nu[\ell L_{\text{seq}}] = \eta |\mathcal{R}_{h_k^\varepsilon, h_\ell}[\ell L_{\text{seq}}]| = \eta |\mathcal{F}_{h_\ell}(F_\varepsilon)|, \quad (94)$$

$$\mu_{y_{\ell,1}}[\ell L_{\text{seq}}] = \sigma \sqrt{\frac{\pi}{2}} L_{1/2} \left(\frac{-\nu^2[\ell L_{\text{seq}}]}{2\sigma^2} \right), \quad (95)$$

$L_{1/2}(x)$ is the Laguerre polynomial of order 1/2.

$$L_{1/2}(x) = \exp\left(\frac{x}{2}\right) \left[(1-x) I_0\left(\frac{-x}{2}\right) - x I_1\left(\frac{-x}{2}\right) \right]. \quad (96)$$

And, the mean of $y'_1[0]$ is

$$\begin{aligned} \mu_{y'_1} &\equiv \mu_{y'_1}[0] = \sum_{\ell=0}^{N_{\text{seq}}-1} \mu_{y_{\ell,1}}[\ell L_{\text{seq}}] \\ &= \sigma \sqrt{\frac{\pi}{2}} \sum_{\ell=0}^{N_{\text{seq}}-1} L_{1/2} \left(\frac{-\eta^2 |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2}{2\sigma^2} \right). \end{aligned} \quad (97)$$

Now, the variance is derived. First, the variance of the Rice RV $y_1[\ell L_{\text{seq}}]$, then the variance of $y'_1[0]$ is expressed as the sum of individual variances of $y_1[\ell L_{\text{seq}}]$.

$$\text{Var}[y_{\ell,1}[\ell L_{\text{seq}}]] = 2\sigma^2 + \nu^2[\ell L_{\text{seq}}] - \mu_{y_{\ell,1}}^2[\ell L_{\text{seq}}], \quad (98)$$

$$\begin{aligned} \sigma_{y'_1}^2 &\equiv \text{Var}[y'_1[0]] = \sum_{\ell=0}^{N_{\text{seq}}-1} \text{Var}[y_{\ell,1}[0 + \ell L_{\text{seq}}]] \\ &= 2\sigma^2 N_{\text{seq}} + \eta^2 \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2 \\ &\quad - \sigma^2 \frac{\pi}{2} L_{1/2}^2 \left(\frac{-\eta^2}{2\sigma^2} |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2 \right). \end{aligned} \quad (99)$$

The transmitted signal is normalized to unit energy, so $E_{h_0} = 1/N_{\text{seq}}$. Using $\sigma^2 = \sigma_w^2 \|h_0\|^2/2$, and $r_{1,\text{thr}}$ in (56), with parameter b in (53), we adjust $p_{\text{md},1}$ in (83). The final expression for $p_{\text{md},1}$ is then (105)[bottom of this page].

Next, we look at $p_{\text{md},2}$; specifically at partial expressions involved in (85).

$$\begin{aligned} \nu^2[0] &= \sum_{\ell=0}^{N_{\text{seq}}-1} \eta^2 |\mathcal{R}_{h_k^\varepsilon, h_\ell}[\ell L_{\text{seq}}]|^2 \\ &= \eta^2 \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{F}_{h_\ell}(0)|^2, \end{aligned} \quad (100)$$

$$\lambda_0 = \frac{1}{\sigma_w^2 \|h_0\|^2}. \quad (101)$$

We adjust the first and the second argument of the Marcum Q-function in (85).

$$\sqrt{2\lambda_0} \nu^2[0] = \sqrt{2 \frac{N_{\text{seq}}}{\sigma_w^2}} \nu^2[0], \quad (102)$$

$$\begin{aligned} \sqrt{2\lambda_0} r_{2,\text{thr}} &= \sqrt{2 \frac{1}{\sigma_w^2 \|h_0\|^2} \|h_0\|^2 \sigma_w^2 Q_\Gamma^{-1}(N_{\text{seq}}, p_{\text{fa},2})} \\ &= \sqrt{2 Q_\Gamma^{-1}(N_{\text{seq}}, p_{\text{fa},2})}. \end{aligned} \quad (103)$$

The probability of misdetection is then

$$\begin{aligned} p_{\text{md},2} &= 1 - Q_{M, N_{\text{seq}}} \left(\sqrt{2 \frac{\eta^2 N_{\text{seq}}}{\sigma_w^2}} \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2, \right. \\ &\quad \left. \sqrt{2 Q_\Gamma^{-1}(N_{\text{seq}}, p_{\text{fa},2})} \right). \end{aligned} \quad (104)$$

$Q_{M,k}(a, b)$ is Marcum Q-function of order k . The formulas for evaluation of probability of misdetection $p_{\text{md},1}$: (105),

$$p_{\text{md},1} = \Phi \left(\frac{\sqrt{\left(\frac{\Gamma(N_{\text{seq}}+1)}{\Gamma(N_{\text{seq}}+\frac{1}{2}}) \right)^2 Q_\Gamma^{-1}(N_{\text{seq}}, p_{\text{fa},1}) - \sum_{\ell=0}^{N_{\text{seq}}-1} L_{1/2} \left(\frac{-N_{\text{seq}}}{\sigma_w^2} |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2 \right)}}{\sqrt{\frac{4}{\pi} N_{\text{seq}} + \frac{4N_{\text{seq}}}{\pi \sigma_w^2} \sum_{\ell=0}^{N_{\text{seq}}-1} |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2 - \sum_{\ell=0}^{N_{\text{seq}}-1} L_{1/2}^2 \left(\frac{-N_{\text{seq}}}{\sigma_w^2} |\mathcal{F}_{h_\ell}(F_\varepsilon)|^2 \right)}} \right). \quad (105)$$

and $p_{\text{md},2}$: (104), are implemented in the function `get_p_md` in [12].

For $N_{\text{seq}} = 1$, and $F_{\varepsilon} = 0$ both $p_{\text{md},1}$, and $p_{\text{md},2}$ reduce to the same expression,

$$p_{\text{md,ref}} = 1 - Q_{M,1} \left(\eta \sqrt{\frac{2E_h}{\sigma_w^2}}, \sqrt{2(-1) \ln(p_{\text{fa}})} \right). \quad (106)$$

We refer to the case (for $N_{\text{seq}} = 1$, and $F_{\varepsilon} = 0$) as the reference case. It corresponds to the standard approach with one matched filter and perfectly known F_{ε} .

3.3 Probabilities for the Detector that also Estimates a Reference Level

In the subsection, we shortly inspect the detector with reference-level estimators. First one estimates noise-variance from the received noise, the other two (mAv, mMed) are mentioned at the end of Sec. 2. Namely, we consider the three reference-level estimators:

1. Noise-variance estimator. The estimator should take samples of the received signal $x[n]$ well before a frame arrives.

$$\text{var}_w[n] := \sum_{k=1}^L |x[n-k]|^2. \quad (107)$$

The estimator outputs a sequence $\text{var}_w[n]$ that estimates current noise variance $\text{Var}[w[n]] = \sigma_w^2$.

2. Plateau-level estimators using moving average $y'_{1,\text{mAv}}[n]$, $y'_{2,\text{mAv}}[n]$. See (26), and (28).
3. Plateau-level estimators using moving median $y'_{1,\text{mMed}}[n]$, $y'_{2,\text{mMed}}[n]$. See (27), and (29).

The probability of misdetection is evaluated just numerically; we could not get an analytical formula. The noise-independent threshold for the required false-alarm probability is set numerically; for details, see function `set_thr_forGiven_pFa` in [12].

4. Numerical Results

The section presents simulation results. The transmitted signal/preamble $h[n]$ is normalized to unit energy $1 = \sum_n |h[n]|^2$. We set channel parameters $\eta = 1$, $\varphi = 0$, and $\text{SNR} := 1/\sigma_w^2$. In simulation, we have tested 10 different preambles summarized in Tab. 1.

N_{seq} was chosen from the set $N_{\text{seq}} \in \{1, 2, 3, 10\}$, $F_{\varepsilon} \in \{0, 1 \cdot 10^{-3}, 2 \cdot 10^{-3}, \dots, 10 \cdot 10^{-3}\}$; and $p_{\text{fa}} \in \{10^{-3}, 10^{-4}\}$. We have generated 120 000 test-frames and evaluated the probability of detection with the threshold set according to (56), (57), assuming the noise variance σ_w^2 is known. For all the simulated cases, analytical formulae (105), (104) agree with the simulation very well. We plot just 6 cases given by the tuple [preamble ID, N_{seq} , $F_{\varepsilon} \in \{0.002, 0.009\}$].

id:01	wifi (IEEE 802.11) Short Training Field (STF) preamble, $N_h = 160$ samples,
id:02	Zadoff-Chu (ZC) sequence, $\text{ZC}(N_h = 100, u = 1)$ (with parameter u),
id:03	$\text{ZC}(N_h = 100, u = 2)$,
id:04	complex-valued white Gaussian noise sequence $\text{PN}(N_h = 100)$,
id:05	$\text{ZC}(N_h = 200, u = 1)$,
id:06	$\text{ZC}(N_h = 200, u = 2)$,
id:07	$\text{PN}(N_h = 200)$,
id:08	$\text{ZC}(N_h = 400, u = 1)$,
id:09	$\text{ZC}(N_h = 400, u = 2)$,
id:10	$\text{PN}(N_h = 400)$,

Tab. 1. Preamble types used in the simulation.

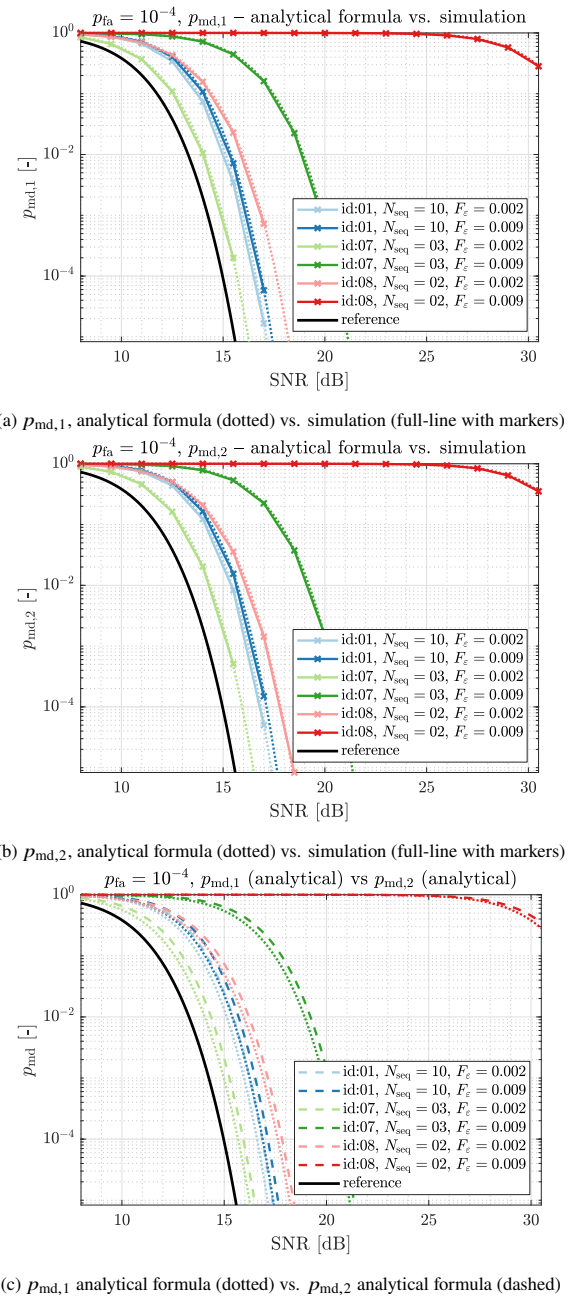


Fig. 6. Probability of misdetection. Comparison of analytical expression (no markers) and simulation results (with 'x' markers).

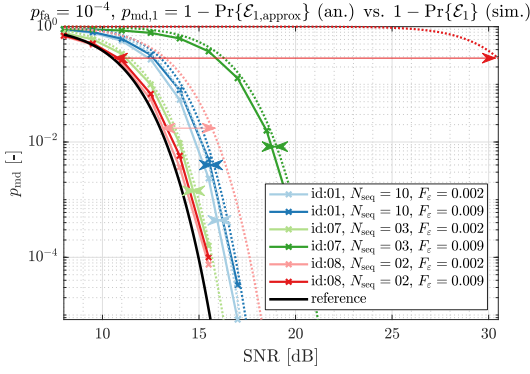
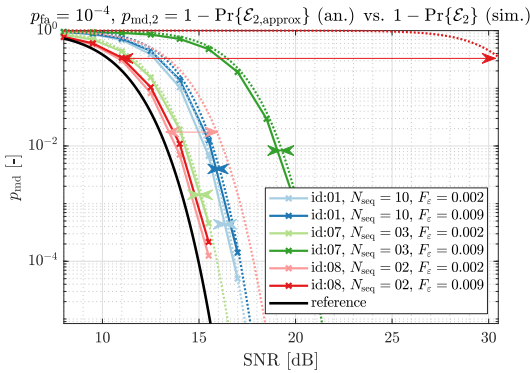
(a) $p_{md,1}$, analytical formula (dotted) vs. $1 - \Pr\{\mathcal{E}_1\}$ sim. (full-line with markers)(b) $p_{md,2}$, analytical formula (dotted) vs. $1 - \Pr\{\mathcal{E}_2\}$ sim. (full-line with markers)

Fig. 7. Probability of misdetection for the detection event $\mathcal{E}_{approx}$ vs. true $\mathcal{E}$. Comparison of analytical expression (no markers) and simulation results (with 'x' markers).

In all figures, lines for analytical expression don't use markers, whereas lines for simulated results use 'x' marker. The black full-line denotes the reference case – the standard approach that uses one matched filter, and the F_ϵ is perfectly known. The p_{md} of the reference case is given by 106.

- Figure 6(a) shows $p_{md,1}$, analytical vs. simulation results.
- Figure 6(b) shows $p_{md,2}$, analytical vs. simulation results.
- Figure 6(c) compares analytical results for $p_{md,1}$ (dotted) vs $p_{md,2}$ (dashed).
- Figure 7(a) compares analytical results for $p_{md,1} = 1 - \Pr\{\mathcal{E}_{1,approx}\}$ vs simulation result for $1 - \Pr\{\mathcal{E}_1\}$. See (63) vs (61).
- Figure 7(b) compares analytical results for $p_{md,2} = 1 - \Pr\{\mathcal{E}_{2,approx}\}$ vs simulation result for $1 - \Pr\{\mathcal{E}_2\}$. See (64) vs (62).

Figures 6(a) and 6(b) show the good match between the analytical expression and the simulation of p_{md} . From the Fig. 6(c) we see that $p_{md,1}$ performs better by about 0.25 dB with respect to $p_{md,2}$ in all simulated cases (dotted lines are left from the corresponding dashed ones).

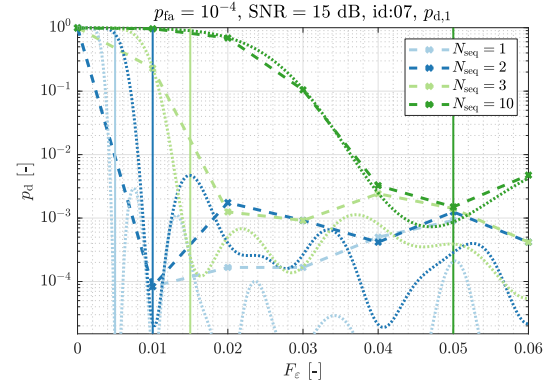
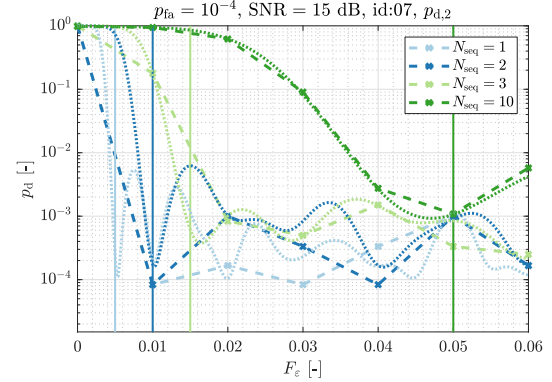
(a) $p_{d,1}$, analytical formula (dotted) vs. simulation (dashed with 'x' marker)(b) $p_{d,2}$, analytical formula (dotted) vs. simulation (dashed with 'x' marker)

Fig. 8. Probability of detection for the preamble id:07. Comparison of ideal-case when noise variance is known (analytical expression, no markers) and simulation (dashed with 'x' marker). p_d as a function of F_ϵ . Vertical lines indicate $F_\epsilon = N_{seq}/N_h = 1/L_{seq}$.

Figures 7(a), and 7(b) show that the true detector performance (detection events $\mathcal{E}_1, \mathcal{E}_2$ instead of $\mathcal{E}_{1,approx}, \mathcal{E}_{2,approx}$) is often significantly better than the one of detectors assuming the approximate detection events (for which we have the analytical expression). The difference between p_{md} for event $\mathcal{E}$, and p_{md} for event $\mathcal{E}_{approx}$ is highlighted by arrow. We point out that the detector just triggers the search of arg-max of the detector's metric in the near vicinity of the detection event. So, the good performance of the detector for events $\mathcal{E}_1$, and $\mathcal{E}_2$ does not guarantee it will correctly estimate the sample where the frame has started.

Figure 8 shows p_d (analytical vs. simulation) for the preamble id:07 as a function of F_ϵ , for SNR = 15 dB, and changing N_{seq} . The simulation used 12 000 test frames. Vertical lines indicate $F_\epsilon = N_{seq}/N_h = 1/L_{seq}$ – the approximate width of the main lobe (for constant-amplitude signals, the main lobe width is exact). For F_ϵ greater than the main lobe width, the analytical expression does not approximate the simulation results closely. Further, we notice that the analytical expression for $p_{d,2}$ in Fig. 8(b) does not decrease below the p_{fa} , whereas the analytical expression for $p_{d,1}$ in Fig. 8(a) does decrease below p_{fa} (worse approximation than for $p_{d,2}$). We see that the width of the main lobe of p_d widens as the N_{seq} increases.

Here, we emphasize the obtained result and interpret it for the continuous-time equivalent model. Denote the width of the main lobe as $F_{\varepsilon, \max}$. For constant-amplitude $h[n]$, $|\mathcal{F}_{h_0}(F_{\varepsilon})| = |\sin(\pi L_{\text{seq}} F_{\varepsilon}) / (N_{\text{seq}} \sin(\pi F_{\varepsilon}))|$. The formula has the first zero at $F_{\varepsilon, \max} = 1/L_{\text{seq}} = N_{\text{seq}}/N_h$ – that corresponds with our observation. Mark the length of continuous-time $h(t)$ as $T_h = N_h \cdot T_{\text{sa}}$. Then, continuous-time equivalent of $F_{\varepsilon, \max}$ is given as

$$f_{\varepsilon, \max} = F_{\varepsilon, \max} \cdot f_{\text{sa}} = \frac{N_{\text{seq}}}{N_h T_{\text{sa}}} = \frac{N_{\text{seq}}}{T_h}. \quad (108)$$

In practical systems designs, we can define an “allowable” width of the main lobe as a fraction of $f_{\varepsilon, \max}$. System specifications can require the maximal offset frequency that the detector should tolerate. The designer selects sufficiently high N_{seq} to ensure the required maximal offset frequency is within the “allowable” range.

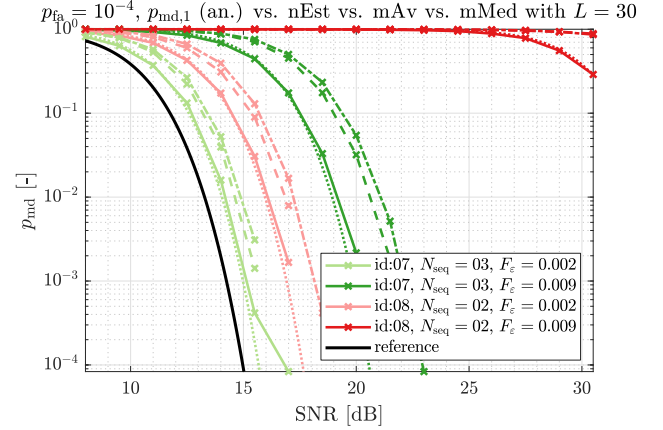
4.1 Detectors with a Reference-level Estimator

The subsection presents p_{md} of detectors that also estimate a reference level. The threshold of these detectors does depend just on p_{fa} , the explicit dependence on the current noise variance is removed.

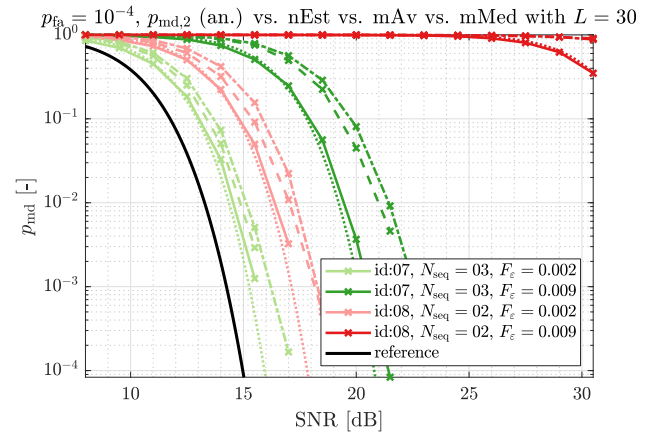
The detectors are mentioned in the Sec. 3.3. p_{md} of the detectors is shown in Fig. 9. Again, we show just a few cases specified by a tuple [preamble ID, N_{seq} , $F_{\varepsilon} \in \{0.002, 0.009\}$]. A different color differentiates the cases. For each case, there are four lines,

1. Dotted line (with no marker) – for the analytical result (noise variance is known).
2. Full line (with 'x' marker) – for the detector that estimates noise variance (nEst). Threshold is set using (56), and (57) using the current estimation of σ_w^2 .
3. Dashed line (with 'x' marker) – for the detector that uses moving average (mAv) to estimate a reference level. The threshold is set using the function `set_thr_forGiven_pFa` [12].
4. Dot-dashed line (with 'x' marker) – for the detector that uses moving median (mMed) to estimate a reference level. The threshold is set using the function `set_thr_forGiven_pFa` in [12].

The averaging length is set to $L = 30$ samples for all cases. 12 000 test frames have been used. We observe that the order of lines (in each case) is the same. Specifically, the order of lines from left (best) to right (worse) is: analytical, nEst, mAv, mMed. Another observation is that $L = 30$ is quite enough for nEst-detector to match with the ideal/analytical one – the full line is very close to the corresponding dotted one (analytical expression). In simulations, we have also tested $L = 10, 20, 40$. For $L = 30$ the 'nEst'-lines are within ~ 0.5 dB (for the selected F_{ε}) from the analytical expression (with exactly known noise variance).



(a) $p_{\text{md},1}$, analytical formula (dotted) vs. simulation nEst (full), mAv (dashed), mMed (dot-dashed)



(b) $p_{\text{md},2}$, analytical formula (dotted) vs. simulation nEst (full), mAv (dashed), mMed (dot-dashed)

Fig. 9. Probability of misdetection. Comparison of ideal case – when noise variance is known (analytical expression, no markers) – and detectors with a reference-level estimator (simulation results, 'x' marker). Averaging length $L = 30$.

4.2 Practical Over-the-air Tests

We have evaluated the detector metric $y'_2[n]$ on two over-the-air examples. One is the reception of the Global Positioning System (GPS), the other is the reception of a non-orthogonal superposition of 3 modified-wifi signals.

The GPS signal was received on AdaIM Pluto SDR [15] from Analog Devices. Sample rate was set to 10 megasamples per second (MSaps). The signal was then resampled to 10-times chip-rate ($f_{\text{chip}} = 1.023$ MHz). The Pseudo Random Noise (PRN) preambles were 10-times up-sampled to match the rate with the resampled received signal. We have analyzed the signal using classical correlation methods with trial frequency offsets to find the present satellites.

In Fig. 10, the classical metric with estimated F_{ε} is shown as the black (first) line (the metric values are scaled). The metric $y'_2[n]$ was evaluated for $N_{\text{seq}} = 10$, and $N_{\text{seq}} = 1$ and plotted as the red (second), and blue (third) line. The red line ($y'_2[n]$, with $N_{\text{seq}} = 10$) is maximal at the same time-index as the classical metric with previously estimated offset frequency.

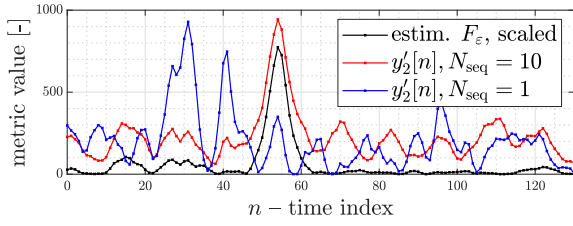


Fig. 10. Metric $y'_2[n]$ for a real over-the-air received GPS signal.

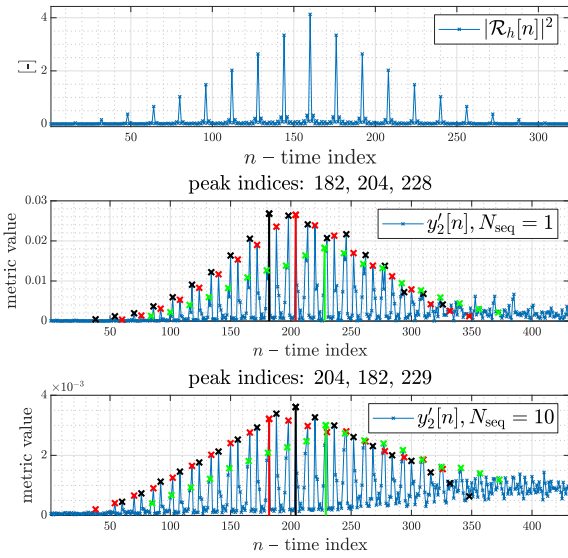


Fig. 11. Autocorrelation of the wifi preamble, and the metric $y'_2[n]$ for the three simultaneously transmitting sources.

In the second example, we transmitted a wifi signal (pregenerated in Matlab) using three Ettus N210 [16] with daughter board SBX-40. The three sources were not (externally) frequency-synchronized, the sample rate was set to 1 MHz (not 5, 10, or 20 MHz as in the wifi standard), and transmitted at the frequency 1.5 GHz. The signal was received using RTL SDR [17] (RTL2838UHIDIR, with tuner R820T from Realtek).

Figure 11 shows the squared magnitude of preamble autocorrelation $|R_h[n]|^2$ in the upper subfigure. The following (middle and lower) figures show the metric $y'_2[n]$ for $N_{\text{seq}} = 1$, and $N_{\text{seq}} = 10$ evaluated on the received signal. The metric is used to estimate the relative delays of the three sources. The indices of the peaks corresponding to the three sources are shown in the subfigure's title. The metric for $N_{\text{seq}} = 10$ should be more reliable. The example should demonstrate that the detector is also suitable for non-orthogonal multiple access scenarios.

5. Conclusion

We have proposed a frame-start detector resilient to frequency offset between the transmitter and the receiver. Although the idea is elementary – approximate absolute value of a sum by the sum of absolute values – the implications

are enormous. We can estimate frame-start alone, i.e., without jointly estimating it with the frequency offset. We have also derived a precise analytical expression for misdetection probability of the detector and compared it with simulation results. Further, we have given a simple design criterion for the selection of the necessary N_{seq} based on the maximal frequency offset that the system needs to manage. The detector has been tested in two over-the-air scenarios and has been shown to be useful. As a minor contribution, we have adjusted the formula for the sum of Rayleigh random variables (RVs) from [14].

Acknowledgments

The work has been supported by grants from the Czech Technical University (CTU) in Prague; grant numbers SGS24/059/OHK3/1T/13 and SGS25/055/OHK3/1T/13. And also from the Czech national TAČR project TN02000054.

References

- [1] CHOI, Z. Y., LEE, Y. H. Frame synchronization in the presence of frequency offset. *IEEE Transactions on Communications*, 2002, vol. 50, no. 7, p. 1062–1065. DOI: 10.1109/TCOMM.2002.800815
- [2] ZHOU, H., WANG, J., ZHANG, P., et al. A constant false alarm rate frame detector for satellite-based automatic identification system. *International Journal of Satellite Communications and Networking*, 2018, vol. 36, no. 1, p. 1–13. DOI: 10.1002/sat.1189
- [3] SCHMIDL, T., COX, D. Robust frequency and timing synchronization for OFDM. *IEEE Transactions on Communications*, 1997, vol. 45, no. 12, p. 1613–1621. DOI: 10.1109/26.650240
- [4] VAN DE BEEK, J., SANDELL, M., BORJESSON, P. ML estimation of time and frequency offset in OFDM systems. *IEEE Transactions on Signal Processing*, 1997, vol. 45, no. 7, p. 1800–1805. DOI: 10.1109/78.599949
- [5] LEE, D., CHEUN, K. Coarse symbol synchronization algorithms for OFDM systems in multipath channels. *IEEE Communications Letters*, 2002, vol. 6, no. 10, p. 446–448. DOI: 10.1109/LCOMM.2002.804247
- [6] LIU, G., WANG, J., BAN, T. Frame detection based on cyclic autocorrelation and constant false alarm rate in burst communication systems. *China Communications*, 2015, vol. 12, no. 5, p. 55–63. DOI: 10.1109/CC.2015.7112044
- [7] YING, W., LIU, X., CHEN, Z., et al. A modified frame synchronization algorithm based on average likelihood ratio testing. In *IEEE International Conference on Signal Processing (ICSP)*. Beijing (China), 2020, p. 417–422. DOI: 10.1109/ICSP48669.2020.9320932
- [8] CROCETTI, L., PAGANI, E., BERTOLUCCI, M., et al. Scalable hardware-efficient architecture for frame synchronization in high-data-rate satellite receivers. *Electronics*, 2024, vol. 13, no. 3, p. 1–20. DOI: 10.3390/electronics13030668
- [9] MARTINEZ, A. B., KUMAR, A., CHAFII, M., et al. Sequence design for frame detection based on autocorrelation. In *IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. Helsinki (Finland), 2021, p. 1–5. DOI: 10.1109/VTC2021-Spring51267.2021.9448669

- [10] PERELS, D., HAENE, S., BURG, A., et al. A frame-start detector for a 4x4 MIMO-OFDM system. In *IEEE International Conference on Acoustics Speed and Signal Processing Proceedings (ICASSP)*. Toulouse (France), 2006, p. 425–428. DOI: 10.1109/ICASSP.2006.1660996
- [11] KIMMER, M. *Physical Layer Testbed for Communication V2X Systems in 5.9GHz Band*. Czech Technical University in Prague, Master Thesis, 2024. [Online]. Available at: <https://dspace.cvut.cz/handle/10467/114801>
- [12] LUKAC, J. Matched-filter based frame-start detector resilient to frequency offset, Matlab-codes. *Zenodo*, 2025. DOI: 10.5281/zenodo.15775778
- [13] ZOU, J., XU, C. Frequency offset tolerant synchronization signal design in NB-IoT. *Sensors*, 2018, vol. 18, no. 11, p. 1–10. DOI: 10.3390/S18114077
- [14] HU, J., BEAULIEU, N. C. Accurate simple closed-form approximations to rayleigh sum distributions and densities. *IEEE Communications Letters*, 2005, vol. 9, no. 2, p. 109–111. DOI: 10.1109/LCOMM.2005.02003
- [15] ANALOG DEVICES. *ADALM-PLUTO Evaluation Board*. [Online]. Available at: <https://www.analog.com/en/resources/evaluation-hardware-and-software/evaluation-boards-kits/adalm-pluto.html>
- [16] ETTUS RESEARCH. *N200/N210 - Ettus Knowledge Base*. [Online]. Available at: <https://kb.ettus.com/N200/N210>
- [17] REALTEK, *SDR (Software Defined Radio) – RTL-SDR*. [Online]. Available at: <https://osmocom.org/projects/rtl-sdr/wiki/Rtl-sdr>

About the Authors ...

Jozef LUKAC (corresponding author) was born in Slovakia. He has received his M.Sc. from Electrical Engineering in 2019. He is currently at Ph.D. study at Czech Technical University (CTU) in Prague. His research interests include communication in dense radio networks, especially the non-orthogonal Multiple Access Channel (MAC) phase.

Michael KIMMER was born in Czechia. He has received his M.Sc. from Electrical Engineering in 2024 at CTU in Prague.

Jan SYKORA was born in Czechia. He has received his M.Sc. and Ph.D. from Electrical Engineering at CTU in Prague. Currently, he teaches several courses in coding, signal processing, theory of information and digital communication at the university.

Appendix A: Scaled Noncentral χ -squared Distribution

Here we define a scaled version of the non-central χ -squared distribution. Assume RVs $X_1, X_2, \dots, X_k$ are independent and distributed according to the normal distribution with possibly different means and unit variance, $X_i \sim \mathcal{N}(\mu_{X_i}, 1)$. Define RV A and a parameter λ as

$$A := \sum_{i=1}^k X_i^2, \quad \lambda := \sum_{i=1}^k \mu_{X_i}^2. \quad (\text{A1})$$

Then $A \sim \chi_k'^2(\lambda) - A$ is distributed according to the non-central χ -squared distribution with k degrees of freedom and the noncentrality parameter λ .

Assume a constant $\sigma^2 > 0$, define $B := \sigma^2 A$; and assign two auxiliary constants, λ_0 , and ν^2 .

$$\lambda_0 := \frac{1}{2\sigma^2}, \quad \nu^2 := \lambda\sigma^2, \quad (\text{A2})$$

$$\begin{aligned} \text{pdf}_B(\zeta) &= \frac{1}{\sigma^2} \text{pdf}_A\left(\frac{\zeta}{\sigma^2}\right), \\ &= \lambda_0 e^{-\lambda_0(\zeta + \nu^2)} \left(\frac{\zeta}{\nu^2}\right)^{\frac{k-2}{2}} I_{\frac{k}{2}-1}(2\lambda_0 \nu \sqrt{\zeta}) \mathbb{1}(\zeta), \end{aligned} \quad (\text{A3})$$

$$\text{cdf}_B(\zeta) = 1 - Q_{M, \frac{k}{2}}\left(\frac{\nu}{\sigma}, \frac{\sqrt{\zeta}}{\sigma}\right). \quad (\text{A4})$$

We refer to the distribution of B as $B \sim \chi_k'^2(\nu^2, \lambda_0)$ or $B \sim \chi_k'^2(\nu^2, \sigma^2)$. The RV B can be seen as a sum of k squared normal RVs with variance σ^2 :

$$Y_i := \sigma X_i, \quad (\text{A5})$$

$$\mu_{Y_i} := \sigma \cdot \mu_{X_i}, \quad \text{Var}[Y_i] = \sigma^2 \cdot \text{Var}[X_i], \quad (\text{A6})$$

$$Y_i \sim \mathcal{N}(\mu_{Y_i}, \sigma^2), \quad (\text{A7})$$

$$B = \sum_{i=1}^k Y_i^2 = \sum_{i=1}^k (\sigma X_i)^2 = \sigma^2 A, \quad (\text{A8})$$

$$\nu^2 = \sum_{i=1}^k \mu_{Y_i}^2 = \sigma^2 \lambda. \quad (\text{A9})$$

For $k = 2\ell$, $\ell \in \mathbb{N}$, B can be seen as a sum of ℓ abs-squared circularly symmetric complex-normal RVs:

$$Z_i \sim \mathcal{CN}(\mu_{Z_i}, \underbrace{2\sigma^2}_{1/\lambda_0}), \quad (\text{A10})$$

$$\text{Re}\{Z_i\} \sim \mathcal{N}(\text{Re}\{\mu_{Z_i}\}, \sigma^2), \quad (\text{A11})$$

$$\text{Im}\{Z_i\} \sim \mathcal{N}(\text{Im}\{\mu_{Z_i}\}, \sigma^2), \quad (\text{A12})$$

$$B = \sum_{i=1}^{\ell} |Z_i|^2 = \sum_{i=1}^{\ell} \text{Re}\{Z_i\}^2 + \sum_{i=1}^{\ell} \text{Im}\{Z_i\}^2, \quad (\text{A13})$$

$$\nu^2 = \sum_{i=1}^{\ell} \text{Re}\{\mu_{Z_i}\}^2 + \sum_{i=1}^{\ell} \text{Im}\{\mu_{Z_i}\}^2 = \sum_{i=1}^{\ell} |\mu_{Z_i}|^2, \quad (\text{A14})$$

$$\text{pdf}_B(\zeta) = \lambda_0 e^{\lambda_0(\zeta + \nu^2)} \left(\frac{\zeta}{\nu^2}\right)^{\frac{\ell-1}{2}} I_{\ell-1}(2\lambda_0 \nu \sqrt{\zeta}) \mathbb{1}(\zeta). \quad (\text{A15})$$