A highly efficient receiver for satellite-based automatic identification system signal detection

INTERNATIONAL JOURNAL OF SATELLITE COMMUNICATIONS AND NETWORKING
Int. J. Satell. Commun. Network. 2014; **00**:1–15
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sat

A Highly Efficient Receiver for Satellite-Based Automatic Identification System Signal Detection‡‡

Giulio Colavolpe[1]*, Tommaso Foggi[2], Alessandro Ugolini[1], Juan Lizarraga[3], Stefano Cioni[3] and Alberto Ginesi[3]

[1]*DII - University of Parma, Viale delle Scienze 181/A, 43124 Parma, Italy*
[2]*CNIT, Research Unit of Parma, Viale delle Scienze 181/A, 43124 Parma, Italy*
[3]*ESA-ESTEC, Noordwijk, The Netherlands*

SUMMARY

An innovative receiver architecture for the satellite-based Automatic Identification System (AIS) has been recently proposed. In this paper, we describe a few modifications that can be introduced on the algorithms for synchronization and detection, that provide an impressive performance improvement. The receiver architecture has been designed for an on-board implementation and a prototype has been implemented by the University of Parma and CGS S.p.A. Compagnia Generale per lo Spazio under the ESA project FENICE. A few modifications are also here described that could allow a further performance improvement in case of processing moved to ground-based stations, based on a-priori information there available. Copyright © 2014 John Wiley & Sons, Ltd.

Received . . .

KEY WORDS: Satellite, LEO, AIS, Maritime, Demodulation, Detection.

1. INTRODUCTION

This paper describes an advanced receiver for the reception of Automatic Identification System (AIS) [1] signals from a constellation of Low Earth Orbit (LEO) satellites. A patent application [2] has been filed at the European Patent Office for the receiver described in this paper.

The AIS communication system has been developed to ensure an efficient exchange of information between vessels and shore stations. The exchanged data include information about the identification of each ship, its position, speed, course and other status information. Each station equipped with an AIS transmitter periodically sends short packets that can be received by every AIS receiver within a coverage area of about 40 nautical miles around the transmitter. Each receiver is then able to build a local map of maritime traffic and thus avoid collisions at sea. A system with these characteristics has important ship-monitoring potentialities, with applications to security and search and rescue operations. However, there is a significant limitation caused by the range of coverage: with this system, it is impossible to receive messages from vessels at distances higher than 40 nautical miles from the shore.

*Correspondence to: Giulio Colavolpe, DII - University of Parma, Viale delle Scienze 181/A, 43124 Parma, Italy.
E-mail: giulio@unipr.it
‡‡This work has been conducted by a team under management of CGS S.p.A. and co-funded by the European Space Agency (ESA contract number 4000101838/10/NL/CLP). This paper will be presented in part at the 7th Advanced Satellite Mobile Systems Conference, 13th International Workshop on Signal Processing for Space Communications (ASMS/SPSC 2014), Livorno, Italy, September 2014.

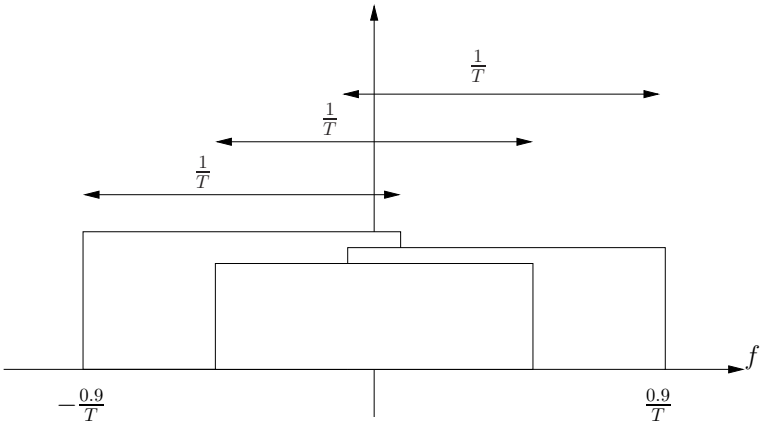

Figure 1. Frequency bandwidths of the three zonal demodulators.

For this reason, there has been recently an increasing interest in systems able to receive messages from ships far from the coastline. One way to achieve this goal is to receive messages from a constellation of LEO satellites, ensuring in this way a global coverage. The reader can refer to [3, 4] for more information and for a description of the system architecture. However, the reception of AIS signals from satellites has to face several issues that are not present in terrestrial AIS. In fact, the standard foresees that vessels operating within the same coverage area transmit according to a self-organized time-division multiple access (SOTDMA) protocol, to avoid message collisions. A satellite orbiting with altitudes ranging from 600 to 1000 km has a field of view that can cover multiple SOTDMA cells, so the probability of message collisions increases.

Some works have already addressed the problem of satellite-based AIS reception. For instance, the receiver described in [4] is based on three zonal demodulators processing different (but overlapping) frequency bandwidths (see Fig. 1 for a schematic representation of the three bandwidths), in order to increase the system diversity. The receiver described in this paper has the same architecture as that in [4]; however, we have obtained an impressive performance improvement by replacing the algorithms for synchronization, detection, and post processing with new ones. The complexity of the new synchronization algorithms is practically the same as those in [4], whereas the complexity of the added post processing is negligible. The detection algorithm here considered has, instead, a much larger complexity but is compatible with an on-board implementation [3]. A prototype has, in fact, been implemented by CGS SpA Compagnia Generale per lo Spazio in the framework of a project cofunded by the European Space Agency. Throughout the paper, we will compare our proposed receiver with that of [4], and we will highlight the differences and the improvements we have introduced into the system.

As far as the detection algorithm is concerned, another solution has been recently proposed in [5]. In that paper, synchronization aspects are not considered and a full-complexity Viterbi-based decoder, taking into account the characteristics of the modulation and the cyclic redundancy check (CRC) present in the standard, is employed. The authors have also extended this solution to include information on some known bits of the data field in [6]. This solution, while it can be considered as a performance limit, has a huge complexity and is certainly out of reach for an on-board implementation, since the Viterbi decoder operates on a trellis of 2^{18} states. In addition, these receivers have the serious flaw that they will always output a valid codeword, hence loosing the error detection capability typical of CRCs, and the only way to establish whether a message is correct or not would be to actually process the data included in the message itself, compare it with a database of information on previously received packets, and see if the new message is coherent with some previous data. Our aim in this paper is to propose an algorithm that has both a good performance and a complexity that makes it suitable for a practical implementation on board.

The remaining of the paper is organized as follows. Section 2 describes the signal model considered throughout the paper. Section 3 details the proposed receiver and Section 4 proposes

DOI: 10.1002/sat

a way to further improve the performance by moving the processing from the satellite to a ground station. Section 5 presents some numerical results in terms of packet error rate and, finally, Section 6 concludes the paper.

2. SYSTEM MODEL AND NOTATION

The complex envelope of a CPM signal can be written as [7]

$$s(t,\boldsymbol{\alpha}) = \sqrt{\frac{2E_S}{T}}\,\exp\left\{j2\pi h\sum_{n=0}^{N-1}\alpha_n q(t-nT)\right\} \tag{1}$$

where E_S is the energy per information symbol, T is the symbol interval, h is the *modulation index*, N is the number of transmitted information symbols, $\boldsymbol{\alpha} = \{\alpha_n\}_{n=0}^{N-1}$ is the information sequence, and $q(t)$ is the *phase pulse*, constrained to be such that

$$q(t) = \begin{cases} 0 & t \le 0 \\ \frac{1}{2} & t \ge LT, \end{cases}$$

L being the *correlation length*. Several examples of commonly used phase pulses are reported in [7].

The modulation index is usually written as $h = r/p$ (where r and p are relatively prime integers), and the information symbols belong to the M-ary alphabet $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$, M being a power of two. In this case, it can be shown [8] that the CPM signal in the generic time interval $[nT, (n+1)T]$ is completely defined by the symbol α_n, the *correlative state*

$$\omega_n = (\alpha_{n-1}, \alpha_{n-2}, \ldots, \alpha_{n-L+1})$$

and the *phase state* ϕ_n, which can be recursively defined as

$$\phi_n = [\phi_{n-1} + \pi h \alpha_{n-L}]_{2\pi}\ , \tag{2}$$

where $[\cdot]_{2\pi}$ denotes the "modulo 2π" operator. In other words, we may express the complex envelope of a CPM signal as (*Rimoldi decomposition*)

$$s(t,\boldsymbol{\alpha}) = \sqrt{\frac{2E_S}{T}}\sum_{n=0}^{N-1} s_T(t-nT;\alpha_n,\omega_n)e^{j\phi_n}$$

where $s_T(t-nT;\alpha_n,\omega_n)$ is a slice of signal of length T (with support in $[nT,(n+1)T]$) whose shape only depends on symbol α_n and correlative state ω_n and is independent of the considered symbol interval. For the initialization of recursion (2), we adopt the following conventions

$$\begin{aligned} \phi_0 &= 0 \\ \alpha_n &= 0 \quad \forall n < 0\,. \end{aligned}$$

At any given time epoch n, the correlative state ω_n can assume M^{L-1} different values, while the phase state ϕ_n can assume p different values [8], so that the CPM signal can be described by means of a finite-state machine with pM^{L-1} possible values of the state $\sigma_n = (\omega_n, \phi_n)$. When n is even, the p values assumed by the phase state ϕ_n belong to the alphabet $\mathcal{A}_e = \{2\pi hm\,,\ m = 0,1,\ldots,p-1\}$, while, when n is odd, belong to the alphabet $\mathcal{A}_o = \{2\pi hm + \pi h\,,\ m = 0,1,\ldots,p-1\}$.[†] In the remaining parts of the paper, we will adopt the following integer representation for the phase state

[†]When r is even, $\mathcal{A}_o$ and $\mathcal{A}_e$ coincide.

DOI: 10.1002/sat

and information symbols

$$\alpha_n = 2\bar{\alpha}_n - (M-1)$$
$$\phi_n = -\pi h(M-1)n + 2\pi h\bar{\phi}_n$$

so that $\bar{\alpha}_n \in \{0, 1, \ldots, M-1\}$ and $\bar{\phi}_n \in \{0, 1, \ldots, p-1\}$. The integer $\bar{\phi}_n$ can be recursively updated as follows

$$\bar{\phi}_n = [\bar{\phi}_{n-1} + \bar{\alpha}_n]_p \, .$$

Based on *Laurent decomposition*, the complex envelope of a CPM signal (1) may be exactly expressed as [9]

$$s(t, \boldsymbol{\alpha}) = \sum_{k=0}^{F-1} \sum_{n} a_{k,n} p_k(t - nT) \tag{3}$$

where $F = (M-1)2^{(L-1)\log_2 M}$ is the number of linearly modulated pulses $\{p_k(t)\}$, and $\{a_{k,n}\}$ are the so-called pseudo-symbols (hereafter, simply referred to as symbols). The expressions of pulses $\{p_k(t)\}$ and those of symbols $\{a_{k,n}\}$ as a function of the modulation parameters and of the information symbols $\{\alpha_n\}$ can be found in [9]. By truncating the summation in (3) to the first $K < F$ terms, we obtain the approximation

$$s(t, \boldsymbol{\alpha}) \simeq \sum_{k=0}^{K-1} \sum_{n} a_{k,n} p_k(t - nT) \, . \tag{4}$$

Most of the signal power is concentrated in the first $M-1$ components, i.e., those associated with the pulses $\{p_k(t)\}$ with $0 \le k \le M-2$, which are denoted as *principal components* [9]. As a consequence, a value of $K = M-1$ may be used in (4) to attain a very good tradeoff between approximation quality and number of signal components [10]. A nice feature of the principal components is that their symbols $\{a_{k,n}\}_{k=0}^{M-2}$ can be expressed as a function of α_n and $a_{0,n-1}$ only [9].

The Gaussian minimum shift keying (GMSK) modulation format [11] is a binary CPM (hence, $M = 2$, $\alpha_n \in \{\pm 1\}$, and $E_S = E_b$, where E_b is the energy per information bit) with modulation index $h = 1/2$ and phase pulse mathematically described in [11]. The derivative of this phase pulse can be obtained by filtering a rectangular pulse of length T with a Gaussian filter of proper -3 dB-bandwidth B. In the case of AIS, the value of B normalized to the symbol rate is $BT = 0.4 \div 0.5$. Although in this case the correlation length is in principle unlimited, $L = 2 \div 3$ can be assumed (for our simulations we used $L = 3$, but no appreciable difference has been observed with $L = 2$).

Considering now the Laurent representation of a GMSK signal, in this case there is only $M - 1 = 1$ principal component and (4) reads

$$s(t, \boldsymbol{\alpha}) \simeq \sum_{n} a_{0,n} p_0(t - nT) \, . \tag{5}$$

To simplify the notation, in the following we will use the following definitions:

$$\begin{aligned} a_n &= a_{0,n} \\ p(t) &= p_0(t) \, . \end{aligned}$$

Symbol a_n can be recursively computed as [9]

$$a_n = j\alpha_n a_{n-1}$$

and is related to the phase state ϕ_n by the following equation

$$a_{n-1} = e^{j\phi_n} \, .$$

DOI: 10.1002/sat

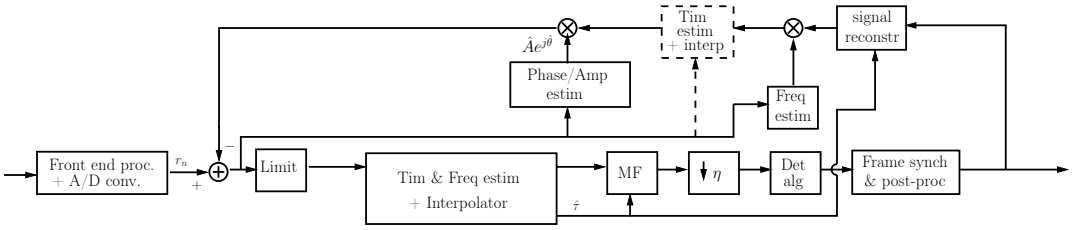

Figure 2. Zonal demodulator.

3. RECEIVER STRUCTURE

The receiver architecture described in this paper has also been considered in [4], and is composed of three zonal demodulators processing different (but overlapping) frequency bandwidths (see Fig. 2 for the block diagram of a zonal demodulator). As described in [4], the signal received from the VHF antenna is first processed by an analog front end and then is subject to an A/D conversion. The resulting discrete-time signal is then properly shifted in frequency and sent to the three zonal demodulators.[‡] For a detailed description of the receiver architecture, the reader can refer to [4].

Let us now consider a single zonal demodulator. First of all, timing and frequency estimations are performed on the received samples in order to provide the estimates necessary for detection. The estimators work on a window of L_0 symbols and $\eta = 3$ samples per symbol are used. As mentioned, to exploit the frequency diversity given by the Doppler spread, the receiver consists of 3 zonal demodulators, each of which is specifically designed to process only one slice of the AIS channel and to achieve the target performance within that slice. Since the estimation range of the frequency estimator is slightly less than $\pm\frac{0.21}{T}$ and taking into account that, due to a maximum Doppler shift of ± 4 kHz and a maximum frequency uncertainty of transmit and receive oscillators of ± 1.8 kHz, the maximum value of the frequency uncertainty is ± 5.8 kHz$= \pm\frac{0.604}{T}$, it is suggested to center the zonal demodulators at the nominal frequency, at the nominal frequency $+\frac{0.4}{T}$ and at the nominal frequency $-\frac{0.4}{T}$ (see Fig. 1). The next sections will provide a detailed analysis of the components of the zonal demodulator, whose block scheme is represented in Fig. 2.

Each of the three zonal demodulators included in the receiver is composed of four main sub-blocks properly interconnected.

1. *The pre-detection synchronization unit*. This unit performs a preliminary estimation of all channel parameters that need to be compensated before detection. The accuracy of these estimates must be higher than the sensitivity of the detection algorithm to an uncompensated error.
2. *The detection algorithm*.
3. *The post-processing unit*. This unit exploits CRC for frame synchronization and error correction.
4. *The post-detection synchronization unit and the digital re-modulator*. This unit performs a fine estimation of the channel parameters needed for the reconstruction and cancellation of the detected signal. In general, the estimation accuracy must be greater than that of the pre-detection synchronization unit.

In the following sections, we will describe the algorithms employed in these four units.

[‡]Instead of a parallel implementation, to reduce the hardware complexity, the same zonal demodulator can be reused. Obviously, the latency will increase.

DOI: 10.1002/sat

3.1. Pre-Detection Synchronization

Aim of this first synchronization stage is to estimate, in a non-data-aided (NDA) mode,[§] and compensate the frequency offset F and the timing offset τ that affect the received signal, whose complex envelope $r(t)$ is modeled by the following equation

$$r(t) = As(t-\tau, \boldsymbol{\alpha}) \exp\{j\theta\} \exp\{j2\pi Ft\} + w(t) \tag{6}$$

where a constant amplitude A, a constant phase offset θ, and a complex additive white Gaussian noise (AWGN) process $w(t)$ with two-sided power spectral density $2N_0$, modeling the noise baseband equivalent, are also accounted for. Note that the interfering users are not included in the model (6), since the interference has been neglected in the receiver design. On the other hand, the impact of the interferers on the receiver performance can be evaluated by means of extensive computer simulations, and we will show that the proposed receiver can perform better than that proposed in [4] in both presence and absence of interfering signals. We also point out that the pre-detection synchronization stage does not attempt to recover the phase offset θ, possibly time-varying, since the selected detection algorithm, described in the next section, can manage the presence of this phase uncertainty (see the relevant description).

The timing and carrier synchronization functions are carried out over a window of only L_0 symbols (bits) of the SOTDMA slot. The reason for this is that the current message alignment is not known at this point, so the synchronization blocks have to be activated on the portion of the slot where transmitted messages would have energy for sure. It turns out that, given the maximum differential delay between messages in the coverage area, only $L_0 = 128$ symbols of the slot can be used when packets of length 224 bits are considered, as those in the AIS 1 and 2 channels. When the shorter packets of length 152 bits, as those in the AIS 3 channels, are considered, it can be assumed $L_0 = 88$.

We empirically found that, in the presence of interference, a performance improvement can be obtained when samples r_n, after cancellation of the previously detected signal, are normalized to unit amplitude. We will say that a *limiter* is applied to the received samples. The advantage of such a transformation in the presence of interference is much higher than the performance degradation resulting in the absence of interference.

Let us now briefly describe the algorithms we have adopted for the pre-detection synchronization stage. The main building block is the synchronization algorithm proposed in [12], and is a modification of that originally selected in [4]. After the limiter, the received samples $\{r_n\}$ are filtered by means of a low-pass filter (LPF), implemented through a finite impulse response (FIR) filter with a limited number of coefficients, having bandwidth B_{LP}, which is a design parameter of the synchronization algorithm. Let $\{z_n\}$ be such filtered samples, indexed from $n = 0$ to $n = \eta L_0 - 1$, which correspond to L_0 signaling intervals. Next, the following coefficients are computed

$$\hat{R}_m(i) = \frac{1}{L_0 - m} \sum_{n=m}^{L_0-1} \left[z_{n\eta+i} z^*_{(n-m)\eta+i} \right]^2$$

for $i \in \{0, 1, \ldots, \eta-1\}$ and $m \in \{1, 2, \ldots, M_{\mathrm{t}}\}$, M_{t} being a design parameter of the synchronization algorithm (we selected a value of $M_{\mathrm{t}} = 20$). The estimate $\hat{\tau}$ of the time offset is then computed, using only the even autocorrelation terms, as

$$\hat{\tau} = -\frac{T}{2\pi} \arg \left\{ \sum_{i=0}^{\eta-1} \sum_{\substack{m=1 \\ m \text{ even}}}^{M_{\mathrm{t}}} A_1(m) |\hat{R}_m(i)| \exp\{-j2\pi i/\eta\} \right\} \tag{7}$$

where the terms $\{A_1(m)\}$ are real coefficients that can be precomputed off-line, based only on the modulation format, as explained in [12]. The samples at the limiter output are then interpolated,

[§]Data-aided solutions do not seem to be viable, because of the very low number of known symbols in the transmitted sequence.

DOI: 10.1002/sat

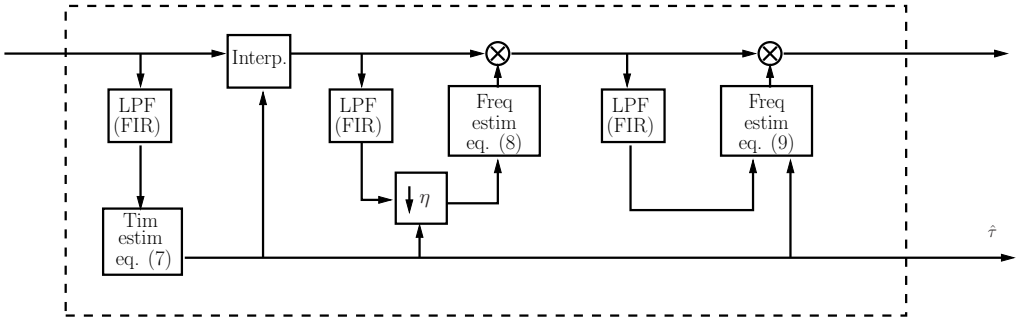

Figure 3. Pre-detection synchronization algorithm.

filtered again, and downsampled, obtaining samples $\{y_n\}$. These latter samples are then employed for frequency estimation by using the algorithm described in [13]. First, the following coefficients are computed

$$R(m) = \frac{1}{L_0 - m} \sum_{n=m}^{L_0-1} v_n v_{n-m}^*$$

for $m \in \{1, 2, \ldots, M_\mathrm{f}\}$, where M_f is a design parameter not greater than $L_0/2$ (we selected $M_\mathrm{f} = L_0/2$), and $v_n = (-1)^n y_n^2$. Then, the first estimate $\hat{F}_1$ of the frequency offset can be expressed as

$$\hat{F}_1 = \frac{1}{4\pi M_\mathrm{f} T} \sum_{m=1}^{M_\mathrm{f}} w(m) \left[\arg\{R(m)\} - \arg\{R(m-1)\}\right]_{2\pi} \tag{8}$$

where

$$w(m) = \frac{3\left[(L_0 - m)(L_0 - m + 1) - M_\mathrm{f}(L_0 - M_\mathrm{f})\right]}{M_\mathrm{f}\left(4M_\mathrm{f}^2 - 6M_\mathrm{f}L_0 + 3L_0^2 - 1\right)}.$$

Finally, a fine frequency estimation is performed by using the algorithm in [12], again adopting only the even autocorrelation terms. Hence, the final estimate $\hat{F}_2$ of the frequency offset is computed as

$$\hat{F}_2 = \frac{1}{4\pi M_\mathrm{t} T} \sum_{\substack{m=1 \\ m \text{ even}}}^{M_\mathrm{t}} \arg\left\{\mu_m \hat{R}_m(i_m) \hat{R}_{m-1}^*(i_{m-1})\right\} \tag{9}$$

where the terms $\{\mu_m\}$ are again real coefficients that can be precomputed off-line [12]—for the modulation format of interest here, we obtain $\mu_m = -1$ for all values of m. In (9), $\hat{R}_0(\cdot)$ is conventionally equal to one, while the terms $\{i_m\}$ are computed as explained in [12], or by simply quantizing the value of $\hat{\tau}/T$ to the closest integer (modulo η). The block diagram of the algorithm is shown in Fig. 3.

With respect to the algorithm described in [4], there are some major differences that allow improving the estimation performance. First of all, the original algorithm operates on the samples before the limiter, employs 10 (even and odd) autocorrelation terms instead of 20 (even) terms used now, and finally it performs only one instance of frequency estimation (i.e., the computation of $\hat{F}_1$ is not performed). The use of two instances of frequency estimations is useful to avoid a bias problem of the frequency estimator that reduces the estimation range. On the other hand, the use of the even autocorrelation terms only and of a different algorithm to compute $\hat{F}_1$, allows to solve a bias problem that appears when a long sequence of zeros in the data field is present (corresponding to particular locations and speeds of the ships).¶

¶Note that the standard [1] regrettably does not foresee a physical layer scrambling process of the transmitted sequence of symbols.

DOI: 10.1002/sat

3.2. Detection Algorithm

In this section, we detail the detection algorithm employed in the new receiver, that has replaced the Viterbi-based detector in [4]. For the selected detection algorithm, the knowledge of the phase shift introduced by the channel is not necessary, since the detector performs an implicit phase estimation. It is based on the Laurent decomposition (5) and was originally proposed in [14].

The received signal, after frequency and timing estimation and compensation, is filtered by means of an oversampled filter matched to the principal pulse of the Laurent decomposition. One sample per symbol interval is retained at the output of the matched filter (MF) using the information provided by the timing synchronizer. We will denote by x_n the sample at discrete-time n. The channel phase is quantized to the Q values of the alphabet $\Psi = \{0, \frac{2\pi}{Q}, \ldots, \frac{2\pi}{Q}(Q-1)\}$, Q being a design parameter to trade performance against complexity (in our simulations we used $Q = 24$). The channel phase probability density function becomes a probability mass function (PMF) and we will denote by $P_{f,n}(\psi_n)$ and $P_{b,n}(\psi_n)$ the estimates of the channel phase PMF in the forward and backward recursion, respectively. We report here the expression of the forward recursion [14].[‖]

$$\begin{aligned}
P_{f,n}(\psi_n) =& H_n(\psi_n)\,[(1-P_\Delta)P\{\alpha_n=-1\}P_{f,n-1}(\psi_n) \\
&+ (1-P_\Delta)P\{\alpha_n=1\}P_{f,n-1}\,(\psi_n+\pi) \\
&+ \frac{P_\Delta}{2}P\{\alpha_n=-1\}P_{f,n-1}\left(\psi_n+\frac{2\pi}{Q}\right) \\
&+ \frac{P_\Delta}{2}P\{\alpha_n=-1\}P_{f,n-1}\left(\psi_n-\frac{2\pi}{Q}\right) \\
&+ \frac{P_\Delta}{2}P\{\alpha_n=1\}P_{f,n-1}\left(\psi_n+\frac{2\pi}{Q}+\pi\right) \\
&+ \frac{P_\Delta}{2}P\{\alpha_n=1\}P_{f,n-1}\left(\psi_n-\frac{2\pi}{Q}+\pi\right)\Bigg]
\end{aligned}$$

where $0 < P_\Delta < 1$ is a design parameter, optimized depending on the speed of variation of the channel phase,[**]

$$H_n(\psi_n) = \exp\left\{\Re\left[\frac{1}{N_0}x_n e^{j\pi h(n+1)}e^{-j\psi_n}\right]\right\} \tag{10}$$

and $P\{\alpha_n=-1\}$, $P\{\alpha_n=1\}$ are the a-priori probabilities of the symbols. In our simulations, we have set the a-priori probabilities to 0.5, but in a case in which some symbols of the transmitted message are known at the receiver, it is possible to significantly improve the performance of the detection algorithm by including in the detection process the a-priori probabilities of the known symbols. The PMF computed during the forward and backward recursions are employed in the final completion giving the symbol a-posteriori probabilities (APPs):

$$\begin{aligned}
P(\alpha_n|\mathbf{x}) = & \sum_{\psi_n\in\Psi} P_{b,n}(\psi_n) \\
& [(1-P_\Delta)P(\bar{\alpha}_n)P_{f,n-1}(\psi_n-\pi\bar{\alpha}_n) \\
& + \frac{P_\Delta}{2}P(\bar{\alpha}_n)P_{f,n-1}\left(\psi_n-\pi\bar{\alpha}_n+\frac{2\pi}{Q}\right) \\
& + \frac{P_\Delta}{2}P(\bar{\alpha}_n)P_{f,n-1}\left(\psi_n-\pi\bar{\alpha}_n-\frac{2\pi}{Q}\right)\Bigg].
\end{aligned}$$

From the APPs $P(\alpha_n=1|\mathbf{x})$ and $P(\alpha_n=-1|\mathbf{x})$, the logarithmic likelihood ratio (LLR)

$$L_n = \ln\frac{P\{\alpha_n=1|\mathbf{x}\}}{P\{\alpha_n=-1|\mathbf{x}\}}$$

[‖]The backward recursion proceeds similarly.

[**]In the case of a residual frequency error, this parameter has to be optimized accordingly.

DOI: 10.1002/sat

is computed. The receiver takes the following decision on symbol α_n

$$\hat{\alpha}_n = \text{sign}\,[L_n]$$

while $|L_n|$ is an estimate of the reliability of this decision—the larger its value the more reliable the corresponding decision.

The described algorithm is more conveniently implemented in the logarithmic domain [15]. It turns out that, in this case, it is required to compute the logarithm of the sum of exponentials (the Jacobian logarithm), which results to be [15]

$$\begin{aligned}\ln(e^{x_1} + e^{x_2}) &= \max(x_1, x_2) + \ln(1 + e^{-|x_1 - x_2|}) \\ &\simeq \max(x_1, x_2)\,.\end{aligned}$$

The complexity of this new detector is equivalent to that of a Viterbi algorithm with $3Q$ transitions per trellis section. It is thus $3Q/4$ times more complex than that in [4]. Its implementation is however compatible with an on-board implementation [3].

This detection algorithm is a soft-input soft-output (SISO) algorithm. This means that an estimate of the ratio between the signal amplitude and the noise power spectral density N_0 must be available (see eqn. (10)). However, in this case of absence of channel coding this is not critical. On the contrary, the availability of soft decisions represents a powerful tool to improve the receiver performance. In fact, although a channel coding scheme is not adopted in the AIS scenario, we can still use the CRC foreseen by the AIS standard to improve the performance of the adopted SISO detection algorithm, by using the simple post-processing methods described in the next section.

3.3. Post-Processing Techniques

Frame synchronization based only on the start flag field of the AIS message would be not reliable enough in interference-limited systems as all messages use the same start flag field. Hence, as in [4], the alignment to the start of the message is achieved by using the CRC. In particular, for all 128 possible starting positions of the message, taking into account the maximum differential delay between messages in the coverage area, once a start flag field is observed except for a couple of possible errors, the CRC is verified. When the right position is found, corresponding to a successul verification of the CRC, the decoded message is passed on to the message parser block, which has the functions of discarding duplicated messages and passing the successfully detected messages to the signal reconstruction block of each zonal demodulator. With respect to the approach in [4], we also adopt one of two possible post-processing techniques that can provide an impressive performance improvement. Every time the CRC verification fails, these techniques attempt an error correction. This procedure does not change in the presence of bit stuffing. In fact, in the AIS standard, it is foreseen that if five consecutive ones are found in the bit stream to be transmitted, a zero should be inserted after the five consecutive ones. As a consequence, at the receiver, when five consecutive ones are found followed by a zero, the burst length must be increased by 1 and the initial bit of the CRC field translated accordingly.

Before describing the technique for error correction, we need to explain a property of the AIS modulation format. CPMs are characterized by an intrinsic differential encoding. This means that, at high signal-to-noise ratio (SNR) values, errors occur in couples of consecutive bits. Considering the additional stage of differential encoding foreseen by the AIS standard, the error patterns at high SNR values are in the form “wcw”, where “w” represents a wrong bit and “c” a correct one. In addition, at high SNR values, when a packet is wrong, usually a single couple of bit errors occurs. Hence, the two following alternative post-processing stages can be employed.

3.3.1. Bit flipping When the CRC says that the packet is wrong, we assume that only one couple of bits is wrong, so we search for the couple of bits with the lowest reliability and revert them. This operation is performed sequentially on the two least reliable couples by checking every time if a valid codeword (i.e., a codeword satisfying the CRC check) is found after flipping.

DOI: 10.1002/sat

3.3.2. Syndrome decoding The syndrome of any valid codeword is always equal to a constant value (it is not zero since the CRC foreseen by the standard is not a linear code due to the particular employed initialization), and the syndrome of an invalid codeword depends only on the error sequence, and is independent of the transmitted sequence. For our purposes, the syndrome corresponds to the CRC of the received sequence. To apply this kind of post-processing, all error patterns containing 1 and 2 couples of wrong bits are tested off-line and the corresponding syndromes are saved into a memory. When receiving a sequence from the detection stage, the following steps are performed:

- the CRC is computed from the received sequence;
- if the computed syndrome equals the syndrome of a correct codeword, the sequence is declared correct;
- otherwise, a search is performed for the computed syndrome among those corresponding to the saved error patterns, starting from those derived from a single couple of errors;
- if a correspondence is found, the word is corrected, otherwise it is declared wrong.

To further improve the reliability of this correction strategy, when searching for errors corresponding to 2 couples of wrong bits, only the couples of bits whose LLRs don't exceed a fixed threshold are corrected (this threshold is a system parameter and, in our simulations, we chose the value of 15).

The second technique outperforms the first one, at the price of an increased computational complexity; however, both techniques have a complexity that is limited when compared to that of the detection algorithm. An important aspect of both processing techniques is that they maintain the error detection capability of the CRC, since they declare an error if they are unable to correct the received message. In other words, we verified through computer simulations that both techniques negligibly increase the probability that a valid codeword is declared in the presence of errors.

3.4. Post-Detection Synchronization and Digital Re-Modulation

In this section, we describe the new post-detection synchronization unit and we highlight the differences between this new unit and the one in [4].

When a packet is correctly detected, it can be re-modulated and subtracted from the received signal in order to try to detect other packets. However, the corresponding (time-invariant) amplitude and (time-varying) channel phase, unnecessary to perform detection, must be estimated. In addition, a refined frequency estimate must be also computed since the frequency uncertainty after the pre-detection synchronization stage is larger than that required for a reliable interference cancellation. Through computer simulations, we verified that one of the most critical tasks is represented by the frequency estimation. In fact, in this case a very large accuracy is required. In order to have a limited performance loss with respect to the case of perfect cancellation, the residual frequency error must be lower than $10^{-4}/T$, thus much lower than the frequency error of $10^{-2}/T \div 1.5 \cdot 10^{-2}/T$ tolerated by the described detection algorithm.

Although very efficient data-aided (DA) algorithms based on the *whole* packet are adopted in [4] for frequency, (time-invariant) phase and amplitude estimation, a non negligible performance loss with respect to perfect cancellation is experienced. In order to improve the performance of the estimation algorithms with respect to those proposed in [4], we suggest the following modifications.

First of all, we perform the post-detection synchronization based on the oversampled received signal instead of on the matched filter output. The advantage is that, contrarily to what happens at the output of the matched filter, noise is white and intersymbol interference (ISI) is removed. In other words, it is avoided that ISI and the colored noise degrade the performance. Therefore, we suggest to reconstruct an oversampled version of the detected packet, also necessary to perform cancellation, with time shift provided by the pre-detection stage and arbitrary amplitude and phase. This can be simply done through a discrete-time modulator and an interpolator. Thus, in Fig. 2, the block "signal reconstruction" is, in practice, a discrete-time CPM modulator followed by a quadratic interpolator that, taking into account the timing estimate performed in the predetection stage, tries to align the reconstructed signal and the received samples. Since the discrete-time CPM modulator has to produce 3 samples for each couple (α_n, ϕ_n) (the correlative state is absent in the case of GMSK

DOI: 10.1002/sat

and the phase state ϕ_n takes on 2 values), it can be conveniently implemented through a look-up table. We will see that, contrarily to what done in [4], it is not necessary to employ the frequency estimate obtained in the pre-detection stage since our post-detection frequency estimator algorithm, that we will now describe, has a sufficiently large estimation range.

Let us denote by $\{\hat{s}_{n\eta+m}\}$ the samples of this reconstructed packet. Frequency estimation is then performed on samples $z_{n\eta+m} = r_{n\eta+m}\hat{s}^*_{n\eta+m}$ by using the DA Mengali and Morelli algorithm [13]. The use of the Mengali and Morelli algorithm, which has the same performance as the Luise and Reggiannini estimator [16] suggested in [4], allows also to remove the main limitation of this latter algorithm. In fact, the Luise and Reggiannini algorithm has an estimation range which depends on the number of symbol intervals observed by the estimator—the larger this number the more limited the estimation range. Considering that the initial frequency uncertainty (after the pre-detection stage) is of $\pm 1.5 \cdot 10^{-2}/T$, the estimator can work by using a very limited number of symbol intervals thus providing a very limited estimation accuracy. To address this problem, in [4] it is suggested to perform frequency synchronization in two steps by using a frequency estimator working on a limited number of symbols in the first step and a second estimator (still based on the Luise and Reggiannini algorithm) working on a larger number of symbols to increase the accuracy. Since the Mengali and Morelli algorithm has an estimation range larger than $\pm 0.2/T$ independently of the number of observed symbol intervals, we can use the whole packet to obtain the most accurate estimate. We verified that, for a given number of observed symbols, the Mengali and Morelli algorithm has the same performance as the Luise and Reggiannini algorithm for both the AWGN scenario and the interference-limited scenario. In addition, they reach the modified Cramer-Rao lower bound (MCRB) in the AWGN scenario. So there is no room for improving the post-detection frequency synchronization. We also verified that there is no performance loss in the frequency estimation when in the presence of the residual timing error before the post-detection frequency synchronization.

Post-detection phase and amplitude estimation can then be performed jointly by using the maximum likelihood (ML) technique. To simplify the notation, we denote by $S_{n\eta+m}$ sample $\hat{s}_{n\eta+m}$ after post-detection frequency estimation and compensation. Denoting by $\hat{\theta}$, and $\hat{A}$, the estimates of phase and amplitude, respectively, the time-varying channel phase is updated using a DA first-order phase-locked loop (PLL) with error signal given by $\Im[r_{n\eta+m}S^*_{n\eta+m}e^{-j\hat{\theta}_{n\eta+m}}]$ whereas the amplitude is estimated as[††]

$$\hat{A} = \frac{\left|\sum_{n=0}^{N-1}\sum_{m=0}^{\eta-1} r_{n\eta+m}S^*_{n\eta+m}e^{-j\hat{\theta}_{n\eta+m}}\right|}{N\eta}$$

where N is the number of symbol intervals considered for the estimation. Since the complexity is very limited, we have considered estimates based on the whole packet. In our opinion there is no need to use the noncoherent post detection integration suggested in [4] to perform the amplitude estimation since post-detection frequency estimation and compensation has already been performed and an algorithm robust to uncompensated frequency offsets is not required.

It is observed that the cancellation can be improved, thus obtaining a (limited) performance improvement, when timing estimation is refined after post-detection frequency estimation and compensation. We propose to perform this task using the following DA algorithm. First of all, the

[††]To leave out of consideration the initialization of the PLL, a forward and a backward PLL can be employed. The first one is used to estimate the phase in the second half of a packet, whereas the second one will be employed to estimate the phase in the first half packet.

DOI: 10.1002/sat

following quantities are computed:

$$\begin{aligned}\gamma_0 &= \sum_{n=0}^{N-1}\sum_{m=0}^{\eta-1} r_{n\eta+m} S^*_{n\eta+m} e^{-j\hat{\theta}_{n\eta+m}} \\ \gamma_1 &= \sum_{n=0}^{N-1}\sum_{m=0}^{\eta-1} r_{n\eta+m} S^*_{n\eta+m+1} e^{-j\hat{\theta}_{n\eta+m+1}} \\ \gamma_{-1} &= \sum_{n=0}^{N-1}\sum_{m=0}^{\eta-1} r_{n\eta+m} S^*_{n\eta+m-1} e^{-j\hat{\theta}_{n\eta+m-1}} .\end{aligned}$$

The refined timing estimate is computed in closed form as

$$\tau = \frac{\eta}{T} \frac{\Re\left[\gamma_0^* \left(\gamma_1 - \gamma_{-1}\right)/2\right]}{\left|\gamma_1 - \gamma_{-1}\right|^2/4 + \Re\left[\gamma_0^* \left(\gamma_1 + \gamma_{-1} - 2\gamma_0\right)\right]}$$

It must be evaluated if this timing refinement and the following quadratic interpolation has a complexity which deserves to be spent considering the limited performance improvement. As can be observed from Fig. 2, the post-detection estimation must be performed using the samples before the limiter.

Finally, we would like to mention the fact that the AIS standard foresees a few symbols of ramp-up and ramp-down at the beginning and at the end of a packet. This must be taken into account during the cancellation, i.e., the reconstructed signal must have appropriate ramp-up and ramp-down intervals. From the analysis of real received AIS packets, we were able to observe the power profile corresponding to the ramp-up and ramp-down sections. Hence, it is possible to estimate the parameters of the power profile and reconstruct the waveform combining these estimated profiles with the reconstructed packet based on detected symbols.

4. FURTHER ON-GROUND PROCESSING

The receiver described in this paper has been designed to work on board of satellites, where limited power and processing capabilities are available. At this point, we may ask ourselves whether a better performance can be achieved by moving our processing from the satellite to a ground station, where additional computational power can be used. Since the receiver has a complexity suitable for an on-board implementation [3], no improvement is possible by only moving the processing to the ground. However, it is possible to exploit there some available information on previously processed packets. This kind of information cannot be available on satellites, hence this further processing stage can be performed only by ground stations. Consequently, the detection algorithm can be fed with a-priori information on some of the bits, thus obtaining a further performance improvement. In particular, the steps the receiver can follow in this case are:

1. Whenever a packet arrives, detection is attempted.
2. If the decoding process fails then, from the time the satellite received the data, the satellite's field of view at the transmission time is estimated, thus characterizing the set of possible transmitters thanks to a database of ships information which is continuously updated based on the detected AIS messages.
3. The message identification number is searched among the possible transmitters and, if found, detection is reattempted including the a-priori information.
4. If detection succeeds, data is extracted from the message, otherwise the packet is discarded.

Besides some fixed fields of transmitted packets, that can be assumed perfectly known at the receiver, the nature of AIS system allows to use some additional bits coming from the latitude and longitude fields. Their number and confidence level can be determined based on the latest reported position, speed and heading. Given the limited distance that can be traveled by a ship between two

DOI: 10.1002/sat

consecutive AIS reports, the most significant bits of its coordinates can be considered reliable with a certain confidence level. Since AIS is a memoryless system, all information about the position of a ship is included in the latest successfully decoded message. At the moment this report is received, the position is known with virtually no uncertainty but, as time passes, this point transforms into a growing region around the initial position. Since a message includes also speed and course of each vessel, this region is not uniform around the starting point.

At each time instant, the ship is then expected to be in a certain position, with a certain confidence given by a probability density function, whose parameters are related to the vessel's maximum speed; we have modeled this uncertainty by means of a Gaussian distribution centered in the expected position, and we have evaluated the reliability information for a series of test cases. One example is reported in Section 5.

5. NUMERICAL RESULTS

We assess the performance of the proposed receiver and compare it with that in [4]. The performance will be shown in terms of packet error rate (PER) versus E_b/N_0. A detailed comparison by using a complex system simulator described in [3] is on going in the framework of a project funded by the European Space Agency. The results, not publicly available at the moment, are in line with those here described, i.e., the receiver here proposed is actually that with the best performance.

In Fig. 4, we consider the case of presence of at most a single interferer in addition to the reference signal (which is that with the highest power), with different values of signal-to-interference power ratio (SIR). Both the reference signal and the interferer have a random normalized Doppler frequency uniformly distributed in the interval $[0, 0.22]$. We report the PER values related to the detection of the reference signal without taking into account the possibility to detect the interferer after successful detection and cancellation of the reference signal. Both our proposed post-processing techniques ensure a performance gain of several dBs in both presence and absence of interference from another signal, and the syndrome decoding algorithm gains a further dB with respect to bit flipping. In the same figure, we also report an example of a curve obtained when considering in the detection stage some a-priori information on some of the message bits. For this case, we have considered a ship sailing in the open sea and, through probabilistic considerations, we were able to find some level of reliability information on 14 bits from the coordinates fields.

Fig. 5 compares the proposed receiver with that of [4], in the presence of five interfering signals (in this case, the denominator of the SIR must be interpreted as the sum of the powers of all interfering signals). Having a larger number of interferers, their sum tends to have a Gaussian distribution and this is for sure a worse case. Thus, performance degrades with respect to the single interferer case, yet the gain over [4] is still remarkable. Finally, Fig. 6 can give an idea on the performance in the presence of cancellation. It is assumed that a further user is present, with a power 10 dB larger than that of the reference user. In this scenario, we assume that a more powerful signal has already been received and correctly detected; this signal can then be canceled from the overall received signal before attempting the detection of the reference user. Since cancellation is not ideal, this causes a performance degradation, but results show that detection could still be possible even in this case. Syndrome decoding has been applied to obtain the results in Fig. 6, where at most a single interfering signal is present.

6. CONCLUSIONS

This paper proposes several improvements to the receiver described in [4], designed for the reception of AIS signals from LEO satellites. Many of the blocks of [4] have been replaced by more efficient and performing ones. The receiver exhibits excellent performance with respect to [4], and it is suitable for implementations both on board of satellites and in ground-based stations, with the possibility, in the latter case, to exploit available information to further improve performance.

DOI: 10.1002/sat

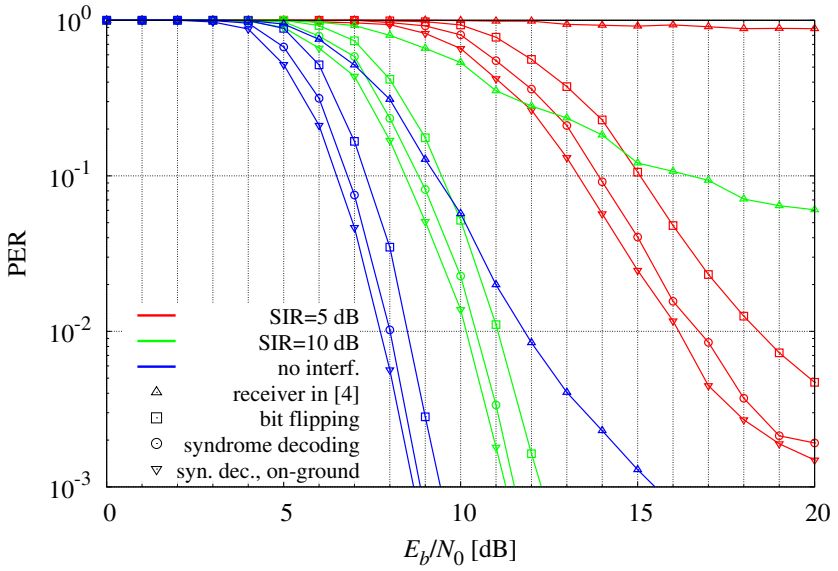

Figure 4. Performance comparison between the proposed receiver and the receiver in [4], for a single interfering signal.

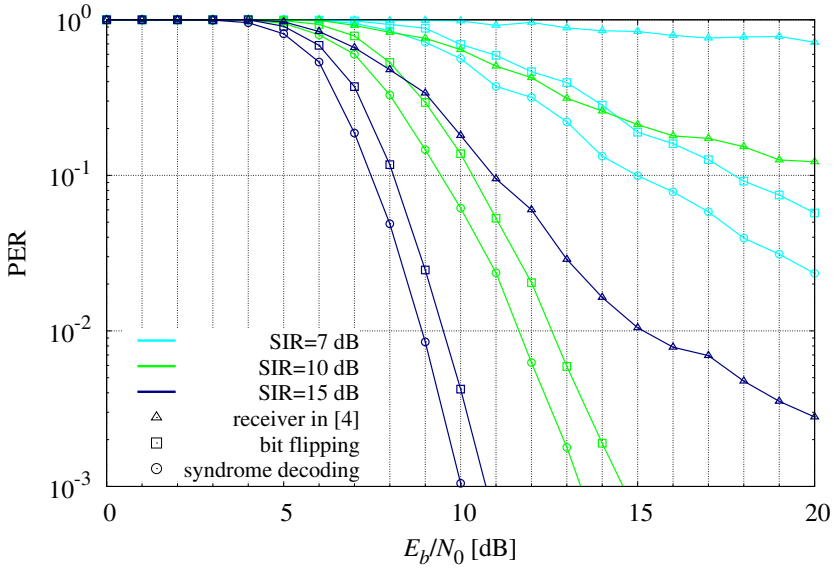

Figure 5. Performance comparison between the proposed receiver and the receiver in [4], for 5 interfering signals.

References

1. International Telecommunication Union, "Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile band," *Recommendation ITU-R M.1371-4*, Apr 2010.
2. G. Colavolpe, T. Foggi, A. Ugolini, J. Lizarraga, S. Cioni, and A. Ginesi, "Receiving method and receiver for satellite-based automatic identification system," Jan. 2014. assigned to ESA-ESTEC, The Neederlands. International patent application n. PCT/EP2014/051273.
3. A. Scorzolini, V. De Perini, E. Razzano, G. Colavolpe, S. Mendes, P. Fiori, and A. Sorbo, "European enhanced space-AIS system study," in *Proc. 5th Advanced Satellite Mobile Systems Conference and 11th International Workshop on Signal Processing for Space Communications (ASMS&SPSC 2010)*, pp. 9–16, Sept. 2010.

DOI: 10.1002/sat

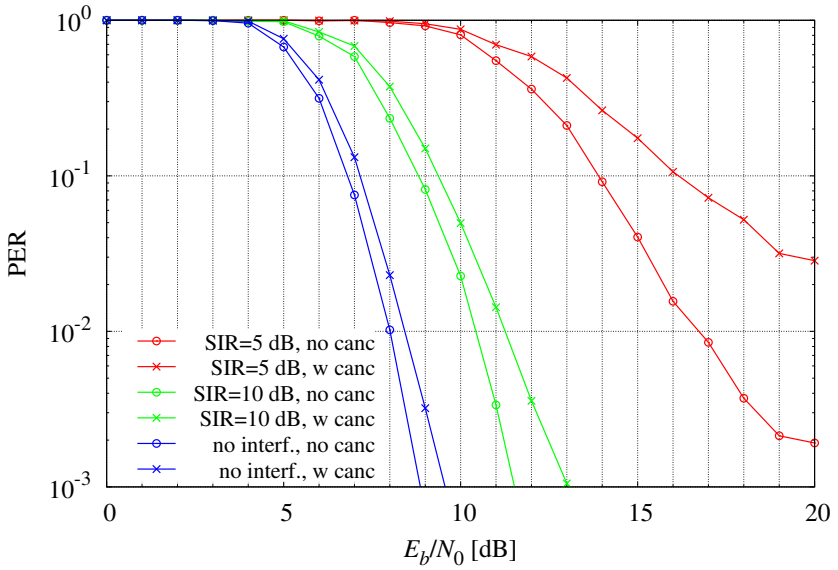

Figure 6. Performance comparison for syndrome decoding with and without cancellation of a more powerful previously detected signal, in the presence of a single interfering signal.

4. P. Burzigotti, A. Ginesi, and G. Colavolpe, "Advanced receiver design for satellite-based automatic identification system signal detection," *International Journal of Satellite Communications and Networking*, vol. 30, pp. 52–63, March/April 2012.
5. R. Prévost, M. Coulon, D. Bonacci, J. LeMaitre, J.-P. Millerioux, and J.-Y. Tourneret, "CRC-assisted error correction in a trellis coded system with bit stuffing," in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pp. 381–384, June 2011.
6. R. Prévost, M. Coulon, D. Bonacci, J. LeMaitre, J.-P. Millerioux, and J.-Y. Tourneret, "Extended constrained Viterbi algorithm for AIS signals received by satellite," in *Satellite Telecommunications (ESTEL), 2012 IEEE First AESS European Conference on*, pp. 1–6, Oct 2012.
7. J. B. Anderson, T. Aulin, and C.-E. W. Sundberg, *Digital Phase Modulation*. New York: Plenum Press, 1986.
8. B. E. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inform. Theory*, vol. 34, pp. 260–270, Mar. 1988.
9. U. Mengali and M. Morelli, "Decomposition of M-ary CPM signals into PAM waveforms," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1265–1275, Sept. 1995.
10. G. Colavolpe and R. Raheli, "Reduced-complexity detection and phase synchronization of CPM signals," *IEEE Trans. Commun.*, vol. 45, pp. 1070–1079, Sept. 1997.
11. K. Murota and K. Hirade, "GMSK modulation for digital mobile radio telephony," *IEEE Trans. Commun.*, vol. 29, pp. 1044–1050, July 1981.
12. M. Morelli and U. Mengali, "Joint frequency and timing recovery for MSK-type modulation," *IEEE Trans. Commun.*, vol. 47, pp. 938–946, June 1999.
13. U. Mengali and M. Morelli, "Data-aided frequency estimation for burst digital transmission," *IEEE Trans. Commun.*, vol. 45, pp. 23–25, Jan. 1997.
14. A. Barbieri and G. Colavolpe, "Simplified soft-output detection of CPM signals over coherent and phase noise channels," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2486–2496, July 2007.
15. P. Roberston, E. Villebrun, and P. Hoeher, "Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding," *European Trans. Telecommun.*, vol. 8, pp. 119–125, March/April 1997.
16. M. Luise and R. Reggiannini, "Carrier frequency recovery in all-digital modems for burst-mode transmissions," *IEEE Trans. Commun.*, vol. 43, pp. 1169–1178, Mar. 1995.

DOI: 10.1002/sat