Data Integrity for GPS and Galileo Signals used by Civil Aviation

Axel Javier Garcia Peña, Olivier Julien

Data Integrity for GPS and Galileo Signals used by Civil Aviation

Axel Garcia-Pena and Olivier Julien
SIGnal processing and NAVigation (SIGNAV) research group
TELECOM Laboratory
Ecole Nationale de l'Aviation Civile (ENAC)

BIOGRAPHY (IES)

Axel Garcia Pena is a researcher/lecturer with the SIGnal processing and NAVigation (SIGNAV) research group of the TELECOM lab of ENAC (French Civil Aviation University), Toulouse, France. His research interests are GNSS navigation message demodulation, optimization and design, GNSS receiver design and GNSS satellite payload. He received his double engineer degree in 2006 in digital communications from SUPAERO and UPC, and his PhD in 2010 from the Department of Mathematics, Computer Science and Telecommunications of the INPT (Polytechnic National Institute of Toulouse), France.

Olivier Julien is the head of the SIGnal processing and NAVigation (SIGNAV) research group of the TELECOM lab of ENAC (French Civil Aviation University), Toulouse, France. His research interests are GNSS receiver design, GNSS multipath and interference mitigation, and interoperability. He received his engineer degree in 2001 in digital communications from ENAC and his PhD in 2005 from the Department of Geomatics Engineering of the University of Calgary, Canada.

ABSTRACT

GNSS signals coupled with control mechanisms at the receiver end must guarantee that the information transmitted through the navigation message is provided unaltered to the final end user. From this premise the concept of Data Integrity is presented in this paper - measure of the trust that can be placed in the invariance of the received message information with respect to the transmitted information- and is quantified through 3 parameters, probability of erroneous message, probability of undetected error and Data Integrity Risk (DIR).

The Data Integrity concept is special important in Civil Aviation where the provision of altered Clock offset correction and satellite Ephemeris Data (CED) to the end user can provoke an out of tolerance position error which would directly impact the Integrity Risk (IR) set by the civil aviation standards, $IR = 10^{-7}$/operation. Therefore, this paper discusses a requirement for the DIR of CED information for the civil aviation application, DIR_{CA} as well as presents the principle of calculation of the CED information DIR, DIR_{CED}, (and any other generic field, DIR_F) of any couple GNSS signal/control mechanism.

Several DIR calculation expressions are provided and commented in this paper. Moreover, they are applied, with their required modifications, to two GNSS signals, Galileo E1 OS and GPS L1 C/A for two different control mechanisms – CED information is only read once and the same CED information must be read twice in a row to be provided to the final end user. The C/N_0 thresholds for which both GNSS signals/control mechanisms couples are compliant with the civil aviation DIR requirement, DIR_{CA}, are provided and compared.

INTRODUCTION

GNSS signals are designed to fulfill the specific needs of the system: to provide the receiver with precise synchronization in a precise pseudo-range measurement whilst also broadcasting essential information such as the satellite ephemeris, clock offset correction parameters and other data. The combination of these two elements allows a GNSS to provide the user with the ability to compute its PVT (position, velocity, time).

The broadcast of the essential information is achieved through the transmission of a navigation message. The navigation message varies between the different GNSS signals since, among other reasons, the system level objectives associated to each signal differ. However, any navigation message design must ensure that no erroneous information is used by the end user, giving special attention to the information which leads to an out of tolerance position (or error for a more general scope). Remark that this work will only consider the information to be erroneous when the received information provided to the end user is different from the transmitted one. Any other source of error, such as information incorrectly generated by the ground station of information incorrectly received by the GNSS satellite from its transmission to the ground station, will not be taken into account.

The importance to ensure that no erroneous information is used by the end user depends on the application. And this objective is especially important in Safety-of-Life (SoL) applications such as civil aviation where the integrity of the calculated position is critical to ensure safe air operations [1]. Therefore, if the probability of guaranteeing that no erroneous information is provided to the end user is seen as a requirement, the numerical value of this requirement is application dependent.

Moreover, since this probability mainly depends on the navigation message/signal characteristics of each signal and on the control mechanisms implemented in the receiver, it can be concluded that these elements should be designed in order to guarantee that the probability of providing erroneous information to the end user is lower than the targeted application requirement.

The last two paragraph have introduced two notions which are the basis of the Civil Aviation integrity concept [1]: event leading to an out of tolerance error position and error probability lower than a set requirement. Therefore, this article focus on introducing or formalizing the Data Integrity concept using the Civil Aviation integrity concept as model.

Two benefits can be expected from the introduction of the Data Integrity concept:

1- To evaluate if an existing GNSS signal can meet the requirement of a given application with the existing control mechanism at the receiver
2- To optimize the navigation message design and associated control mechanism at the receiver end to fulfill the requirement of an application

This paper is organized as follows. First a formal definition of the Data Integrity concept is introduced. Second, a discussion of a possible value for the Data Integrity Risk for Civil Aviation is discussed. Third, the principle of Data Integrity Risk calculation for a generic signal is provided. Fourth, the GPS L1 C/A and Galileo E1 OS signal characteristics are presented. Fifth, the DIR_{CED} of GPS L1 C/A and Galileo E1 OS signals are shown. Finally, the conclusions are given.

DATA INTEGRITY CONCEPT

The Data Integrity concept is defined by extrapolating the civil aviation integrity concept as follows.

Data Integrity is a measure of the trust that can be placed in the correctness/invariance of the received message information with respect to the transmitted information: an information is considered to be erroneous when the transmitted information is different from the transmission provided to the end user by the receiver. Therefore, Data Integrity also includes the ability of the receiver to detect and discard erroneous/altered information. The conditions of losing the data integrity are to provide the end user with one or more erroneous/altered information bits.

The data integrity can thus be defined with the following 3 parameters.

Probability of erroneous message (P_e): It is defined as the probability that the received information message is erroneous/altered. It depends on the propagation channel, on the signal modulation (BPSK for GNSS signals), the inner channel code used for Forward Error Correction (FEC) purposes, on the tracking performance and on the received C/N_0 value.

Probability of an undetected error (P_{und}): It is defined as the probability that the receiver is unable to detect an erroneous information message. It depends on the outer channel code used for detection purposes and on the receiver control mechanisms. However, it is completely independent of the received C/N_0.

Data integrity risk (DIR): It is defined as the probability that the received message information is not the same as the transmitted one, that the receiver is unable to detect the alteration of the information and thus that the receiver use this misleading information during any operation.

Two DIR types are defined depending on whether the DIR is an application requirement or it is the result of the analysis of the Navigation Message structure, signal characteristics impacting the information data and the Control Mechanism implemented at the receiver (NMCM) over a specific field of the navigation message carrying the targeted information:

a) **DIR Field, $\boldsymbol{DIR_F}$**: It is defined as the probability guaranteed by a NMCM to provide the end user with no erroneous (invariant) information over a specific information field. It is a function of P_e and of P_{und}. $DIR_F = f(P_e, P_{und})$
b) **DIR Application Requirement, $\boldsymbol{DIR_{AR}}$**: It is defined as the maximum probability of providing the end user with erroneous (altered) information which is tolerated by an application. Therefore, its value depends on the application.

Finally, the Data Integrity concept has as main objective either:

- to verify that the DIR of the NMCM of an existent navigation message/receiver couple calculated over a given field F is smaller than the DIR of a targeted application (see equation (1)), or
- to design a NMCM which provides a DIR of a given field smaller than the DIR of a targeted application (see equation (1)).

$$DIR_F < DIR_{AR} \quad (1)$$

DATA INTEGRITY RISK FOR CIVIL AVIATION

In civil aviation, the airborne receiver must be certain that clock offset corrections and satellites ephemeris are error free before using them. In the event that they are not erroneous free, they could cause the receiver to erroneously calculate its PVT and even to cause and out of tolerance position error [1]. In this last case, the receiver must guarantee that the probability of committing this out of tolerance position error is lower than the probability required by the standards [1], the well-known Integrity Risk (IR_{CA}) which is equal to 10^{-7}/operation.

There are two possibilities to provide the final user with erroneous information. The first one is that the transmitted information data is already incorrect: information data generated by the ground station is incorrect, transmission from the ground station to the satellite is corrupted and/or the navigation message generated by the GNSS satellite is incorrect. This case is covered by the probability of faulty satellite equal to 10^{-5}/h [1], it is thus already taken into account on the $IR_{CA} = 10^{-7}$/operation value and, consequently, it is not covered in this paper. The second possibility is that the information data transmitted by the satellite is different from the information data received by the user. This possibility corresponds to the Data Integrity Risk definition given on the previous section and it is not specifically tackle on today civil aviation standards [1]. Therefore, a derivation must be done in order to evaluate the Data Integrity Risk for Civil Aviation, DIR_{CA}.

A direct and simple way to provide a DIR_{CA} value without heavy mathematical derivations is to set a value small enough to have a negligible impact on IR_{CA}. In doing so, any calculation and any value provided on the standards should not be affected by the formal introduction of DIR_{CA}. Therefore, this paper proposes to set as a preliminary value until a formal mathematical derivation is done of $DIR_{CA} \approx 10^{-10}$.

Finally, there are two main choices for the field of interested over which the DIR_F will be calculated for a given NMCM. At the beginning of the section, it was said that and out of tolerance position error could be produced when the Clock offset corrections and satellites Ephemeris Data (CED) provided to the final end user were erroneous. Therefore, the first possible choice is to chosen as the targeted field the CED, and thus the targeted Data Integrity Risk is referred as DIR_{CED}. However, another field which significantly affects the final position calculation and which could cause and out of tolerance position error is the Ionospheric corrections. Therefore, the second possible choice is to choose as the targeted field the CED plus the Ionospheric corrections, and thus the targeted Data Integrity Risk is referred as $DIR_{CED\&I}$. In any case, regardless of the selected field, it must be guaranteed in order to fulfill the Data Integrity requirements and thus the Integrity requirements of Civil Aviation that:

$$DIR_{CED} < DIR_{CA} \quad or \quad DIR_{CED\&I} < DIR_{CA} \qquad (2)$$

PRINCIPLE OF DATA INTEGRITY RISK CALCULATION

In this section, the principle of calculation of the DIR over a specific field for a given NMCM is presented. The principle presented in this section is valid for a generic signal; however it has to be usually customized for a specific NMCM in order to take into accounts its unique characteristics.

This section is divided into six parts. The first one presents the navigation message model of a generic GNSS signal. The second one presents the generic model of a GNSS receiver from the demodulation point of view. The third part presents the DIR exact calculation of a word, DIR_w. The fourth and fifth present a tight and loose approximations of the DIR_w. The sixth and last part presents the DIR_F derivation from the DIR_w carrying the field, F.

Navigation message model of a generic GNSS signal

Figure 1 presents the navigation message structure of a generic GNSS signal. A navigation message is constructed from the application of two different types of channel codes, the parity-check (PC) channel code which is used for detection purposes and the Forwards Error Correction (FEC) channel code which is used for corrections purposes. The construction of the navigation message is given next.

First the information bits are generated where $\boldsymbol{u^j}$ represent the information bit generated at epoch j. Second, the information bits are grouped into groups of $\boldsymbol{k_o}$ information bits in order to generate the outer information word, $\mathbf{w_o} = \{\boldsymbol{u_{o,1}}, \dots, \boldsymbol{u_{o,k_o}}\}$ where $\boldsymbol{u_{o,k}}$ represents the kth information bit of $\mathbf{w_o}$. Once the outer information word is generated, the PC encoding is applied to generated the outer codeword of $\boldsymbol{n_o}$ bits and referred as $\mathbf{c_o} = \{\boldsymbol{c_{o,0}}, \dots, \boldsymbol{c_{o,n_o-1}}\}$, where $\boldsymbol{c_{o,n}}$ represents the nth coded bit of $\mathbf{c_o}$, and the PC encoding function is denoted as g_{PC}. Afterwards, the coded bits of $\mathbf{c_o}$ are regrouped (if necessary) into blocks of $\boldsymbol{k_i}$ bits in order to generate the inner information word $\mathbf{w_i} = \{\boldsymbol{u_{i,1}}, \dots, \boldsymbol{u_{i,k_i}}\}$ where $\boldsymbol{u_{i,k}}$ represents the kth information bit of $\mathbf{w_i}$, and the bit regrouping function is denoted as q.

$$[\mathbf{w}_{\mathrm{i},1}, \dots, \mathbf{w}_{\mathrm{i,P}}] = q(\mathbf{c}_{\mathrm{o}}) \qquad (3)$$

Where P is the number of inner information words generated by the "bit grouping function" q, and $\mathbf{w_{i,p}}$ is the pth inner information word generated by q.

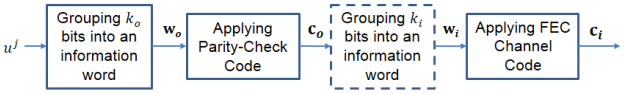

Figure 1 - Navigation message Structure of a generic GNSS signal

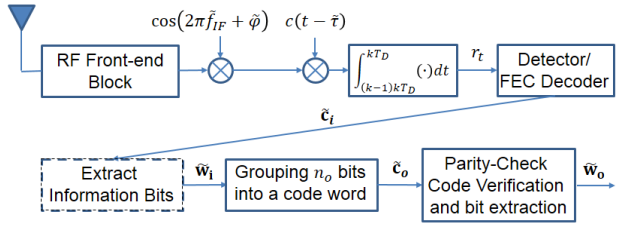

Figure 2 – Receiver chain from a demodulation point of view

Then, the FEC encoding is applied to each $\mathbf{w_i}$ in order to generate the inner codeword of $\boldsymbol{n_i}$ bits and referred as $\mathbf{c_i} = \{\boldsymbol{c_{i,1}}, \dots, \boldsymbol{c_{i,n_i}}\}$, where $\boldsymbol{c_{i,n}}$ represents the n[th] coded bit of $\mathbf{c_i}$, and the FEC encoding function is denoted as g_{FEC}. Finally, the $\boldsymbol{c_{i,n}}$ are fed to the modulator block and thus they are the bits (after modulation, etc) transmitted through the propagation channel.

Receiver chain from a demodulation point of view

Figure 2 presents the receiver chain from a demodulation point of view. The explanation of this figure is skipped until the in-phase prompt correlators outputs which contain the navigation message data: r_t represents the in-phase prompt correlator output at epoch t. The correlators outputs are fed to a FEC decoder which estimates the most probable inner codeword, $\tilde{\mathbf{c}}_i = \{\tilde{c}_{i,1}, \dots, \tilde{c}_{i,n_i}\}$, where $\tilde{c}_{i,n}$ represents the n[th] coded bit of $\tilde{\mathbf{c}}_i$ and the FEC decoding function is denoted as h_{FEC}. Afterwards, an estimation of the inner information word $\tilde{\mathbf{w}}_i = \{\tilde{w}_{i,1}, \dots, \tilde{w}_{i,k_i}\}$ is generated by extracting the estimated inner information bits, $\tilde{w}_{i,k}$, from $\tilde{\mathbf{c}}_i$, where $\tilde{w}_{i,k}$ is the k[th] information bit of $\tilde{\mathbf{w}}_i$ and the FEC bit extraction function is denoted as f_{FEC}. Then, an estimation of the most probable outer codeword, $\tilde{\mathbf{c}}_o = \{\tilde{c}_{o,1}, \dots, \tilde{c}_{o,n_o}\}$, is obtained by regrouping the bits of $\tilde{\mathbf{w}}_i$, where $\tilde{c}_{o,n}$ represents the n[th] coded bit of $\tilde{\mathbf{c}}_o$, and the bit regrouping function is denoted as q^{-1} since it is the inverse function q.

$$q^{-1}\left(\tilde{\mathbf{w}}_{i,1}, \dots, \tilde{\mathbf{w}}_{i,P}\right) = \tilde{\mathbf{c}}_o \qquad (4)$$

Where $\tilde{\mathbf{w}}_{i,p}$ is the p[th] estimated inner information word used in q^{-1} to obtain $\tilde{\mathbf{c}}_o$. Note that if a FEC channel code was not implemented during the generation of the navigation message, a "Detector block" would have been implemented instead of the "FEC decoder", "Extract information bits" and "Grouping n_o bits into a codeword" blocks in order to directly estimate the most probable outer codeword, $\tilde{\mathbf{c}}_o$.

Finally, the PC decoder block is applied in order to determine if $\tilde{\mathbf{c}}_o$ contains any errors. If it is determined that $\tilde{\mathbf{c}}_o$ is error free, an estimation of the outer information word, $\tilde{\mathbf{w}}_o = \{\tilde{w}_{o,1}, \dots, \tilde{w}_{o,k_o}\}$, where $\tilde{w}_{o,k}$ is the k[th] information bit of $\tilde{\mathbf{w}}_o$, is obtained by extracting the bits from $\tilde{\mathbf{c}}_o$. If it is determined that $\tilde{\mathbf{c}}_o$ contains errors, $\tilde{\mathbf{w}}_o$ is discarded. The estimated information bits, $\tilde{w}_{o,k}$, are the bits provided to the end user.

Denoting h_{PC} as the PC verification test, and denoting f_{PC} as the "bit extraction and discard" function of PC, the application of h_{PC} and f_{PC} on $\tilde{\mathbf{c}}_o$ can be mathematically modelled as:

$$h_{PC}(\tilde{\mathbf{c}}_o) = \begin{cases} 1 & if\ \tilde{\mathbf{c}}_o = \mathbf{c}_o \\ 0 & if\ \tilde{\mathbf{c}}_o \neq \mathbf{c}_o, \tilde{\mathbf{c}}_o\ \epsilon\ Detectable \\ 1 & \tilde{\mathbf{c}}_o \neq \mathbf{c}_o, \tilde{\mathbf{c}}_o\ \epsilon\ Non\ Detectable \end{cases} \qquad (5)$$

$$f_{PC}(\tilde{\mathbf{c}}_o, h_{PC}(\tilde{\mathbf{c}}_o)) = \begin{cases} \tilde{\mathbf{w}}_o & if\ h_{PC}(\tilde{\mathbf{c}}_o) = 1 \\ Discard & if\ h_{PC}(\tilde{\mathbf{c}}_o) = 0 \end{cases} \qquad (6)$$

Therefore, the first line of (5) represents the desired verification test success of the estimated outer codeword when the true and the estimated outer codewords are the same. The second line of (5) represents the desired verification tests fail of the estimated outer codeword when the true and the estimated outer codewords are different. Finally the third line of (5) represents the undesired verification test success of the estimated outer codeword when the true and the estimated outer codewords are different: from expression (6) it can be seen that $\tilde{\mathbf{w}}_o$ with $\tilde{\mathbf{w}}_o \neq \mathbf{w}_o$ will be fed to the end user. Therefore, the last case is the one responsible for causing the loss of data integrity.

DIR mathematical expression of an information word

The mathematical expression of the *DIR* of an information word, DIR_w, can be directly given from its definition and from the navigation message structure and receiver chain descriptions. The definition is given in equation (7).

$$DIR_w = P\left(\tilde{\mathbf{w}}_o \neq \mathbf{w}_o\ \cap \tilde{\mathbf{w}}_o\ is\ not\ discarded\right) \qquad (7)$$

Previous equation can be expressed as shown in equation (8) when $P = 1$, which means that $\mathbf{c}_o = \mathbf{w}_i, n_o = k_i$, and when the propagation channel can be modelled as an AWGN channel (see Appendix A):

$$DIR_w = \frac{1}{M_o}\sum_{j=1}^{M_o}\sum_{\substack{v=1 \\ v\neq j}}^{M_o} Q\left(\sqrt{2d_{u(j)u(v)}\frac{C}{N_0}T_D D}\right) \qquad (8)$$

Where $u(j)$ is a bijective function between the M_o elements of the set $j \in [0, \dots, M_o - 1]$ (same for v) and a subset of M_o elements of set $u \in [0, \dots, M_i - 1]$, $M_o = 2^{k_o}$ is the number of outer codewords/information words of the alphabet, $M_i = 2^{k_i}$ is the number of inner codewords/information words of the alphabet, $M_o < M_i$, $d_{u(j)u(v)}$ is the Hamming distance between $\mathbf{c}_i^{u(j)}$ and $\mathbf{c}_i^{u(v)}$, $\mathbf{c}_i^{u(j)}$ and $\mathbf{c}_i^{u(v)}$ are the inner codeword $u(j)$ and the inner codeword $u(v)$ of the inner codewords alphabet which have their associated $\mathbf{c}_o^j = q^{-1}\left(f_{FEC}\left(\mathbf{c}_i^{u(j)}\right)\right)$ and $\mathbf{c}_o^v = q^{-1}\left(f_{FEC}\left(\mathbf{c}_i^{u(v)}\right)\right)$ succeed the PC verification test, $h\left(\mathbf{c}_o^j\right) = 1$ and $h(\mathbf{c}_o^v) = 1$, T_D is the symbol interval, D is the percentage of power provided to the GNSS signal data component and $Q(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$. Note that $\mathbf{c}_i^{u(j)} \subset \mathbf{c}_i^u$ where $\mathbf{c}_i^u$ represents the inner codeword u of the complete inner codeword alaphabet, $u \in [0, \dots, M_i - 1]$.

The main problem with (8) is its heavy computational load: first, all M_o outer codeword must be identified, $\mathbf{c}_o^j$, second, the associated M_o inner codewords must be calculated, $\mathbf{c}_i^{u(j)}$, third ${M_o}^2/2$ hamming distances must be calculated and finally, ${M_o}^2/2\ Q(x)$ functions must be evaluated. Therefore, the number of operations order of magnitude is $\sim M_o(2 + M_o)$ which can be really high when k_o increases.

The computational burden of the previous expression can be reduced by simplifying the previous when taking into account the FEC and PC linear properties [7]. Moreover, the new expression is also based on defining the estimated codewords/information words as the true word modulo-2 addition an error vector. Starting from the estimated inner codeword:

$$\tilde{\mathbf{c}}_{\boldsymbol{i}} = \mathbf{c}_{\boldsymbol{i}} \oplus \mathbf{e}_{\boldsymbol{ic}} \quad \mathbf{e}_{\boldsymbol{iw}} = f_{FEC}(\mathbf{e}_{\boldsymbol{ic}}) \quad \mathbf{w}_{\boldsymbol{i}} = \mathbf{w}_{\boldsymbol{i}} \oplus \mathbf{e}_{\boldsymbol{iw}} \qquad (9)$$

Where $\mathbf{e}_{\boldsymbol{ic}} = \left\{e_{ic,1}, \dots, e_{ic,n_i}\right\}$ is the inner codeword error vector, $e_{ic,n}$ represents the n$^{\text{th}}$ element of e_{ic}, $\mathrm{e}_{ic,n} = 1$ means that a bit estimation error was committed in n$^{\text{th}}$ position, $\oplus$ represent the modulo-2 addition and $\mathbf{e}_{\boldsymbol{iw}} = \left\{e_{iw,1}, \dots, e_{iw,n_i}\right\}$ is the inner information word error vector.

The simplified DIR_w expression which take into account the error vector definition as well as the FEC and PC channel codes linearity is given below (see appendix B):

$$DIR_w = \sum_{b=2}^{M_i} Q\left(\sqrt{2d_b \frac{C}{N_0} T_D D}\right) \cdot h_{PC}\left(q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^b\right)\right)\right) \qquad (10)$$

Where, $\mathbf{e}_{ic}^b$ is the outer codeword error vecor b, $b \in [1, \dots, M_i]$, and d_b is the Hamming weight (number of ones) of the error vector $\mathbf{e}_{ic}^b$.

Note that $Q\left(\sqrt{2d_b \frac{C}{N_0} T_D D}\right)$ represents the probability of an erroneous message, P_e, and that $h_{PC}\left(q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^b\right)\right)\right)$ represents the probability of an undetected error, P_{und}. Therefore, from this expression is quite easy to link the DIR calculation to one of the previous statements given in this paper: $DIR_w = f(P_e, P_{und})$.

The computational load of this expression is $\sim 3M_i$ (M_i calculations of $\mathbf{e}_{ic}^b$, M_i calculations of d_b and M_i evaluations h_{PC}). Therefore if ${M_i < M_o}^2$, this expression has a computational load lower than expression (8). In any case, if M_i is too large, expression (10) cannot still be calculated with a reasonable amount of time.

DIR mathematical expression tight approximation

Expression (10) can be found by rewritten the addition as a function of the errors' Hamming distance:

$$DIR_w = \sum_{d=d_{min}}^{d_{max}} Q\left(\sqrt{2d \frac{C}{N_0} T_D D}\right) \cdot num_d^{PC} \qquad (11)$$

$$num_d^{PC} = \sum_{\mathbf{e}_{ic}^{b,d}} h_{PC}\left(q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)\right) \qquad (12)$$

Where $\mathbf{e}_{ic}^{b,d}$ represent the inner codeword error vector b having a Hamming weight equal to d, d_{min} and d_{max} are respectively the minimum and the maximum distance of any possible inner codeword error vector. Note that $Q\left(\sqrt{2d \frac{C}{N_0} T_D D}\right)$ represents the probability of committing an error having a Hamming weight equal to d and thus this term is common to all the errors having the same d. Parameter num_d^{PC} represents the quantity of inner codeword errors having Hamming weigh d and which generate (individually) an outer codeword error vector, $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$, which cannot be detected by the PC verification test, $h_{PC}\left(\mathbf{e}_{oc}^{b,d}\right) = 1$.

The computational load of expression (11) can be controlled by taking into account that the contribution to the final DIR_w value of the inner codeword error vectors with the higher Hamming weighs, d, is negligible in comparison with the contribution of error with lower Hamming weighs:

$$Q\left(\sqrt{2d_1\frac{C}{N_0}T_D D}\right)num_{d_1}^{PC} > Q\left(\sqrt{2d_2\frac{C}{N_0}T_D D}\right)num_{d_2}^{PC} \quad (13)$$
$$if\ d_1 < d_2$$

This is due to the fact the first term, $Q(Kd)$, decreases much faster than the increase of the second term, num_d^{PC}. Therefore, expression (11) can be limited by setting a new $d'_{max} < d_{max}$ for which the addition from $d'_{max}+1$ to d_{max} is negligible in comparison to addition from d_{min} to d'_{max}.

To sum up, expression (11) can be used to calculate a DIR_w approximation and its computational load can be reduced by limiting d_{max}. However, this bound is still quite computationally heavy due to the fact that individual outer codeword error vectors must be searched for, their Hamming weigh, d, must be calculate and the PC verification, h_{PC}, must be applied on them (see equation (12)).

$$DIR_w \approx \sum_{d=d_{min}}^{d'_{max}} Q\left(\sqrt{2d\frac{C}{N_0}T_D D}\right)\cdot num_d^{PC} \quad (14)$$

Finally, it is important to remark that although equation (11) is the exact expression to calculate the DIR_W of a generic NMCM and that (14) is a tight bound, these expressions should probably be modified to be adapted to the specifics characteristics of a specific targeted NMCM.

DIR mathematical expression loose approximation

In order to avoid the PC verification and thus in order to reduce the computational burden, an approximation can be used to calculate num_d^{PC}. In fact, num_d^{PC} can be calculated by applying the a-priori detection properties of PC. Usually, the PC channel detection properties can be defined in two different forms:

1) Detection properties depend on the Hamming weight, d^o, of the outer codeword error, $\mathbf{e_{oc}}$.
2) Detection properties depend on the burst length, br, of the outer codeword error, $\mathbf{e_{oc}}$. The burst length of an error vector, e, is defined as the length between the first and the last elements of e which are not zero.

From the first type of detection properties, num_d^{PC} can be expressed as:

$$num_d^{PC} = \sum_{d^o=d^o_{min}}^{d^o_{max}} num_{d,d^o}^{e} \quad (15)$$

Where d^o_{min} and d^o_{max} are respectively the minimum and the maximum distance of any possible outer codeword error vector. num_{d,d^o}^{e} is the true number of inner codeword error vectors having a Hamming weigh d which generate (individually) an outer codeword error vector, $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$ having a Hamming weight equal to d^o which succeed the PC verification test.

Therefore, expression (15) can be approximated as:

$$num_d^{PC} \approx \sum_{d^o=d^o_{min}}^{d^o_{max}} \overline{num}_{d,d^o} \quad (16)$$

$$\overline{num}_{d,d^o} = num_{d,d^o}\cdot P\left(h_{FEC,d^o}\right) \quad (17)$$

Where $\overline{num}_{d,d^o}$ is the average number of inner codeword error vectors having a Hamming weigh d which generate (individually) an outer codeword error vector, $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$ having a Hamming weight equal to d^o which succeed the PC verification test. num_{d,d^o} is the number of inner codeword error vectors having a Hamming weigh d which generate (individually) an outer codeword error vector, $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$ having a Hamming weight equal to d^o. Finally, $P\left(h_{FEC,d^o}\right)$ is the detection probability a priori of the PC channel code as a function of outer codeword error vector Hamming weigh, d^o.

Following the same logic for the 2nd type of detection properties codes:

$$num_d^{PC} = \sum_{br=br_{min}}^{br_{max}} num_{d,br}^{e} \quad (18)$$

$$num_d^{PC} \approx \sum_{br=br_{min}}^{br_{max}} \overline{num}_{d,br} \quad (19)$$

$$\overline{num}_{d,br} = num_{d,br}\cdot P\left(h_{FEC,br}\right) \quad (20)$$

Where br_{min} and br_{max} are respectively the minimum and the maximum burst length of any possible outer codeword error vector. $num_{d,br}^{e}$ and $\overline{num}_{d,br}$ are respectively the true and average number of inner codeword error vectors having a Hamming weigh d which generate (individually) an outer codeword error vector, $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$ having a burst length equal to br which succeed the PC verification test. $num_{d,br}$ is the number of inner codeword error vectors having a Hamming weigh d which generate (individually) an outer codeword error vector, $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$ having a burst length equal to br. Finally, $P\left(h_{FEC,br}\right)$ is the a priori detection probability of the PC channel code as a function of outer codeword error vector burst length, br.

However, one must proceed carefully when using expressions (16) and (19) instead of expressions (15) and (18) since they are loose approximations. In fact, the main limitation of these expressions is the use of $P(h_{FEC,br})$ or $P(h_{FEC,d^o})$ as means to obtain $\overline{num}_{d,d^o}$ or $\overline{num}_{d,br}$: $\overline{num}_{d,d^o}$ or $\overline{num}_{d,br}$ are closer to num^e_{d,d^o} or $num^e_{d,br}$ when $num_{d,br}$ or num_{d,d^o} are of the same order or larger than the inverse of $P(h_{FEC,br})$ or $P(h_{FEC,d^o})$ (law of large numbers [8]). Unfortunately, $num_{d,br}$ or num_{d,d^o} can be quite small compared to the inverse of $P(h_{FEC,br})$ or $P(h_{FEC,d^o})$. Therefore, in this case, the quality of the approximation can be poor.

DIR mathematical expression of a field

The DIR of a given filed, DIR_F, can be calculated from the DIR of the information words, DIR_w, containing the field. Assuming that N words with the same structure (FEC, PC, size, etc.) carry all the information of a given filed, F, DIR_F can be calculated as:

$$DIR_F = 1 - (1 - DIR_w)^N \approx N \cdot DIR_w \qquad (21)$$

DESCRIPTION OF THE SIGNALS

In this section, GPS L1 C/A and Galileo E1 OS signal are described. Their descriptions are only focused on providing to the reader the elements necessary to determine the DIR_{CED} of each signal.

GPS L1 C/A

Table I presents the GPS L1 C/A characteristics [4] and Table II presents the structure of the CED information inside the complete GPS L1 C/A navigation message structure [4].

From Table I it can be seen that no FEC channel code is implemented. This means that the inner information word and codeword are the same ($n_i = k_i$). The other important characteristic to remark is that $n_o(= 32) > k_i(= 30)$ when they should also be the same due to the lack of FEC channel code. The reason for this difference is that the last two bits of $\mathbf{c_o}$ at epoch t are used as the two first bits of $\mathbf{w_o}$ at epoch $t + 1$. Therefore, two bits of $\mathbf{c_o}$ at epoch $t + 1$ are already known in $\mathbf{c_o}$ at epoch t and thus, they are not sent again to the receiver [4]. Figure 3 and Figure 4 show the PC encoding and decoding process.

The general characteristics of the Extended Hamming (32,26) are the following:

- All outer codeword error patterns, $\mathbf{e_o}$,containing up to 3 errors are detected.
- Not all outer codeword error patterns, $\mathbf{e_o}$, with ≥ 4 errors can be detected → An a priori detection probability, $P(h_{FEC,br})$, of 0.015625 can be set.

Finally, for GPS L1 C/A signal two different control mechanisms are analyzed:

1) CED information is only read once and it is provided to the end user if no information is discarded by the FEC verification. If some information is discarded, the whole CED information must be read again.
2) Current avionics standard requirement [2]: CED information is read twice and it is only provided to the end user if no information is discarded by the FEC verification and if the two demodulated CED sets are exactly the same. A whole CED reading is made again if (a) previous condition(s) are(is) not met.

Galileo E1 OS

Table I presents the Galileo E1 OS characteristics [5] and Table III presents the structure of the CED information inside the complete Galileo E1 OS navigation message structure [5]. Note that the part of page size does not include the 10 synchronization bits (or symbols). Moreover, note that the interleaver/de-interlevaer are not taken into account since their influence is found before the application of the FEC and PC decoding process and that the analyzed channel is the AWGN channel model.

From these two tables it can be seen that $n_o(= 220) > k_i(120)$. This means that a "bit regrouping function", q, is implemented; specifically, for Galileo E1 OS, $[\mathbf{w}_{i,1}, \mathbf{w}_{i,2}] = q(\mathbf{c}_o)$, where $\mathbf{w}_{i,1}$ and $\mathbf{w}_{i,2}$ are the two parts of a page. Figure 5 provides a graphical definition of the q where it can be seen that not all the bits of $\mathbf{w}_{i,1}$ and $\mathbf{w}_{i,2}$ belong to $\mathbf{c}_o$: $\mathbf{c}_o$ is constituted of only fields highlighted in red (CRC only protects) - Even/Odd page bits, Page Type bits, Data bits, Reserved 1 bits, SAR bits and CRC bits [5].

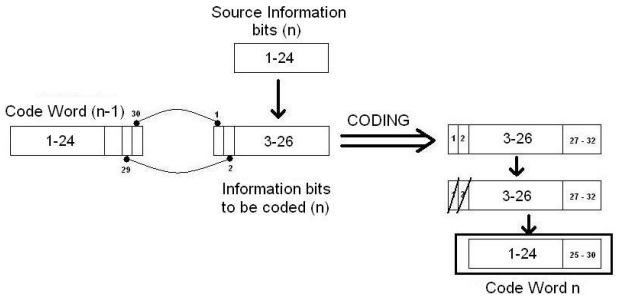

Figure 3 – GPS L1 C/A signal PC encoding process

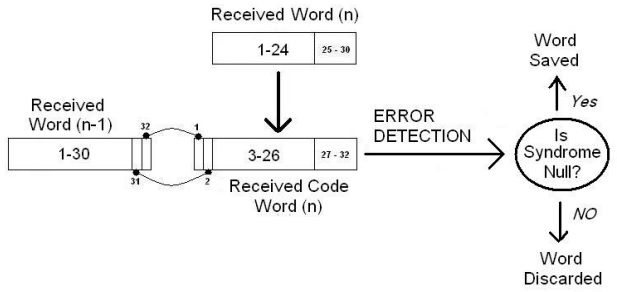

Figure 4 – GPS L1 C/A signal PC decoding process

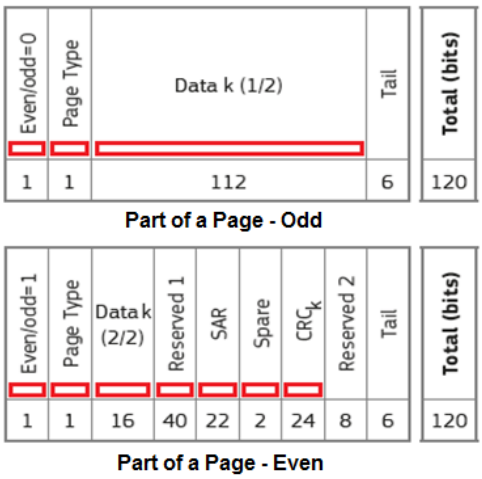

Figure 5 – Galileo E1 OS parts of pages and CRC protected bits

The implemented CRC-24Q of GALILEO E1 OS has the following characteristics [5]:

1) It detects all single bit errors.
2) It detects all double bit error combinations.
3) It detects any odd number of errors.
4) It detects any burst error for which the length of the burst is ≤ 24 bits.
5) It detects most large error bursts with length greater than the parity length r = 24 bits. The fraction of error bursts of length $br > 24$ that are undetected is:
   a) $2^{-24} = 5.96 \times 10^{-8}$, if $b > 25$ bits.
   b) $2^{-23} = 1.19 \times 10^{-7}$, if $b = 25$ bits.

Table I- GPS L1 C/A and Galileo E1 OS signal characteristics

	GPS L1 C/A	Galileo E1 OS
Data component power sharing	1	0.5
Symbol Rate	50 symb/s	250 symb/s
PC Channel Code	Extended Hamming (32,26)	CRC-24Q
(n_o, k_o)	(32,26)	(220,196)
FEC Channel Code	---	Convolutional Code (171,133, 1/2)
(n_i, k_i)	(30,30)	(240,120)

Table II- GPS L1 C/A CED information structure

Word Size	Frame Size	CED Size
30 bits	10 words	3 Frames

Table III- Galileo E1 OS CED information structure

Page Part Size	Page Size	CED Size
240 bits	2 Page Parts	4 Pages

Finally, the control mechanism implemented for Galileo E1 OS is the following one: CED information is only read once and it is provided to the end user if no information is discarded by the FEC verification.

SIMULATION RESULTS

In this section, first the simulation conditions are described. Second, the GPS L1 C/A DIR_{CED} results as well as the customization of DIR_{CED} are presented. Third, the Galileo E1 OS DIR_{CED} results as well as the customization of DIR_{CED} are presented. Fourth and last, a fair comparison between the two signals is presented.

Simulation conditions

The results presented in this paper are calculated from equations (14) and (21). The calculation of equation (14) depends on the parameter num_d^{PC} and on d'_{max}. The value of d'_{max} will be specified for each type of signal. The value of num_d^{PC} has been calculated for each signal using two different approaches:

1) Tight approximation: using equation (12).
2) Loose approximation: using either equation (15) or (18).

In both approaches, simulations have been conducted first by generating all possible $\mathbf{e}_{ic}^{b,d}$ with $d \leq d'_{max}$ and second by generating $\mathbf{e}_{oc}^{b,d} = q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^{b,d}\right)\right)$. Finally, for the first approach, h_{PC} is applied to all $\mathbf{e}_{oc}^{b,d}$ to obtain num_d^{PC}. For the second approach, $P\left(h_{FEC,d^o}\right)$ or $P\left(h_{FEC,br}\right)$ are applied to each corresponding $\mathbf{e}_{oc}^{b,d}$ to obtain num_d^{PC}.

GPS L1 C/A signal

The main difficulty to calculate GPS L1 C/A DIR_w is to take into account on num_d^{PC} calculation the use of the two last bits of epoch t codeword on the FEC encoding process of codeword at epoch $t+1$ [4].

The calculation is thus divided into the calculation of 16 types of codewords depending on whether the epoch t codeword had an error on 0, 1 or 2 bits out of the last two bits (4 cases) and depending on whether the epoch t codeword had an error on 0, 1 or 2 bits out of the last two bits (4 cases).

For GPS L1 C/A signal, d'_{max} was set equal to 30, the size of $\mathbf{w}_i$. Moreover, for the loose approximation, the calculation of num_d^{PC} was obtained by applying equation (16) with:

$$P\left(h_{FEC,d^o}\right) = \begin{cases} 0 & d^o \leq 3 \\ 0.015625 & d^o \geq 4 \end{cases} \qquad (22)$$

Moreover, equation (14) must be modified in order to take into account the second control mechanism analyzed for GPS L1 C/A. In this case, since the only thing to take into account is that the same error vector must appear twice to

be accepted by the control mechanism (if succeeding the PC verification), equation (14) becomes:

$$DIR_w \approx \sum_{d=d_{min}}^{d'_{max}} Q\left(\sqrt{2d\frac{C}{N_0}T_D D}\right)^2 \cdot num_d^{PC} \qquad (23)$$

Figure 6 presents the DIR_{CED} of GPS L1 C/A when the signal is transmitted through an AWGN channel. The DIR_{CED} is presented for the two described control mechanism and its value is calculated using the exact DIR_{CED} expression and a loose approximation.

From Figure 6 it can been observed that the loose approximation sub-estimates the DIR values of GPS L1 C/A but it provides satisfactory results. The reason is that $1/P\left(h_{FEC,d^o}\right) = 0.015625$ is quite low and thus, even for a small $num_{d,br}$, it can be seen that $num_{d,d^o}^e \approx \overline{num}_{d,d^o}$. In fact, the sub-estimation is due that the average value is slightly lower than the true value.

The second observation that can be made is that even for the first type of control mechanism, only one read of the CED information (one set), the $DIR_{CED} \leq DIR_{CA} = 10^{-10}$ for $C/N_0 \geq 24.8$ dB-Hz. Moreover, [3] presents a PLL threshold around 24 dB-Hz for a Civil aviation airplane in normal conditions. Therefore, the necessity of the second type of control mechanism, reading twice the CED information (2 sets) and verifying that their values match, could be questioned from the DIR_{CED} point of view.

Galileo E1 OS signal

The two main difficulties to calculate the GALILEO E1 OS DIR_w are the FEC Convolutional Code and the high number of $\mathbf{e}_{ic}^{b,d}$ to inspect due to the large size of $\mathbf{w}_i$ ($b_{max} = 120$ bits)**.**

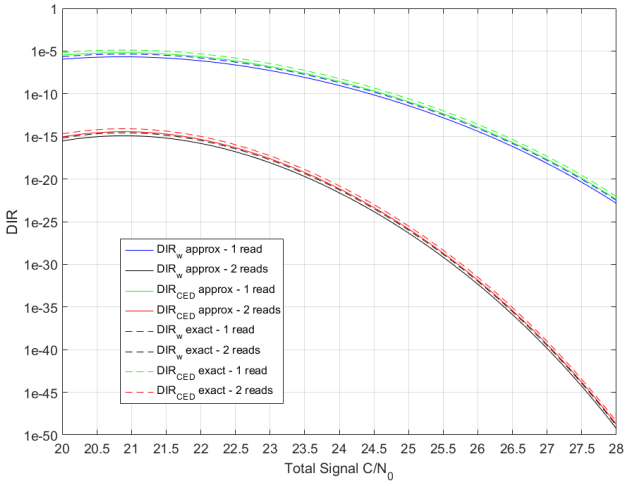

Figure 6 – DIR vs C/N_0 for GPS L1 C/A signal in AWGN channel conditions

First, the formula used in appendix A to calculate the DIR_w is not exactly valid for a convolutional code. In fact, due to the special treillis structure of these type of codes, an exact expression of the probability of a given $\mathbf{e}_{ic}^{b,d}$ cannot be given. However, an overbound of the addition of all the $\mathbf{e}_{ic}^{b,d}$ [6][7] can be found which uses the same principles of appendix A. Therefore, for GALILEO E1 OS, equation (11) becomes (while equations (15) to (21) remain the same):

$$DIR_w \leq \sum_{d=d_{min}}^{d_{max}} Q\left(\sqrt{2d\frac{C}{N_0}T_D D}\right) \cdot num_d^{PC} \qquad (24)$$

Second, in this paper, in order to reduce the number of $\mathbf{e}_{ic}^{b,d}$ to analyze, the following characteristic was used. Due to the special characteristics of a Convolutional Code, $\mathbf{e}_{ic}^{b,d}$ with large burst length, b, and which have at some point 1 bit equal to 1, at least $L-1$ consecutive bits equal to 0 and then another bit equal to 0, $\left[\ldots 1\,\overbrace{0\cdots 0}^{\geq L-1}\,1\ldots\right]$, are considered as two different $\mathbf{e}_{ic}^{b,d}$**;** where the larger $\mathbf{e}_{ic}^{b,d}$ will be denoted as composite error, the shorter $\mathbf{e}_{ic}^{b,d}$ will be denoted as simple errors and L is the FEC constraint length ($L = 7$ for the GALILEO E1 OS FEC channel code). This means that composite error $\mathbf{e}_{ic}^{b,d}$ with very large b but with a low Hamming weigh, d, (and thus error vectors having a high probability of appearing) can be constructed from two simple errors $\mathbf{e}_{ic}^{b,d}$ having shorter b. Therefore, knowing that the FEC convolutional code has a $d_{min} = 10$, the following method was used to generate the different $\mathbf{e}_{ic}^{b,d}$: to combine 1, 2 or 3 simple errors to generate a composite error with simple errors $d'_{max} = 40, 30$ and 16 respectively.

Finally, DIR expression has to be slightly modified to take into account the special structure of a composite error:

$$DIR_w \leq \sum_{d=d_{min}}^{d_{max}} \prod_{v=1}^{v_{comp}} Q\left(\sqrt{2d_v\frac{C}{N_0}T_D D}\right) \cdot num_d^{PC} \qquad (25)$$

Where d_v represents the Hamming weigh of the vth simple error constituting the composite error and v_{comp} is the number of simple errors. Note that a simpler modification consists in just changing in equation (24) the hamming weigh:

$$d = \sum_{v=1}^{v_{comp}} d_v \qquad (26)$$

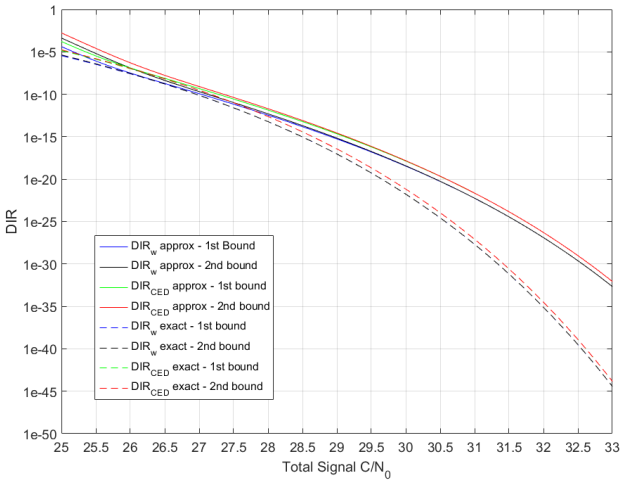

Figure 7 – DIR vs C/N_0 for GALILEO E1 OS signal in AWGN channel conditions

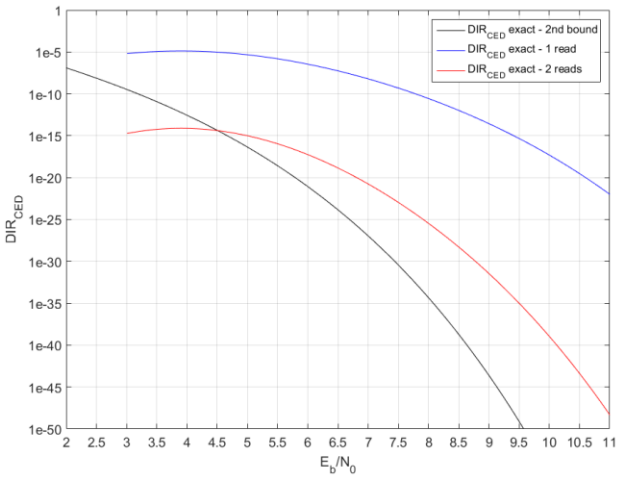

Figure 8 – DIR vs E_b/N_0 comparison between GPS L1 C/A and GALILEO E1 OS signals in AWGN channel conditions

Figure 7 presents the DIR_{CED} of Galileo E1 OS when the signal is transmitted through an AWGN channel. The DIR_{CED} value is calculated using the overbounds presented in equations (24)-(26), denoted as 1st bound, and in (25), denoted as 2ns bound. Moreover, num_d^{PC} was calculated either using the exact formula of equation (12) or the loose approximation of equations (19)-(20). In this last case, $P(h_{FEC,br})$:

$$P(h_{FEC,br}) = \begin{cases} 0 & br \leq 24 \\ 2^{-23} & br = 25 \\ 2^{-24} & br > 25 \end{cases} \quad (27)$$

From Figure 7 it can been observed that the loose approximation over-estimates the DIR values of Galileo E1 OS and that it provides quite average to poor results. The reason is the one stated during the presentation of this approximation: the values of $1/P(h_{FEC,br})$ are very high in comparison with $num_{d,br}$ values for low d (the most probable errors and the ones dominating the DIR_W values).

Therefore, $num_{d,br}^e$ and $\overline{num}_{d,br}$ do not have the same magnitude, especially for the first value different from 0: $num_{d,br}^e \neq 0$ for $d \geq 26$ and $\overline{num}_{d,br} \neq 0$ for $d \geq 16$.

The second observation that can be made is that the results presented for low C/N_0 should be used carefully because the bound presented in equations (24) and (25) is loose for these values [6][7]. In fact, these bounds become tighter to the real value from a $C/N_0 > 27$ dB-Hz. Moroever, it can be seen that the two bounds (either for the loose approximation or for the exact formula) provide the same results except for values between 25-26 dB-Hz.

Finally, it can be observed that for a $C/N_0 \geq 27.2$ dB-Hz, the $DIR_{CED} \leq DIR_{CA} = 10^{-10}$, and thus Galileo E1 OS is compliant with the civil aviation requirement for Data Integrity loss.

Comparison

From the two previous sections, GPS L1 C/A DIR_{CED} and Galileo E1 OS DIR_{CED} thresholds can be compared with a better result for the former signal even when the first control mechanism is used ($C/N_{0,L1\ C/A} = 24.8 > C/N_{0,E1\ OS} = 27.2$). Therefore, the reader could have the wrong impression that GPS L1 C/A contains a better navigation message structure than Galileo E1 OS from the DIR_{CED} point of view when this is not case. In fact, in Figure 8, a fair comparison between GPS L1 C/A DIR_{CED} and Galileo E1 OS DIR_{CED} in AWGN channel conditions is presented as a function of the signal E_b/N_0. DIR_{CED} values are calculated for the exact formula case.

From Figure 8, it can be seen that Galileo E1 OS outperforms GPS L1 C/A when the CED information is read only once. This change of tendency with respect to the DIR_{CED} values expressed as a function of the C/N_0 is due to subtraction of the data component power sharing and the bit rate: Galileo E1 OS has a bit rate 2.5 times higher than GPS L1 C/A and provides 50% of the power to the pilot component. Therefore, it can be concluded that Galileo E1 OS navigation structure is better than GPS L1 C/A from the DIR_{CED} point of view as was expected. Additionally, note that when the second control mechanism is used, GPS L1 C/A only outperforms Galileo E1 OS for $E_b/N_0 \leq 4.5$ dB.

CONCLUSIONS

In this paper, the concept of Data Integrity has been introduced/formalized. The Data Integrity concept was proposed to allow the evaluation/design of a GNSS signal and control mechanism couple. Moreover, this concept allowed to inspect if an existing/designed couple can meet the requirement of a given application to provide the final end user at reception with the same (unaltered) information as the transmitted one.

Three parameters has been proposed to define/quantify the Data Integrity: probability of erroneous message, probability of undetected error and Data Integrity Risk (DIR). The DIR was the parameter used to verify if a GNSS signal/control mechanism design could meet the requirement of a given application by defining a DIR field, DIR_F, associated to the GNNS signal/control mechanism couple and a DIR application, DIR_A.

A DIR value was discussed for Civil Aviation, DIR_{CA}. The value was set to $DIR_{CA} = 10^{-10}$ in order to be integrated inside the Integrity Risk value, $IR_{CA} = 10^{-7}$/operation with a negligible impact. This DIR_{CA} value requirement was set for the Clock offset corrections and Ephemeris satellite Data (CED)

The principle of DIR_W calculation of a generic GNSS signal/control mechanism (NMCM) was presented. An exact formula was derived as well as two approximations, one tight and one loose. Moreover, an expression of the DIR_F was derived from the DIR_W of the words carrying the targeted field. However, these generic formulas have always to be modified to be adapted to the specific characteristics of the targeted NMCM.

The DIR of the CED, DIR_{CED}, for Galileo E1 OS and GPS L1 C/A was calculated. The limitations of the tight and loose calculations were presented for each signal as well as the threshold for which $DIR_{CED} < DIR_{CA}$. For GPS L1 C/A the threshold was equal to 24.8 dB-Hz when the CED information is only read once. Therefore, this result questions the need of applying the current avionics standard control mechanism of reading twice the same exact error free CED information before providing it to the final end user. For Galileo E1 OS the threshold was equal to 27.2 dB-Hz. Nevertheless, GPS L1 C/A only outperforms Galileo E1 OS due to the higher bit rate and the power allocated to pilot component of the latter signal: Galileo E1 OS has a better navigation message structure than GPS L1 C/A from the DIR point of view as was expected.

APPENDIX A

In order to derive the mathematical expression of DIR_w, the first thing to take into account is that PC encoding function, g_{PC}, is bijective. This means that $\mathbf{c}_o^j \underbrace{\leftrightarrow}_{unique} \mathbf{w}_o^j$, where $\mathbf{c}_o^j$ is the outer codeword j of the outer codeword alphabet, $\mathbf{w}_o^j$ is the outer information word j of the outer information word alphabet, $j \in [1, \dots, M_o]$ and $M_o = 2^{k_o}$ is the size of the alphabets. Therefore, equation (7) can be expressed as a function of $\tilde{\mathbf{c}}_o$ and $\mathbf{c}_o$ (using equation (5) and (6)):

$$DIR_w = P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o \cap h_{PC}(\tilde{\mathbf{c}}_o) = 1\right) \qquad \text{(A-1)}$$

$$DIR_w = P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o\right) h_{PC}(\tilde{\mathbf{c}}_o) \qquad \text{(A-2)}$$

Using the law of total probability with the $\mathbf{c}_o$ alphabet, and taking into account the fact that the PC is the tool used to discard $\tilde{\mathbf{c}}_o$ not succeeding its test, expression (A-2) can be expressed as:

$$DIR_w = \sum_{j=1}^{M} P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \mathbf{c}_o = \mathbf{c}_o^j\right) \cdot h_{PC}(\tilde{\mathbf{c}}_o) \qquad \text{(A-3)}$$

Where $P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \mathbf{c}_o = \mathbf{c}_o^j\right)$ represent the probability that the outer codeword $\mathbf{c}_o^j$ was transmitted and that the receiver estimated an outer codeword different from $\mathbf{c}_o^j$.

Note that $P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \mathbf{c}_o = \mathbf{c}_o^j\right)$ represent the probability of an erroneous message, P_e, and that $h_{PC}(\tilde{\mathbf{c}}_o)$ represents the probability of an undetected error, P_{und}.

Expression (A-3) can be further developed if the value of $\tilde{\mathbf{c}}_o$ when $\tilde{\mathbf{c}}_o \neq {\mathbf{c}_o}^j$, $\mathbf{c}_o = {\mathbf{c}_o}^j$ is detailed. To do that, the PC verification inability to detect that $\tilde{\mathbf{c}}_o \neq \mathbf{c}_o$ must be closely inspected. In fact, PC can only determine if $\tilde{\mathbf{c}}_o$ belong to the outer codeword alphabet, $\tilde{\mathbf{c}}_o = \mathbf{c}_o^j, j \in [1, \dots, M_o]$, but it cannot determine if the estimated outer codeword, $\tilde{\mathbf{c}}_o$, is the same as the transmitted one, $\mathbf{c}_o$ [7]. Therefore, PC verification test, h_{PC}, can be further mathematically detailed as presented in equation (A-4) when assuming that $\mathbf{c}_o^j$ was transmitted, $\mathbf{c}_o = \mathbf{c}_o^j$:

$$h_{PC}(\tilde{\mathbf{c}}_o) = \begin{cases} 1 & \tilde{\mathbf{c}}_o = \mathbf{c}_o^j \\ 0 & \tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \tilde{\mathbf{c}}_o \neq \mathbf{c}_o^u \\ 1 & \tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \tilde{\mathbf{c}}_o = \mathbf{c}_o^u \end{cases} \quad u \in [0..M_o - 1] \qquad \text{(A-4)}$$

As well as in (5), the third line represents the conditions for the data integrity loss. This means that $P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \mathbf{c}_o = \mathbf{c}_o^j\right)$ can be expressed as:

$$\begin{aligned} &P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o^j, \mathbf{c}_o = \mathbf{c}_o^j\right) \\ &= P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o, \tilde{\mathbf{c}}_o \neq \mathbf{c}_o^b, \mathbf{c}_o = \mathbf{c}_o^j\right) \\ &\quad + \sum_{\substack{u=1 \\ u \neq j}}^{M} P\left(\tilde{\mathbf{c}}_o = \mathbf{c}_o^u, \mathbf{c}_o = \mathbf{c}_o^j\right) \end{aligned} \qquad \text{(A-5)}$$

Where, $\mathbf{c}_o^b$ is the outer codeword b of the outer codeword alphabet, $b \in [1, \dots, M_o]$, and $P\left(\tilde{\mathbf{c}}_o \neq \mathbf{c}_o, \tilde{\mathbf{c}}_o \neq \mathbf{c}_o^b, \mathbf{c}_o = \mathbf{c}_o^j\right)$ is the probability that the outer codeword $\mathbf{c}_o^j$ was transmitted and that the receiver estimated as the outer codeword a vector different from any $\mathbf{c}_o^b$.

Therefore, after the multiplication by $h_{PC}(\tilde{\mathbf{c}}_o)$, equation (A-3) becomes

$$DIR_w = \sum_{j=1}^{M}\sum_{\substack{u=1\\u\neq j}}^{M} P\left(\tilde{\mathbf{c}}_{\boldsymbol{o}} = \mathbf{c}_o^u, \mathbf{c}_{\boldsymbol{o}} = \mathbf{c}_o^j\right) \quad \text{(A-6)}$$

This formula can be expressed as a function of the estimated inner codewords, $\tilde{\mathbf{c}}_i$. In this case, it is assumed for simplifications purposes that $P = 1$ and that $\mathbf{c}_o = \mathbf{w}_i, n_o = k_i$. However, the total number of possible inner information words allowed by the FEC, $M_i = 2^{k_i}$, is larger than the number of outer codewords, $M_o = 2^{k_o}$, generated by the PC. Therefore, only a subset of possible inner information words represent all the outer codewords $\mathbf{c}_o^j \subset \mathbf{w}_i^u$ with $j \in [1, \dots, M_o], u \in [1, \dots, M_i]$ and $M_o < M_i$. Therefore, denoting $u(j)$ as a bijective function between the M_o elements of the set $j \in [1, \dots, M_o]$ and a subset of M_o elements of the set $u \in [1, \dots, M_i]$, denoting $\mathbf{w}_i^{u(j)} \equiv \mathbf{c}_o^j$ as the inner information word $u(j)$ which is equal to the outer codeword j, and taking into account that the FEC encoder function, g_{FEC}, is also bijective, $\mathbf{c}_i^j \underbrace{\leftrightarrow}_{unique} \mathbf{w}_i^j$ and thus $\mathbf{c}_i^{u(j)} \underbrace{\leftrightarrow}_{unique} \mathbf{w}_i^{u(j)}$, expression (A-3) becomes:

$$DIR_w = \sum_{j=1}^{M_o}\sum_{\substack{v=1\\v\neq j}}^{M_i} P\left(\tilde{\mathbf{c}}_{\boldsymbol{i}} = \mathbf{c}_i^{u(v)}, \mathbf{c}_{\boldsymbol{i}} = \mathbf{c}_i^{u(j)}\right) \quad \text{(A-7)}$$

Then, using $P(A,B) = P(A|B)P(B)$, the expression becomes:

$$DIR_w = \sum_{j=1}^{M_o}\sum_{\substack{v=1\\v\neq j}}^{M_o} P\left(\tilde{\mathbf{c}}_{\boldsymbol{i}} = \mathbf{c}_i^{u(v)} \middle| \mathbf{c}_{\boldsymbol{i}} = \mathbf{c}_i^{u(j)}\right) \cdot P\left(\mathbf{c}_{\boldsymbol{i}} = \mathbf{c}_i^{u(j)}\right) \quad \text{(A-8)}$$

Besides, it is possible to express $P\left(\tilde{\mathbf{c}}_{\boldsymbol{i}} = \mathbf{c}_i^{\boldsymbol{u(v)}} \middle| \mathbf{c}_{\boldsymbol{i}} = \mathbf{c}_i^{u(j)}\right)$ as a function of the signal characteristic and the FEC channel code properties [6][7]:

$$P\left(\tilde{\mathbf{c}}_{\boldsymbol{i}} = \mathbf{c}_i^{u(v)} \middle| \mathbf{c}_{\boldsymbol{i}} = \mathbf{c}_i^{u(j)}\right) = Q\left(\sqrt{2d_{u(j)u(v)}\frac{C}{N_0}T_D D}\right) \quad \text{(A-9)}$$

Where $d_{u(j)u(v)}$ is the Hamming distance between $\mathbf{c}_i^{u(j)}$ and $\mathbf{c}_i^{u(v)}$, T_D is the symbol interval, D is the percentage of power provided to the GNSS signal data component and $Q(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$.

Finally, taking into account that $\mathbf{w}_o^j$ with $j \in [1, \dots, M_o]$ are equiprobable, which implies $P\left(\mathbf{c}_{\boldsymbol{i}} = \mathbf{c}_i^{u(j)}\right) = 1/M_o$ expression (A-8) can be expressed as:

$$DIR_w = \frac{1}{M_o}\sum_{j=1}^{M_o}\sum_{\substack{v=1\\v\neq j}}^{M_o} Q\left(\sqrt{2d_{u(j)u(v)}\frac{C}{N_0}T_D D}\right) \quad \text{(A-10)}$$

APPENDIX B

In order to derive an overbound of expression (A-10), its summations are going to be extended.

First, the summation referring to the possible values of $\tilde{\mathbf{c}}_{\boldsymbol{o}}$ is going to be extended from only the inner codewords associated to an outer codeword, $\mathbf{c}_i^{u(v)}$, to all the possible codewords, $\mathbf{c}_i^b$, but taking into account if the generated estimated outer codeword, $\tilde{\mathbf{c}}_{\boldsymbol{o}} = q^{-1}f_{FEC}\left(\mathbf{c}_i^b\right)$, verifies the PC test, h_{PC}:

$$DIR_w = \frac{1}{M_o}\sum_{j=1}^{M_o}\sum_{\substack{b=1\\b\neq u(j)}}^{M_i} Q\left(\sqrt{2d_{u(j)b}\frac{C}{N_0}T_D D}\right) \cdot h_{PC}\left(q^{-1}f_{FEC}\left(\mathbf{c}_i^b\right)\right) \quad \text{(B-1)}$$

Second, a first overbound is derived by adding all the possible codewords, $\mathbf{c}_i^{b\prime}$, to the possible $\mathbf{c}_i^{u(j)}$ which could be generated by each possible $\mathbf{c}_o^j$**:**

$$DIR_w < \frac{1}{M_o}\sum_{b\prime=1}^{M_i}\sum_{\substack{b=1\\b\neq b\prime}}^{M_i} Q\left(\sqrt{2d_{b\prime b}\frac{C}{N_0}T_D D}\right) \cdot h_{PC}\left(q^{-1}f_{FEC}\left(\mathbf{c}_i^b\right)\right) \quad \text{(B-2)}$$

Expression (B-2) can be further simplified by taking into account that FEC and PC channel codes are linear codes, which means that the hamming distance distribution of all the codewords is the same [7]. Therefore, the DIR_w analysis can be reduced, without loss of generality, to just the case where the null outer information word, $\mathbf{w}_{\boldsymbol{o}} = \mathbf{0}$, was transmitted (note that the extra inner codewords added in (B-2) have been removed):

$$DIR_w = \sum_{b=2}^{M_i} Q\left(\sqrt{2d_{b,\mathbf{0}}\frac{C}{N_0}T_D D}\right) \cdot h_{PC}\left(q^{-1}\left(f_{FEC}\left(\mathbf{c}_i^b\right)\right)\right) \quad \text{(B-3)}$$

Where $\mathbf{c}_i^{b=1}$ is assumed to be the $\mathbf{0}$ inner codeword generated from $\mathbf{w}_{\boldsymbol{o}} = \mathbf{0}$, and $d_{b,\mathbf{0}}$ is the Hamming distance between $\mathbf{c}_i^b$ and $\mathbf{c}_i^{b=1} = \mathbf{0}$.

Using the fact that the estimated inner codeword can be modelled as the modulo-2 addition between the true inner

codeword and an error vector, when assuming $\mathbf{w_o} = \mathbf{0}$ equation (9) becomes:

$$\tilde{\mathbf{c}}_i = \mathbf{e}_{ic} \qquad \widetilde{\mathbf{w}}_i = \mathbf{e}_{iw} \tag{B-4}$$

Therefore:

$$DIR_w < \sum_{b=2}^{M_i} Q\left(\sqrt{2d_b \frac{C}{N_0} T_D D}\right) \cdot h_{PC}\left(q^{-1}\left(f_{FEC}\left(\mathbf{e}_{ic}^b\right)\right)\right) \tag{B-5}$$

Where, $\mathbf{e}_{ic}^b$ is the outer codeword error vector b, $b \in [1, \ldots, M_i]$, and d_b is the Hamming weight (number of ones) of the error vector $\mathbf{e}_{ic}^b$.

REFERENCES

[1] ICAO (2010) Annex 10, Aeronautical Telecommunications, Volume 1 (Radio Navigation Aids), Amendment 86, effective 17 November 2011. GNSS standards and recommended practices (SARPs) are contained in Section 3.7 and subsections, Appendix B, and Attachment D.

[2] RTCA DO-229D, Minimum Operational Performance Standards for Global Positioning System/Wide Area Augmentation System Airborne Equipment, December 13, 2006.

[3] RTCA DO-235B Assessment of Radio Frequency Interference Relevant to the GNSS L1 Frequency Band, March 13, 2008.

[4] ARINC Engineering Services, "Navstar GPS space Segment/Navigation User segment interfaces, IS-GPS-200F", September 21, 2011.

[5] European Space Agency, "Galileo OS SIS ICD Issue 1.2", November 2015.

[6] J.G Proakis and M.Salehi, "Digital Communications" 5th ed, McGraw-Hill, 2008.

[7] Shu Lin and Daniel J. Costello Jr, "Error Control Coding 2nd edition", Pearson Prentice Hall, 2004

[8] M. Cover and Joy A. Thomas, "Elements of Information Theory 2nd Edition", Thomas, Wiley-Interscience, 2006.