Non-coherent rate-splitting for multibeam satellite forward link: practical coding and decoding algorithms

N. Noels⋆ M. Moeneclaey⋆ T. Ramírez† C. Mosquera† M. Caus$ A. Pastore$

⋆Digital communications and information processing group, Ghent University, Ghent, Belgium

†atlanTTic research center, University of Vigo, Galicia, Spain

$Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, Barcelona, Spain

***Abstract*—Non-Coherent Rate-Splitting (NCRS) was recently proposed as a practical multiuser coding and decoding scheme to increase the spectral efficiency of multibeam satellite communication systems. In this paper, we further study the practical realization of NCRS. We propose a modified coding scheme (NCRS*) that is robust to a nonzero time offset among beams. In NCRS*, as opposed to NCRS, the beams send independently channel encoded and modulated waveforms.**

We asses the performance of NCRS* in terms of the achievable rate region. It is shown that NCRS* performs worse than NCRS, but better than or comparable to other competing schemes, which, as opposed to NCRS*, require flexible bandwidth allocation or perfect synchronization at the transmitter. We also propose a new N-MAP algorithm for the practical implementation of NCRS* receivers. Similar to the existing U-MAP algorithm, N-MAP takes into account the modulation used by, and the time offset between, the signals received from the different beams. In most cases, however, N-MAP has a significantly lower complexity than U-MAP.

I. Introduction

Conventional multibeam satellite communication systems employ a conservative frequency reuse pattern that allows to simply ignore the inter-beam co-channel interference (CCI) at the receiver. However, in search of novel technologies to meet the very high throughput demands of the integrated satellite-terrestrial communication networks of the future (such as 5G), the satellite community is looking at multibeam systems in which adjacent spot beams use the same frequency band and polarization. In such systems, CCI is a major issue [1].

Recently, a novel CCI management technique for multibeam satellite down-link communication, called Non-Coherent Rate-Splitting (NCRS), has been proposed [2]–[5]. A key feature of NCRS is that, as opposed to precoding, no channel phase information is required at the transmitter. NCRS is a non-orthogonal multiuser transmission scheme, in which multiple co-channel signals cooperate to simultaneously serve a group of users. To this end the transmitted messages are split into private and public components, and successive cancellation decoding (SCD) [6] is applied at the receive terminals to extract messages that use the same physical resources. By means of a theoretically achievable rate analysis, NCRS has been shown to offer increased spectral efficiency when compared to competing schemes [3]. However, the practical implementation of NCRS has not yet been made very concrete. The purpose of the current paper is to address this concern, while specifically taking into account that the signals received from different satellite beams are usually not perfectly synchronized in time.

First, we propose a new practical variant of the NCRS coding scheme, further referred to as NCRS*. In NCRS*, separate channel coding and modulation for each beam is considered. This is interesting because the corresponding transmitter does not have to take into account the difference in timing between the beams. Moreover, it allows to use a simpler receiver structure. To asses the implication of using NCRS* on the system throughput, we derive the theoretically achievable rate region for NCRS* and compare it to the results from [2]. In [2], the achievable rate region for NCRS was derived and compared to competing schemes. Our analysis indicates that using NCRS* rather than NCRS may result in a significant performance degradation for specific channel magnitude values, but can still be expected to yield an average system performance that is better than, or comparable to, other competing schemes.

As a second contribution, we propose a novel N-MAP algorithm to perform the SCD in NCRS* receivers. In the literature there is only a limited amount of related work. A practical receiver for the optimal detection of a desired signal in the presence of a symbol asynchronous CCI signal was first considered in [7] assuming uncoded BPSK and rectangular modulation pulses. More recently, a turbo receiver architecture for quasi-optimal joint maximum-a-posteriori (MAP) co-channel detection was studied in [8]; here, the multiuser detector and the single user decoders are treated as separate concatenated blocks that iteratively exchange soft information. In [9], an approximate joint MAP co-channel detector for interfering multipath transmissions was developed using the factor graph (FG) and sum-product algorithm (SPA) framework. Neither [8] or [9] consider a receiver that is only interested in a subset of the co-channel signals, as is the case with NCRS*. The general term for this concept is non-unique decoding (NUD). In [10], a practical FG-based implementation of a receiver with NUD of low-density parity-check coded co-channel signals is employed but in this work only synchronous co-channel reception is considered and the focus is on code design rather than on implementation complexity.

The outline of the paper is as follows. Section II briefly reviews the NCRS basics and presents the proposed NCRS* scheme. Section III derives and analyzes the achievable rate region of NCRS*. Section IV proposes algorithms for the practical implementation of NCRS*. Finally, Section V summarizes the main conclusions.

In the following, small bold letters denote row vectors, the kth element of $\boldsymbol{a}$ is $a[k]$, capital bold letters denote matrices and the (k,l)th element of $\boldsymbol{A}$ is $A_{k,l}$. Furthermore, $\overset{a}{\propto}$ denotes equality within a constant not depending on $\boldsymbol{a}$ and a function is said to be $O(a(n))$ if it is bounded below by $a(n)$ asymptotically for large n.

II. From NCRS to NCRS*

Considered is a system that serves one user per beam cell using the same physical resources. Only channel *magnitude* information is available at the transmitter and there is a per-beam power constraint. To limit the complexity, the system is divided into 2-beams 2-users subsystems and a CCI management technique is applied to each subsystem separately. The relevant channel can be categorized as an equivalent Gaussian multiple-input single-output (MISO) broadcast (BC) channel with two inputs (X_1, X_2) and two outputs (Y_1, Y_2). We use the convention that the message communicated to the user observing Y_i is M_i.

We first review the CCI management technique proposed in [2]–[5] and referred to as NCRS. NCRS operates in a time sharing manner. There are two operation modes OM 1 and OM 2. Each OM i is in effect for a fraction α_i of the time, with $\alpha_1 = 1-\alpha_2 = \alpha$ and $\alpha \in [0,1]$. In both OMs, rate splitting (RS) is performed to create a triplet of independent messages $\left(M_{1_p}, M_c, M_{2_p}\right)$. In OM 1, M_1 is split to create M_{1_p} and M_c, while M_2 equals M_{2_p}. In OM 2, it is just the other way around: M_2 is split into M_{2_p} and M_c, while M_1 equals M_{1_p}. The messages M_{1_p} and M_{2_p} are encoded into X_{1_p} and X_{2_p}, respectively, with $\mathbb{E}\left[\left|X_{1_p}\right|^2\right] = \mathbb{E}\left[\left|X_{2_p}\right|^2\right] = 1$. Further, M_c is encoded into a symbol vector $\boldsymbol{X}_c = \left[X_{1_c}\, X_{2_c}\right]^T$, with $\mathbb{E}\left[\left|X_{1_c}\right|^2\right] = \mathbb{E}\left[\left|X_{2_c}\right|^2\right] = 1$. Finally, superposition coding [6] is applied to transmit the symbol pairs $\left(X_{1_p}, X_{1_c}\right)$ and $\left(X_{2_p}, X_{2_c}\right)$ through beams 1 and 2, respectively. We have: $X_\beta = \sqrt{\dot{\lambda}_\beta} X_{\beta_c} + \sqrt{\lambda_\beta} X_{\beta_p}$, for $\beta = 1$ or $\beta = 2$, where $\dot{\lambda}_\beta = 1-\lambda_\beta$ and $\lambda_\beta \in [0,1]$ is a power allocation factor.

The signal observed at user β is $Y_\beta = h_{\beta,1}X_1 + h_{\beta,2}X_2 + W_\beta$, where $W_\beta \sim \mathcal{CN}\left(0, \sigma_\beta^2\right)$ is the thermal noise at receiver β and $\begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{bmatrix}$ is the complex valued channel matrix. For further use we define the vector $\boldsymbol{\Gamma} = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{2,1}, \gamma_{2,2})$, with$\gamma_{\beta,\beta'} = \left|h_{\beta,\beta'}\right|^2/\sigma_\beta^2$. Receivers 1 and 2 perform simultaneous NUD to recover M_1 and M_2 from Y_1 and Y_2, respectively. The practical decoding strategy proposed in [3] is the following. To recover M_β, receiver β performs *two-stage* SCD. Upon receiving Y_β, the receiver recovers M_c treating M_{1_p} and M_{2_p} as part of the noise. The receiver then subtracts $h_{\beta,1}\sqrt{\dot{\lambda}_1}X_{1_c} + h_{\beta,2}\sqrt{\dot{\lambda}_2}X_{2_c}$ from Y_β and decodes the result to recover M_{β_p} treating $M_{(3-\beta)_p}$ as part of the noise.

As far as the communication of M_c is concerned, the channels to users 1 and 2 can be considered as equivalent MISO channels with two transmit antennas and one receive antenna, incomplete channel state information at the transmitter (CSIT) and a per-antenna power constraint. It is well-established that, if cross-antennas coding is applied to transmit M_c, this can significantly increase the reliability (For the considered [11]). A well-known example is Alamouti space-time block coding, which does not require CSIT and which is known to be capacity achieving for two transmit antennas and one receive antenna [12], [13]. However, an important disadvantage of cross-antenna coding is that it usually relies on perfect antenna synchronization which is difficult to realize in the multibeam satellite context [14], [15].

In [16], it was shown that antennas that simply send independent symbols at each time instant to a single-antenna receiver do not incur a loss in single-user capacity under perfect CSIT[1]. This inspires us to propose an alternative NCRS-type scheme, further referred to as NCRS*. The main difference between NCRS and NCRS*, is that in NCRS*, M_c is further split into independent components M_{1_c} and M_{2_c} that are separately encoded into X_{1_c} and X_{2_c}, respectively. This modification is the key to increase the robustness to time offsets and paves the way to apply separate decoding of M_{1_c} and M_{2_c} with soft interference cancellation.

NCRS* resembles the Han-Kobayashi (HK) scheme, which yields the best known inner bound on the capacity of a single-input single-output (SISO) interference channel [6]. The main difference is the coordination at the transmitter. With NCRS* it is possible to transmit the messages M_1 and M_2 through both beams; this is not allowed with HK.

III. Achievable Rate Region

The values of the power allocation parameters λ_1 and λ_2 and the value of the time-sharing parameter α determine the rates r_1 and r_2 that can be allocated to users 1 and 2, respectively. A collection of rate pairs (r_1, r_2) that can be theoretically achieved for given $\boldsymbol{\Gamma} = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{2,1}, \gamma_{2,2})$ is referred to as an achievable rate region.

Achievable rate regions for NCRS were first presented in [2]. To derive them point-to-point Gaussian channel coding was assumed for $X_{1_p}$and X_{2_p} and a Gaussian 2x1 MISO channel capacity achieving coding scheme was considered for $\boldsymbol{X}_c$ [12], [13]. The resulting achievable rate region consists of the convex hull of the union of the sets

$$\left\{(r_1, r_2);\; \begin{array}{c} 0 \le r_1 \le R_1(\lambda_1, \lambda_2, \alpha), \\ 0 \le r_2 \le R_2(\lambda_1, \lambda_2, \alpha) \end{array}\right\} \tag{1}$$

over all $(\lambda_1, \lambda_2, \alpha)$ in $[0,1]^3$, where

$$R_\beta(\lambda_1, \lambda_2, \alpha) = R_{\beta_p}(\lambda_1, \lambda_2) + \alpha_\beta R_c(\lambda_1, \lambda_2), \tag{2}$$

[1]With more than one single-antenna receiver (multicast) and fixed channel magnitudes known at the transmitter, the capacity region is reduced, as we show further on.

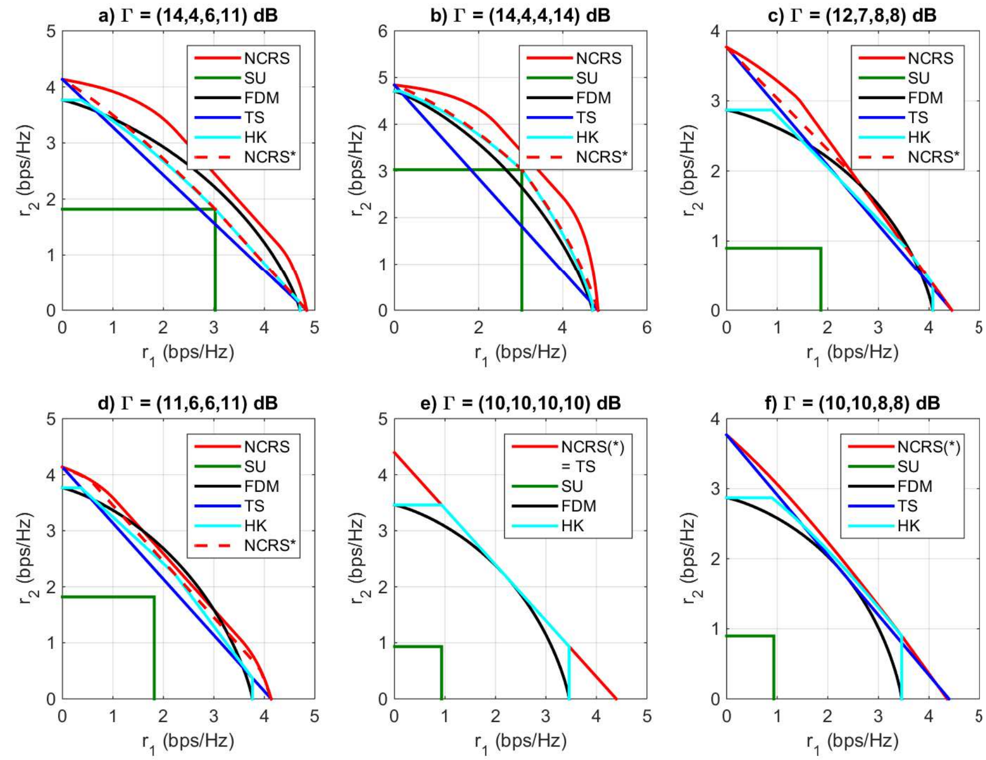

Figure 1. Achievable rate region for NCRS, NCRS*, HK, FDM, TS and SU for different values of $\mathbf{\Gamma} = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{2,1}, \gamma_{2,2})$.

with $\alpha = \alpha_1 = 1 - \alpha_2$ and

$$R_{\beta_p}(\lambda_1, \lambda_2) = \log_2\left(1 + \frac{\lambda_\beta \gamma_{\beta,\beta}}{1 + \lambda_{3-\beta}\gamma_{\beta,3-\beta}}\right), \quad (3)$$

$$R_c(\lambda_1, \lambda_2) = \min_{\beta \in \{1,2\}} \log_2\left(1 + \frac{\dot{\lambda}_1 \gamma_{\beta,1} + \dot{\lambda}_2 \gamma_{\beta,2}}{1 + \lambda_1 \gamma_{\beta,1} + \lambda_2 \gamma_{\beta,2}}\right). \quad (4)$$

Here, R_{β_p} denotes the maximum rate at which M_{β_p} can be reliably transmitted to user β and R_c is the maximum rate at which M_c can be reliably transmitted to both users. Rc is confined in the intersection of two simultaneous decoding rate regions. The expression for R_c tacitly assumes SCD in two stages at user β.

We now compute achievable rate regions for NCRS*. For this, point-to-point Gaussian channel coding is assumed for X_{1_p}, X_{2_p}, X_{1_c} and X_{2_c}. Moreover, the receivers are assumed to apply a *three-stage* SCD strategy, in which the two common message parts are both recovered prior to the private message. The order in which the common messages M_{1_c} and M_{2_c} are recovered does not have to be the same in both receivers. We distinguish four possible strategies.

- Both receivers decode M_{1_c} first.
- Receiver β decodes M_{β_c} first, with $\beta \in \{1, 2\}$.
- Receiver β decodes $M_{(3-\beta)_c}$ first, with $\beta \in \{1, 2\}$.
- Both receivers decode M_{2_c} first.

Decoded messages are removed from the observation prior to starting the subsequent decoding stage. Messages that have not yet been decoded are treated as additional noise. A schematic view of receiver i, decoding M_{j_c} first, is provided in Fig. 2. The corresponding rate region is the convex hull of the union of (1) over all $(\lambda_1, \lambda_2, \alpha)$ in $[0, 1]^3$, with R_β as in (2) and R_{β_p} as in (3), but with R_c given by (5) rather than (4). We have

$$R_c = \max_{j \in \{1,2,3,4\}} R_{c,j}, \quad (5)$$

with

$$R_{c,1} = \min\left(R_{1_c,1}, R_{1_c,2}\right) + \min\left(\mathring{R}_{2_c,1}, \mathring{R}_{2_c,2}\right),$$

$$R_{c,2} = \min\left(R_{1_c,1}, \mathring{R}_{1_c,2}\right) + \min\left(\mathring{R}_{2_c,1}, R_{2_c,2}\right),$$

$$R_{c,3} = \min\left(\mathring{R}_{1_c,1}, R_{1_c,2}\right) + \min\left(R_{2_c,1}, \mathring{R}_{2_c,2}\right),$$

$$R_{c,4} = \min\left(\mathring{R}_{1_c,1}, \mathring{R}_{1_c,2}\right) + \min\left(R_{2_c,1}, R_{2_c,2}\right),$$

where

$$R_{\beta_c,i} = \log_2\left(1 + \dot{\lambda}_\beta \gamma_{i,\beta} / \left(1 + \lambda_\beta \gamma_{i,\beta} + \gamma_{i,3-\beta}\right)\right),$$

$$\mathring{R}_{\beta_c,i} = \log_2\left(1 + \dot{\lambda}_\beta \gamma_{i,\beta} / \left(1 + \lambda_1 \gamma_{i,1} + \lambda_2 \gamma_{i,2}\right)\right).$$

In a number of special cases, it is possible to further simplify (5) analytically. For example, it is not difficult to

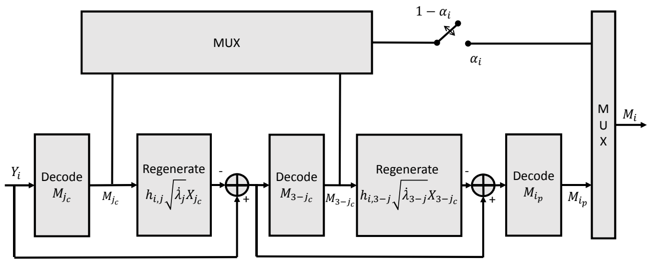

Figure 2. Block diagram of receiver i, decoding M_{j_c} first.

show that (5) equals (4), for $\lambda_1 = 1$ or $\lambda_2 = 1$. The same holds for showing that (5) equals (4) for any (λ_1, λ_2), if $(\gamma_{1,1} = \gamma_{2,1}, \gamma_{1,2} = \gamma_{2,2})$ or $(\gamma_{1,1} = \gamma_{1,2}, \gamma_{2,1} = \gamma_{2,2})$. [2] On the other hand, it can also be shown that, if $(\gamma_{1,1} = \gamma_{2,2} = G, \gamma_{1,2} = \gamma_{2,1} = g)$ with $G > g + g^2$, (5) is strictly smaller than (4), for any $1 > \lambda_1 \geq 0$ and $1 > \lambda_2 \geq 0$.[3]

Fig. 1 shows numerically evaluated achievable rate regions of NCRS, NCRS* and other relevant multiuser transmission schemes, for various $\mathbf{\Gamma}$. Here, SU refers to the single user approach: the receiver considers all interference (including the interference from within the beam pair (1,2)) as additional noise, TS refers to basic time sharing: both co-channel signals are used to transmit M_i, during a fraction $\chi_i \in [0,1]$ (with $\chi_1 + \chi_2 = 1$) of the time, and FDM stands for frequency division multiplexing: the paired users are simultaneously served in non-overlapping frequency bands, whereby the transmission of M_i uses only a fraction ϵ_i (with $\epsilon_1 + \epsilon_2 = 1$) of the bandwidth. For more details on these schemes and the corresponding achievable rate region expressions, see [2]. Denoting the achievable rate region of scheme X as $\mathcal{R}(\mathrm{X})$, we make the following observations:

- Because independent per-antenna coding is sub-optimal, $\mathcal{R}(\mathrm{NCRS}*) \subseteq \mathcal{R}(\mathrm{NCRS})$.
- Because HK does not allow to transmit M_1 and M_2 via both beams, $\mathcal{R}(\mathrm{HK}) \subseteq \mathcal{R}(\mathrm{NCRS}*)$. The maximum achievable sum-rate is the same for HK and NCRS*.
- In all considered scenarios, $\mathcal{R}(\mathrm{SU}) \subseteq \mathcal{R}(\mathrm{NCRS}*)$ and $\mathcal{R}(\mathrm{TS}) \subseteq \mathcal{R}(\mathrm{NCRS}*)$.
- For $\mathbf{\Gamma}$ = (14,4,6,11) dB, $\mathcal{R}(\mathrm{FDM}) \subseteq \mathcal{R}(\mathrm{NCRS})$ but $\mathcal{R}(\mathrm{FDM}) \nsubseteq \mathcal{R}(\mathrm{NCRS}*)$.
- For $\mathbf{\Gamma}$ = (12,7,8,8) dB and $\mathbf{\Gamma}$ = (11,6,6,11) dB, $\mathcal{R}(\mathrm{FDM}) \nsubseteq \mathcal{R}(\mathrm{NCRS})$.
- For $\mathbf{\Gamma}$ = (14,4,4,14) dB, $\mathbf{\Gamma}$ = (10,10,10,10) dB and $\mathbf{\Gamma}$ = (10,10,8,8) dB, $\mathcal{R}(\mathrm{FDM}) \subseteq \mathcal{R}(\mathrm{NCRS}*)$.

Here, $\mathcal{R}(\mathrm{X}) \subseteq \mathcal{R}(\mathrm{Y})$ ($\mathcal{R}(\mathrm{X}) \nsubseteq \mathcal{R}(\mathrm{Y})$) indicates that $\mathcal{R}(\mathrm{X})$ is (is not) a subset of $\mathcal{R}(\mathrm{Y})$. Taking into account that in a realistic multibeam communication satellite system there is a multitude of user pairs (i,j) with a large variety of channel magnitude conditions $(\gamma_{i,i}, \gamma_{i,j}, \gamma_{j,i}, \gamma_{j,j})$, our results (including a large amount of simulations not reported here) indicate that NCRS* can provide an average system performance that is obviously worse than that of NCRS, but better than TS and HK, and at least comparable to that of FDM. It should further be noted that NCRS* does not require flexible bandwidth allocation (as opposed to FDM) nor perfect synchronization at the transmitter (as opposed to NCRS and TS). Moreover, just as in the case of plain NCRS, in NCRS*, the parameter α can be used to modulate the contribution of the common message to each user. This allows a seamless rate allocation which can be tailored to the user demands while avoiding time-multiplexing of encoding schemes and the corresponding synchronization needs at the receivers. The value α can be changed across time without increasing the complexity of the transmit and receive architectures.

IV. Practical implementation and algorithms

A. Coding and modulation

In practice, NCRS* shall be implemented with good practical point-to-point coded modulation schemes rather than Gaussian coding. Each message M_q, $q \in Q = \{1_c, 2_c, 1_p, 2_p\}$, is represented by a stream of information words and, for each $q \in Q$, every information word $\boldsymbol{b}_q$ is encoded into a codeword $\boldsymbol{x}_q$ at rate r_q. Thus, $r_1 = r_{1_p} + \alpha(r_{1_c} + r_{2_c})$ and $r_2 = r_{2_p} + (1-\alpha)(r_{1_c} + r_{2_c})$. The symbols $x_q[k]$ take values in a discrete alphabet A_q of size $|A_q| \equiv \mathcal{X}_q$ with $\sum_{x \in \mathcal{A}_q} |x_q[k]|^2 = 1$. The code constraints determine the joint probability mass function (pmf) $P_q(\boldsymbol{b}_q, \boldsymbol{x}_q)$ of $\boldsymbol{b}_q$ and $\boldsymbol{x}_q$, $q \in Q$. The codewords $\boldsymbol{x}_{1_c}$, $\boldsymbol{x}_{2_c}$, $\boldsymbol{x}_{1_p}$ and $\boldsymbol{x}_{2_p}$ are used to generate the waveforms $s_1(t)$ and $s_2(t)$, transmitted by beams 1 and 2, respectively; we have

$$s_\beta(t) = \sum_k x_\beta[k]\, p(t - kT_s). \tag{6}$$

Here, T_s denotes the symbol period, $p(t)$ is a real-valued even square-root Nyquist pulse with $\int p(u)\, p(u - mT_s)\, \mathrm{d}u = \delta_m$ and

$$x_i[k] = \sqrt{\dot{\lambda}_i} x_{i_c}[k] + \sqrt{\lambda_i} x_{i_p}[k]. \tag{7}$$

To avoid inter-codeword interference, a guard interval is assumed between subsequent codewords. As a result, we can further consider single codeword transmission.

B. Channel

The signal $y_\beta(t)$ received at user β can be modeled as:

$$y_\beta(t) = w_\beta(t) + u_\beta(t), \tag{8}$$

with $w_\beta(t)$ a standard normal random process and

$$u_\beta(t) = \sqrt{\gamma_{\beta,1}} s_1(t) + e^{j\phi_\beta} \sqrt{\gamma_{\beta,2}} s_2(t - \tau_\beta), \tag{9}$$

[2] $(\gamma_{1,1} = \gamma_{2,1}, \gamma_{1,2} = \gamma_{2,2})$ is the typical situation for co-located users. $(\gamma_{1,1} = \gamma_{1,2}, \gamma_{2,1} = \gamma_{2,2})$ is the typical situation for unbalanced users that both experience a very high level of CCI.

[3] This is the typical case for symmetric user locations and limited CCI.

where τ_β (ϕ_β) is the time offset (the phase offset) between $s_1(t)$ and $s_2(t)$ upon arrival at user β, and $\gamma_{i,j}$ is defined as in (**??**). Using $\boldsymbol{x}$ as short-hand for $\left(\boldsymbol{x}_{1_c}, \boldsymbol{x}_{2_c}, \boldsymbol{x}_{1_p}, \boldsymbol{x}_{2_p}\right)$, the likelihood function of $\boldsymbol{x}$, given the observation of $y_\beta(t)$, is

$$p\left(y_\beta(t)\,|\boldsymbol{x}\right) \overset{\boldsymbol{x}}{\propto} \exp\left(-\int |y_\beta(t) - u_\beta(t)|^2 \,\mathrm{d}t\right). \quad (10)$$

C. Decoding

Following the NCRS* strategy, three-stage SCD is adopted at users $\beta = 1$ and $\beta = 2$ to recover $\boldsymbol{b}_{1_c}$, $\boldsymbol{b}_{2_c}$ and $\boldsymbol{b}_{\beta_p}$ from $y_\beta(t)$. For simplicity, we will focus on the first stage of a receiver for user 1 that adopts the decoding order $M_{1_c} \to M_{2_c} \to M_{1_p}$. This involves decoding $y_1(t)$ to recover $\boldsymbol{b}_{1_c}$, treating $\boldsymbol{x}_{2_c}$, $\boldsymbol{x}_{1_p}$ and $\boldsymbol{x}_{2_p}$ as part of the noise.

We will consider the MAP bit-by-bit recovery of $\boldsymbol{b}_{1_c}$, which is optimum in the sense that it minimizes the bit error probability. Each bit $b_{1_c}[l]$ is recovered as 0 if $p\left(b_{1_c}[l] = 0\,|y_1(t)\right)$ is larger than $p\left(b_{1_c}[l] = 1\,|y_1(t)\right)$, and as 1 otherwise. Here, $p\left(b_{1_c}[l]\,|y_1(t)\right)$ is the a posteriori probability (APP) of $b_{1_c}[l]$ for given $y_1(t)$.

To efficiently compute the required bit APPs, the receiver is assumed to use SPA message passing on a FG representing a factorization of $p\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}, \boldsymbol{v}\,|y_1(t)\right)$, with $\boldsymbol{v}$ a well-chosen set of additional variables [17]. Using the chain rule, we have

$$p\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}, \boldsymbol{v}\,|y_1(t)\right) \overset{\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}, \boldsymbol{v}}{\propto} p\left(y_1(t)\,|\boldsymbol{x}_{1_c}, \boldsymbol{v}\right) P_{1_c}\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}\right). \quad (11)$$

Practical detection and decoding algorithms result from a further decomposition of the factors $p\left(y_1(t)\,|\boldsymbol{x}_{1_c}, \boldsymbol{v}\right)$ and $P_{1_c}\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}\right)$, respectively.

In the following, we assume that an efficient decoding algorithm is available, so we further focus on the detection algorithm. We compare three approaches to FG-based NCRS* receiver algorithm design. The approaches differ in the way that they translate the concept of "treating the interference as part of the noise" into a practical detection and decoding algorithm.

The time offset [14], [15] has a major impact on the receiver complexity. For further use we decompose τ_β as $\tau_\beta = K_\beta T_s + \kappa_\beta$, with K_β integer-valued and $\kappa_\beta \in \left[-\frac{T_s}{2}, \frac{T_s}{2}\right]$.

D. N-MAP decoding algorithm

Inspired by the theoretical concept of Gaussian channel coding, we model the interfering symbols $\left\{x_{2_c}[k], x_{1_p}[k], x_{2_p}[k]\right\}$, as independent Standard Normal random variables. The advantage of this approach is that it allows a closed-form derivation of $p\left(y_1(t)\,|\boldsymbol{x}_{1_c}\right)$ from $p\left(y_1(t)\,|\boldsymbol{x}\right)$ (given by (10)). As a result, we can compute the bit APPs $p\left(b_{1_c}[k]\,|y_1(t)\right)$ using a FG representation of

$$p\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}\,|y_1(t)\right) \overset{\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}}{\propto} L_1\left(\boldsymbol{x}_{1_c}\right) P_{1_c}\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}\right), \quad (12)$$

where

$$L_1\left(\boldsymbol{x}_{1_c}\right) = p\left(y_1(t)\,|\boldsymbol{x}_{1_c}\right) \overset{\boldsymbol{x}_{1_c}}{\propto} \prod L_{1,k}\left(\boldsymbol{x}_{1_c}\right)$$

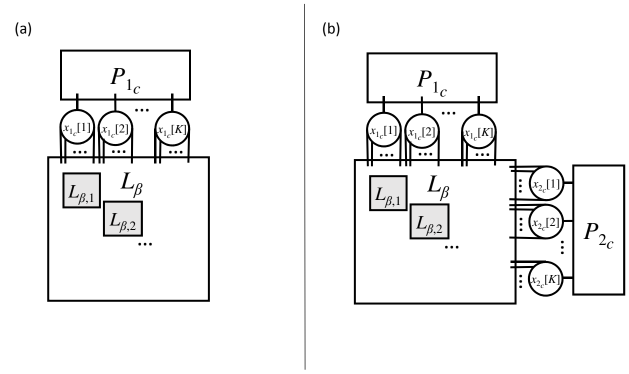

Figure 3. FG to compute bit APPs required for recovering M_{1_c}.

and

$$\begin{aligned} L_{1,k}\left(\boldsymbol{x}_{1_c}\right) = &\exp\left(\Re\left\{x_{1_c}[k]\, z_1^*[k]\right\}\right) \\ &\cdot \exp\left(\Re\left\{x_{1_c}[k]\, \tilde{z}_1^*[k]\right\}\right) \\ &\cdot \prod_l \exp\left(-x_{1_c}[k]\, C_{k,l} x_{1_c}^*[l]\right), \end{aligned} \quad (13)$$

with

$$z_1[k] = \sum_l A_{k,l} \int y_1(t)\, p(t - lT_s)\,\mathrm{d}t \quad (14)$$

and

$$\tilde{z}_1[k] = \sum_l B_{k,l} \int y_1(t)\, p(t - lT_s - \tau_1)\,\mathrm{d}t. \quad (15)$$

In (13)-(15), $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are Wiener class Toeplitz matrices [18]. For conciseness, we omit almost all further details about $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$. For the current discussion, it suffices to know that $C_{k,l}$ typically vanishes for $|k - l + K_1| > e$, with e a small positive integer value. As a result, $L_{1,k}\left(\boldsymbol{x}_{1_c}\right)$ from (13) can be approximated as $L_{1,k}\left(\boldsymbol{s}_{1_c}[k]\right)$, with $\boldsymbol{s}_{1_c}[k] = \left(x_{1_c}[k - K_1 - e], ..., x_{1_c}[k - K_1 + e]\right)$.

The FG corresponding to (12)-(13) is shown in Fig. 3(a), where the connections between the nodes $L_{1,k}$ are not specified because they depend on wether or not the states $\boldsymbol{s}_{1_c}[k]$ are introduced as internal auxiliary variables in L_β. An essential feature of (13) is that, for $e > 0$, the variable $x_{1_c}[k]$ appears in more than two factors (in $2e+1$ to be precise). This is a result of the inter-symbol-interference (ISI) that is caused by a non-zero time offset τ_1. A similar situation was encountered in [9] and the different methods outlined in [9] can also be applied here to derive practical SPA-based algorithms for MAP bit detection. For the remainder of our discourse, we only have to recall that any of the methods from [9] yields a detector with a computational burden of $O\left(\mathcal{X}_{1_c}^{2e'+1}\right)$ per symbol period (with $e' \leq e$ a design parameter.).

E. U-MAP decoding algorithm

In [10], a different approach was taken to implement NUD. Instead of Standard Normally distributed, the interfering symbols are assumed independent and Uniformly distributed over their respective alphabets. The resulting algorithms will be referred to as U-MAP (U: Uniform). Adopting this model, we have

$$p\left(\boldsymbol{b}_{1_c}, \boldsymbol{x} \left| y_1(t) \right.\right) \overset{\boldsymbol{b},\boldsymbol{x}}{\propto} L_1(\boldsymbol{x}) P_{1_c}\left(\boldsymbol{b}_{1_c}, \boldsymbol{x}_{1_c}\right), \tag{16}$$

with

$$L_1(\boldsymbol{x}) = p\left(y_1(t) \left| \boldsymbol{x} \right.\right).$$

It follows immediately from (10) that

$$p\left(y_1(t) \left| \boldsymbol{x} \right.\right) \overset{\boldsymbol{x}}{\propto} \prod_k L_{1,k}(\boldsymbol{x}),$$

where

$$\begin{aligned} L_{1,k}(\boldsymbol{x}) &= \exp\left(2\Re\left\{\sqrt{\gamma_{1,1}} x_1[k] \zeta_1^*[k]\right\}\right) \\ &\cdot \exp\left(2\Re\left\{\sqrt{\gamma_{1,2}} x_2[k] \tilde{\zeta}_1^*[k]\right\}\right) \\ &\cdot \exp\left(-\gamma_{1,1}\left|x_1[k]\right|^2\right) \cdot \exp\left(-\gamma_{1,2}\left|x_2[k]\right|^2\right) \\ &\cdot \prod_l \exp\left(-2\Re\left\{e^{j\phi_1}\sqrt{\gamma_{1,1}\gamma_{1,2}} f_{k,l}(\boldsymbol{x})\right\}\right), \end{aligned} \tag{17}$$

with $x_1[k]$ and $x_2[k]$ as in (7),

$$\zeta_1[k] = \int y_1(t) p(t - kT_s) \,\mathrm{d}t, \tag{18}$$

$$\tilde{\zeta}_1[k] = \int y_1(t) p(t - kT_s - \tau_1) \,\mathrm{d}t$$

$$g(l;\tau) = \int p(u) p(u - lT_s - \kappa) \,\mathrm{d}u$$

and where $f_{k,l}(\boldsymbol{x})$ takes either of the following two forms:

$$f_{k,l}(\boldsymbol{x}) = \begin{cases} x_1[k] g(k - l + K_1; \kappa_1) x_2^*[l] & , (i) \\ x_1[l] g(l - k + K_1; \kappa_1) x_2^*[k] & , (ii) \end{cases}. \tag{19}$$

For $\kappa \in \left[-\frac{T_s}{2}, \frac{T_s}{2}\right]$, $g(l;\kappa)$ typically becomes negligibly small for $|l| > e$, with e a small positive integer value. As a result, it is safe to approximate $L_{1,k}(\boldsymbol{x})$ as $L_{1,k}(\boldsymbol{s}[k])$, with

$$\boldsymbol{s}[k] = \begin{cases} \left(x_{1_c}[k], x_{1_p}[k], \boldsymbol{s}_{2_c}[k], \boldsymbol{s}_{2_p}[k]\right) & , (i) \\ \left(\boldsymbol{s}_{1_c}[k], \boldsymbol{s}_{1_p}[k], x_{2_c}[k], x_{2_p}[k]\right) & , (ii) \end{cases}$$

and $\boldsymbol{s}_q[k] = \left(x_q[k - K_1 - e], ..., x_q[k - K_1 + e]\right)$. The FG corresponding to (16)-(19) is shown in Fig. 3(b). Again, the connections between the nodes $L_{1,k}$ are not specified because they depend on the selection of internal detector variables. Similar to in (13), a non-zero time offset τ_1 causes ISI, which in its turn causes the variables $x_{2_c}[k]$ and $x_{2_p}[k]$ for case (i) and $x_{1_c}[k]$ and $x_{1_p}[k]$ for case (ii) to appear in more than two factors (in $2e + 1$ to be precise). As for N-MAP, the techniques from [9] can be applied to derive a practical SPA-based U-MAP detection algorithm. The computational burden that results from adopting any of these methods is $O\left(\mathcal{X}_{1_c}\mathcal{X}_{2_c}^{2e'+1}\mathcal{X}_{1_p}\mathcal{X}_{2_p}^{2e'+1}\right)$ for case (i) and $O\left(\mathcal{X}_{1_c}^{2e'+1}\mathcal{X}_{2_c}\mathcal{X}_{1_p}^{2e'+1}\mathcal{X}_{2_p}\right)$ for case (ii) per symbol period ($e' < e$ is again a design parameter).

F. S-MAP decoding algorithm

In conventional multibeam systems all a priori information about the structure of the interfering signal components is simply ignored. I.e., not just the channel coding scheme as with N-MAP and U-MAP, but also the entire modulation scheme. In that case, assuming that $\mathbb{E}\left[x_q[k] x_{q'}[k']\right]$ equals 1 if $(q,k) = (q',k')$ and 0 otherwise (which is the usual case), (12) applies with

$$L_1\left(\boldsymbol{x}_{1_c}\right) = \tilde{p}\left(y_1(t) \left| \boldsymbol{x}_{1_c} \right.\right) \tag{20}$$

and

$$\begin{aligned} &\tilde{p}\left(y_1(t) \left| \boldsymbol{x}_{1_c} \right.\right) \\ &\overset{\boldsymbol{x}_{1_c}}{\propto} \exp\left(-\frac{1}{1 + \lambda_1\gamma_{1,1} + \gamma_{1,2}} \int \left|y_1(t) - u_{1,1_c}(t)\right|^2 \mathrm{d}t\right), \end{aligned}$$

where

$$u_{1,1_c}(t) = \sqrt{\dot{\lambda}_1\gamma_{1,1}} \sum_k x_{1_c}[k] p(t - kT_s).$$

It easily follows that

$$\tilde{p}\left(y_1(t) \left| \boldsymbol{x}_{1_c} \right.\right) \overset{\boldsymbol{x}_{1_c}}{\propto} \prod_k L_{1,k}\left(x_{1_c}[k]\right), \tag{21}$$

with

$$\begin{aligned} L_{1,k}\left(x_{1_c}[k]\right) &= \exp\left(-\frac{\dot{\lambda}_1\gamma_{1,1}}{1 + \lambda_1\gamma_{1,1} + \gamma_{1,2}} \left|x_{1_c}[k]\right|^2\right) \\ &\cdot \exp\left(\frac{2}{1 + \lambda_1\gamma_{1,1} + \gamma_{1,2}} \Re\left\{\sqrt{\dot{\lambda}_1\gamma_{1,1}} x_{1_c}[k] \zeta_1^*[k]\right\}\right), \end{aligned} \tag{22}$$

where $\zeta_1[k]$ is defined as in (18). The FG representing (12) and (20)-(17) is the one from Fig. 3(a). However, in this case, every variable in L_1 appears in a single factor only, which makes the application of the SPA straightforward. The complexity of S-MAP is $O\left(\mathcal{X}_{1_c}\right)$ per symbol period.

It is easily verified that in the case of perfectly synchronous beams (i.e., $\tau_1 = 0$ and $e = 0$) N-MAP is equivalent to S-MAP. However, for $\tau_1 \neq 0$ and therefore $e \neq 0$ (and in general $e' > 0$), there is a substantial difference between N-MAP and S-MAP: N-MAP is equivalent to modeling the interfering signal components $\left(u_1(t) - u_{1,1_c}(t)\right)$ as a colored Gaussian random process, as opposed to white in the case of S-MAP.

G. Discussion

We now discuss the overall complexity of the receiver pair, assuming that N-MAP, S-MAP or U-MAP is used in decoding stages 1, 2 and 3 of users 1 and 2. To facilitate comparison, Table IV-G summarizes the complexity orders obtained in the previous sections.

Assuming that for non-zero time offset the appropriate value for the design parameter e' is always strictly larger than 0, we can draw the following conclusions:

1) In the absence of a time offset, N-MAP reduces to S-MAP, and always has a lower order of complexity than U-MAP.

Complexity	General case.	Case $\mathcal{X}_q \equiv \mathcal{X}$, $q \in Q$.
N-MAP	$O\left(\left(\max_{q\in Q} \mathcal{X}_q\right)^{2e'+1}\right)$	$O\left(\mathcal{X}^{2e'+1}\right)$
U-MAP	$\min\left(O\left(\mathcal{X}_{1_c} \cdot \mathcal{X}_{2_c}^{2e'+1} \cdot \mathcal{X}_{1_p} \cdot \mathcal{X}_{2_p}^{2e+1}\right), O\left(\mathcal{X}_{1_c}^{2e+1} \cdot \mathcal{X}_{2_c} \cdot \mathcal{X}_{1_p}^{2e+1} \cdot \mathcal{X}_{2_p}\right)\right)$	$O\left(\mathcal{X}^{4e'+4}\right)$
S-MAP	$O\left(\max_{q\in Q} \mathcal{X}_q\right)$	$O\left(\mathcal{X}\right)$

Table I
COMPUTATIONAL COMPLEXITY OF THE RECEIVER PAIR PER SYMBOL PERIOD.

2) A time offset significantly increases the complexity order of both N-MAP and U-MAP. [4]
3) In the presence of a time offset, it is not guaranteed that N-MAP yields a lower order of complexity than U-MAP. Everything depends on modulation orders used. If the same modulation order is employed for all component signals ($\mathcal{X}_q \equiv \mathcal{X}$, $q \in Q$), the complexity order of N-MAP is significantly lower than that of U-MAP.

V. CONCLUSION

We have considered NCRS-based interference management in multibeam multiuser satellite communication systems, in which adjacent spot beams employ the same frequency band and polarization. More specifically, we have explored the idea of using a simplified scheme (termed NCRS*) that transmits independent message components in every beam. To asses the associated theoretical performance loss as compared to conventional NCRS and other competing schemes, the achievable rate region of NCRS* has been derived. Our numerical results provide a clear indication for the effectiveness of NCRS*. Then, using the FG and SPA framework, a practical N-MAP receiver algorithm has been proposed for NCRS*. It was shown that a time offset between the signals received from different beams significantly increases the complexity. N-MAP has been contrasted against the U-MAP and S-MAP algorithms from the literature. In general, N-MAP can be expected to be more accurate than S-MAP (since it involves less approximations) and less complex than U-MAP. Further study is required to extensively evaluate the error performance of the corresponding receiver structures.

ACKNOWLEDGMENT

Part of this work has been supported by the European Space Agency funded activity SatNEx IV Contract No.4000113177/15/NL/CLP. The views of the authors of this paper do not reflect the views of ESA.

REFERENCES

[1] E. Lutz, “Co-channel interference in high-throughput multibeam satellite systems,” in *IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 885–891.
[2] M. Caus, A. Pastore, M. Navarro, T. Ramírez, C. Mosquera, N. Noels, N. Alagha, and A. I. Perez-Neira, “Exploratory analysis of superposition coding and rate splitting for multibeam satellite systems,” in *15th Int. Symp. Wireless Commun. Syst.*, Lisbon, Portugal, Aug. 2018.
[3] T. Ramírez, C. Mosquera, M. Caus, A. Pastore, M. Navarro, and N. Noels, “Message splitting for interference cancellation in multibeam satellite systems,” in *9th Advanced Satellite Multimedia Syst. Conf. (ASMS)*, Berlin, Germany, Sept. 2018.
[4] T. Ramírez, C. Mosquera, M. Caus, A. Pastore, N. Alagha, and N. Noels, “Adjacent beams resource sharing to serve hot-spots: a rate-splitting approach,” in *36th Int. Commun. Satellite Syst. Conf. (ICSSC)*, Niagara Falls, Canada, Oct. 2018.
[5] N. Noels, M. Moeneclaey, T. Ramírez, C Mosquera, M. Caus, and A. Pastore, “Symbol asynchronous transmission in multibeam satellite user downlink: rate region for novel superposition coding schemes,” in *6th IEEE Global Conf. Signal Inform. Process.*, Nov. 2018.
[6] A. El Gamal and Y.-H. Kim, *Network information theory*, Cambridge University Press, 2011.
[7] R. Kwan and C. Leung, “Optimal detection of a BPSK signal with unsynchronized co-channel interferers,” in *IEEE Int. Conf. Commun.* IEEE, June 1999, vol. 1, pp. 73–77.
[8] T. K. Moon and J. H. Gunther, “Multiple-access via turbo joint equalization,” *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3001–3010, Oct. 2012.
[9] D. J. Jakubisin and R. M. Buehrer, “Approximate joint MAP detection of co-channel signals in non-Gaussian noise,” *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4224–4237, Oct. 2016.
[10] S. Sharifi, A. K. Tanc, and T. M. Duman, “Implementing the Han-Kobayashi scheme using low density parity check codes over Gaussian interference channels,” *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 337–350, Feb. 2015.
[11] C. Mosquera, T. Ramírez, M. Caus, N. Noels, and A. Pastore, “Non-coherent rate splitting for the MISO BC with magnitude CSIT,” https://arxiv.org/abs/1906.03863.
[12] S. Sandhu and Paulraj A., “Space-time block codes: a capacity perspective,” *IEEE Commun. Lett.*, vol. 4, no. 12, pp. 384–386, Dec. 2000.
[13] A. Sezgin and E. A. Jorswieck, “Capacity achieving high rate space-time block codes,” *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 435–437, May 2005.
[14] Pantelis-Daniel Arapoglou, Alberto Ginesi, Stefano Cioni, Stefan Erl, Federico Clazzer, Stefano Andrenacci, and Alessandro Vanelli-Coralli, “DVB-S2X-enabled precoding for high throughput satellite systems,” *Int. J. Satell. Commun. Networking*, vol. 34, no. 3, pp. 439–455, 2016.
[15] Stefano Andrenacci, Symeon Chatzinotas, Alessandro Vanelli-Coralli, Stefano Cioni, Alberto Ginesi, and Bjorn Ottersten, “Exploiting orthogonality in DVB-S2X through timing pre-compensation,” in *Proc. 8th Advanced Satell. Multimedia Sys. Conf. and 14th Sig. Process. Space Commun. Workshop (ASMS/SPSC)*, Palma de Mallorca, Spain, Sept. 2016.
[16] M. Vu, “MIMO capacity with per-antenna power constraint,” *IEEE Trans. Commun.*, vol. 59, no. 5, pp. 1268–1274, May 2011.
[17] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
[18] R. M. Gray, “Toeplitz and circulant matrices: A review,” *Found. Trends Commun. Inform. Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[4]In passing, we point out that a non-zero time offset can be beneficial from a performance perspective, as it is known to increase the signal component separability [5].