Signal Processing for High Throughput Satellites: Challenges in New Interference-Limited Scenarios

Journal:	*IEEE Signal Processing Magazine*
Manuscript ID	SPM-Oct-2018-290.R2
Manuscript Type:	Feature Article
Date Submitted by the Author:	n/a

Complete List of Authors:	Pérez-Neira, Ana; Technical University of Catalonia, Signal Theory and Communications Vazquez, Miguel Angel; Centre Tecnologic de les Telecomunicacions de Catalunya, Mysore R, Bhavani Shankar; Interdisciplinary Centre for Security, Reliability and Trust, Maleki, Sina; TU Delft, Electrical Engineering Chatzinotas, Symeon; SnT-University of Luxembourg,

Keywords:	multiantenna signal processing, satellite communications

Signal Processing for High Throughput Satellites: Challenges in New Interference-Limited Scenarios

Ana I. Perez-Neira*† Miguel Angel Vazquez*, M. R. Bhavani Shankar*, Sina Maleki* and Symeon Chatzinotas*,

* Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA)

† Department of Signal Theory and Communications Universitat Politècnica de Catalunya

* SnT, Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg

Email: aperez@cttc.cat

Abstract

The field of satellite communications is enjoying a renewed interest in the global telecommunications market, and high throughput satellites, with their multiple spot beams, are critical components for delivering the future rate demands. In this article the state-of-the-art and open research challenges in high throughput satellites are presented, with focus on signal processing approaches for efficient interference mitigation. The paper sheds light on how some of the signal processing techniques that have been developed for wireless terrestrial communications can be applied to the satellite context. All the reviewed techniques are essential in empowering satellite systems to support the increasing demands of the upcoming generation of communication networks.

I. Introduction

Satellite communication (SatCom) systems, supported by their inherent wide coverage, are considered essential in satisfying the data traffic ubiquity, which is expected to continue to increase over the coming decades. SatCom systems are a promising solution to provide connectivity in unserved or under-served areas, regardless of the end user being fixed or in a moving platform on the ground, sea or air (e.g., on a train, ship or airplane) [1]. Also, the satellites' capability of addressing wide geographic regions, even continents, using a minimum amount of infrastructure on the ground, is very appealing. The ubiquitous coverage, together with the efficiency of its broadcast nature (i.e. its capacity to deliver the same content to a very large number of users) has been the clear main asset of SatCom. Note, however, that in recent years satellites have also developed broadband services [2], thanks to the new developed technologies that allow one satellite to manage hundreds of narrower beams with smaller coverage. This so-called spot beams can deliver different broadband data to different users. These features improve the area data traffic they can support. Another appealing feature is that SatCom can be viewed as a green communication technology. This is because, once the satellite is in space, it has access to solar energy and can stay in orbit for up to 15 years with no additional source of energy. Thanks to these features, the SatCom ecosystem, on its own, results very interesting in specific private and public sectors such as, for example, resilient overlay communications and disaster relief, governmental services, traffic off-loading and remote cellular backhaul provisioning (particularly for integrated access and backhauling which is a cornerstone of upcoming mobile communications generations), broadcast or multicast[1] services, and SCADA (supervisory control and data acquisition) for tele-supervision of industrial processes.

While such a diversification of satellite-only services is foreseen to bear fruit, maximum benefits are envisaged by integrating satellite and terrestrial communications in the future generations of communications. For instance, the roles and benefits of satellites in the fifth generation (5G) [3] have begun to be studied in 3GPP Release 14, leading to the specific requirement to support satellite access being captured in TS22.261 - Service requirements for next generation new services and markets; Stage 1, recognizing the potential added value that satellite coverage brings, as part of the mix of access technologies for 5G [4], [5]. In this context potential new markets and emerging applications that are currently pursued by the satellite community include ubiquitous broadband access, commercial aeronautical and maritime services, machine-to-machine communications, and smart cache feeding.

In all these applications, signal processing (SP) is challenged to satisfy the corresponding requirements in terms of spectrum and energy consumption. The increase in demand for these new satellite services and systems drives innovative approaches that are moving away from the traditional linear television broadcast (i.e., direct to the home, or DTH). In addition, the new standard developments on the SatCom arena, some of them for their integration with the 5G New Radio standard, help to pave the way to meet demand for faster, and cheaper

[1] In both broadcast and multicast multiple users receive the same information using the same radio resources. While broadcasting refers to same information for the entire coverage area, multicasting refers to a beam or a set of beams. In this article we try to stick to this meaning, however, these terms are often used interchangeably in the SatCom community and broadcast may be sometimes used to encompass both.

SatComs. On the contrary, the conventional commercial SatCom, which were in general custom designed to meet the requirements of a specific mission, lose importance. Notably, custom designs have been a limiting factor in reducing the cost and delivery time of the satellite.

In parallel, in the past few years, an important new trend has been observed in the satellite sector that relies on the vast potential of either very high or high throughput satellites (V/HTS). Many operators are currently upgrading their constellations to deliver higher radio frequency (RF) power, enhanced functionality, and higher frequency reuse with V/HTS technology. With its flexibility, V/HTSs are key in the paradigm shift that we have commented earlier. Their spectral efficiency (number of b/s/Hz) will be multiplied by two to three times with respect to the current non-high throughput satellites. In total, the expected aggregate high throughput will vary according to application served and satellite, but it is anticipated to be in the range of 25-60 Gb/s. For very high throughput satellites the expected aggregate rate is within the range of Tb/s. This superior performance does not necessarily require higher bandwidth or increasing the weight of the payload. Therefore, a superior aggregate rate is provided at a lower cost per bit. As we present in this paper, V/HTS leverages frequency reuse across multiple narrowly focused spot beams; thereby, maximizing the available frequency band. However, high frequency-reuse means higher levels of interference. This new trend poses interesting challenges regarding emerging interference-limited scenarios, and SP offers valuable tools to cope with them.

The purpose of this article is to provide an introduction to the new interference-limited paradigm that V/HTS is facing, together with a thorough review of the recent SP approaches that address this new paradigm. In this article, new directions for research are identified. In particular, we narrow our focus to fixed satellite services (FSS) that are provided by geostationary satellites in the L/Ku/Ka-bands, where SP is needed to attain the promised high rates (at Ku/Ka-bands these rates are in the order of Tb/s). Notably, it is also in the geostationary orbit where V/HTS has originated with well-established waveforms, coding, and modulators defined in the digital video broadcasting over satellite 2 extension (DVB-S2X) standard [6]. Both on ground SP at the gateway and onboard SP at the satellite are addressed. Before that, let us comment, first, about the main features of SatCom systems; and second, about high throughput satellites.

II. Features of Satellite Communication Systems

In the past decades, SatCom systems have exploited new techniques and technologies that were originally implemented in terrestrial communications. For instance, while in the mid-

1980s, advanced analog-to-digital and digital-to-analog converters (ADC and DAC, respectively) were used in delay-sensitive audio/voice applications, satellite systems adapted them into more complex digital signal processing techniques for delay-tolerant video broadcasting/multicasting. However, the customization of the terrestrial communication technologies to the peculiarities of the SatCom systems is not straightforward due to the different satellite orbits, channels, system constraints, and processing. Next, let us comment on these peculiarities.

Today, there are approximately 1300 fully operational communication satellites, and every type of orbit has an important role to play in the overall communication system. Geostationary earth orbit (GEO), at 35,000 km, presents an end-to-end propagation delay of 250 ms; therefore, it is suitable for the transmission of delay-tolerant data. Medium earth orbit (MEO), at 10,000 km, introduces a typical delay of 90 ms; based on that, it can offer a compromise in latency and provide fiber-like data rates. Finally, low earth orbit (LEO) is at between 350 and 1,200 km, and introduces short delays that range from 20 to 25 ms. In all these cases, the satellite acts as a very particular wireless relaying node, whose features lead to a communication system that cannot be always treated as a wireless terrestrial one [2]. Namely, the channel, communication protocols, and complexity constraints of the satellite system create a set of particularities, which are noted below:

- Due to the long distance to be covered from the onground station to the satellite, the satellite communication link may introduce both a high round-trip delay and a strong path loss of hundreds of decibels. To counteract the latter, satellites are equipped with high-power amplifiers (HPA) that may operate close to saturation and create intermodulation and nonlinear impairments (it is much the same way as an audio amplification).
- Usually one HPA controls one antenna element (also so-called feeds) on the satellite. To ensure that there is no distortion, it is critical to satisfy the power limitations on each feed and not only on the average total power that the satellite transmits.
- Satellite communications traverse about 20 km of atmosphere and experience high molecular absorption, which is even higher in the presence of rain and clouds, particularly for frequencies above 10 GHz. Therefore, satellite links have been conventionally designed based on the thermal noise limitations and on the link budget analysis that have to consider large protection margins for additional losses (e.g., rain attenuation).
- In the nongeostationary orbits (i.e., MEO and LEO), there are significant time-domain variations of the channel due to the relative movement of the satellites with respect to the ground station. Variations in Doppler can exceed the range [-200, 200] kHz over

the pass of the satellite (e.g., 45 min for MEO).

- Due to the long distance and carrier frequencies, the satellite antenna feeds are generally seen as a point in the far field, thus precluding the use of conventional spatial diversity schemes onboard the satellite. Also, due to the absence of scatterers near the satellite (i.e., there are no objects in space that create multiple paths) and the strong path loss (i.e., it is a long-distance communication), the presence of a line-of-sight component, which focuses all the transmitted power and is not blocked or shadowed, is required. This contrasts with the terrestrial cellular communications, where the presence of a line-of-sight is not critical. On the positive side, due to the lack of rich scatter, the satellite communications experience higher cross polarization isolation than the terrestrial communication networks.
- The processing complexity onboard the satellite is limited and largely unexplored, as it is highly correlated with its power consumption, mass, and ultimately, with the cost of the system.
- The received signal-to-noise-ratio (SNR) is very low and therefore the user terminal (UT) must have high sensitivity, good receiver antenna gains, and good tracking capabilities to steer the beam of the UT, such that it continuously points to the satellite.
- The practical challenges of the satellite system require solutions that are different from the ones used in the terrestrial wireless communications. An important one is the specific satellite multiuser protocol framing that is defined in the current broadband and mobile interactive standards, DVB-S2X and BGAN, respectively. In these protocols, in order to overcome the satellite channel noise, channel codes are long and, therefore, must take into account data from multiple users. This fact creates a multicast transmission, because the same information has to be decoded by a group of users. Multicast transmission requires specific SP techniques, as this article explains.
- Finally, satellite solutions are generally characterized by a relatively long development and in-orbit validation phases before being operative. This is different from terrestrial solutions, where it is easier to test new technologies *in situ* without incurring in excessive deployment costs.

As we have commented, this article focuses on the recent V/HTS technologies, due to their potential to improve SatCom systems in both performance and cost. Let us next describe them in detail.

III. HIGH THROUGHPUT SATELLITES: A NEW INTERFERENCE-LIMITED PARADIGM

In contrast to monobeam satellites, high throughput satellites partition the service area into multispot beams; this allows a higher aggregate throughput and more service flexibility to satisfy the heterogeneous demand. The system architecture is shown in Figure 1 and consists of a Gateway (GW), a satellite, and multiple UTs. The GW is connected to the core network and serves a set of users that are geographically far away using the satellite. The links from the GW to the satellite and from the satellite to the UT are known as the feeder link and the user link, respectively. In the usual star configuration that is illustrated in Figure 1, the feeder link presents high directivity and gain. As this link presents an SNR that is considerably higher than the one in the user link, in general, it is assumed to be noiseless and perfectly calibrated against channel power variations. Also, depending on the direction of the communication, the link is known as forward link when it originates at the GW to the UT, and reverse link when it goes from the UT to the GW. The mentioned links combine to create the: feeder forward, user forward, feeder return and user return links, each of which usually works in different frequency bands. The frequency selection is driven by considering different aspects, among them coverage and beam size, atmospheric conditions in the served region, and availability of a robust ecosystem of ground equipment technologies. In SatCom, the signals have to cross the atmosphere and they use high frequencies. Otherwise, the signals would get reflected by the atmosphere and could not penetrate to get through to the satellite. For instance, current-generation GEO HTSs typically use the Ka-band, which is less congested than the C/Ku-band. For FSS, this refers to the exclusive satellite band from 19.7 to 21.2 GHz for the forward link and from 29.5 to 31 GHz for the reverse link. Mobile satellite services (MSS) generally use lower frequencies such as the L-band (i.e., from 1.5 to 2.5 GHz) because of their lower attenuation, which enables lower antenna gains at the mobile UT. Note, however, that recently the Ka-band is also being considered to provide in-flight and maritime connectivity. In all these cases the frequencies that are transmitted by the satellite occupy the lower band. The satellite is a light-weight device that cannot support high-power transmission units. Therefore, it transmits at a lower frequency (the higher the frequency, the higher is the transmission power to accommodate losses) as compared to the GW, which can afford to use very high-power transmission.

The HTSs that are currently operative (e.g., Viasat-2, SES-12) provide aggregate data rates of more than 100 Gb/s. These HTS systems use the Ku/Ka-band in both feeder and user link, and serve as much as 200 beams in the user link. For its part, V/HTS systems (e.g., Viasat-3) aim at achieving data rates in the range of Tb/s and, due to that, they need higher frequencies in the Q-band (30-50 GHz), V-band (50-75 GHz), and W-band (75-110 GHz), in order to serve as many as 3000 beams in the user link. For these reasons, advanced SP techniques are required in order to reduce the interference among so many multiple beams, facilitate adaptive coverage, dynamically optimize the traffic, and share the spectrum with terrestrial services, among other functions. Flexibility in the resource allocation per beam can significantly improve the quality of service and bring down the incurred cost of the V/HTS system per transmitted bit.

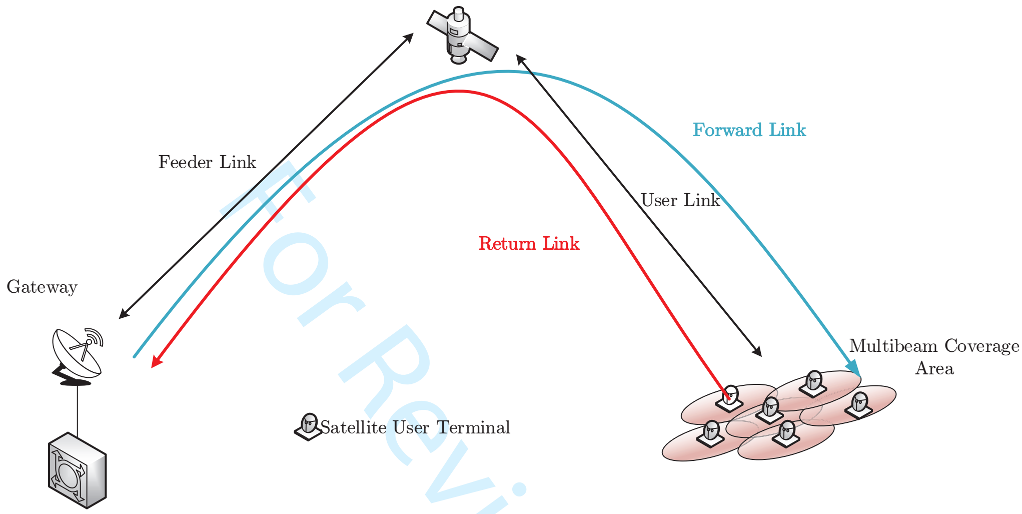

Fig. 1. Scheme of a multibeam satellite system. The forward link goes from the GW to the UTs via the satellite (blue line). The return link goes from the UTs to the GW via the satellite (red line). The feeder link connects the GW with the satellite (bidirectionally), whereas the user link connects the UTs with the satellite (bidirectionally).

Figure 2 shows an example of the classical linguistic beam wide coverage. In contrast, multispot beams allow tessellation of the coverage into much smaller footprints, thus enabling frequency reuse within the geographical area covered by one linguistic beam. As a consequence, per user bandwidth assignment and the aggregate throughput can potentially increase in V/HTS. Multispot beams enable broadband data services in addition to the traditional broadcast services offered by the linguistic beams. Figure 3 shows an example of the footprints of a four-color reuse scheme, where a total bandwidth of 500 MHz is allocated to the user link at the Ka-band. This bandwidth is divided into two sub-bands that, when combined with two orthogonal polarizations, generates the so-called four-color beam pattern across the coverage area. In the Ku/Ka-band, orthogonal polarizations maintain very low cross-polarization and due to that, they can be used as if they were different frequencies. Within each beam, multiple users can be served by using an orthogonal access scheme. Currently, with the common frequency reuse of four colors, the interference power among beams is in the range from 14 to 34 dB below the carrier signal.

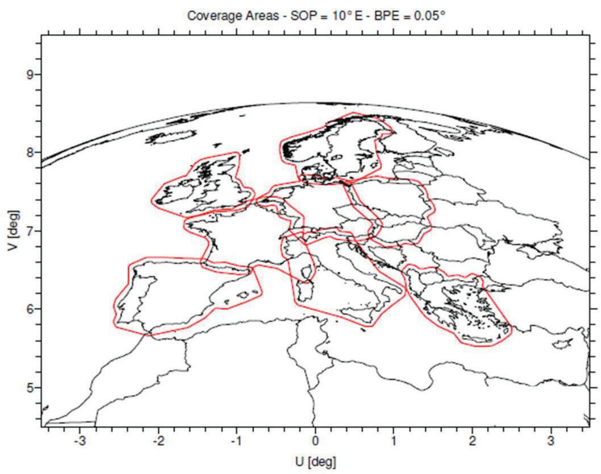

Fig. 2. Broadcasting satellite with eight linguistic beams in the Ka-band (copyright European Telecommunications Standards Institute, 2015; further use, modification, copy and/or distribution are strictly prohibited).

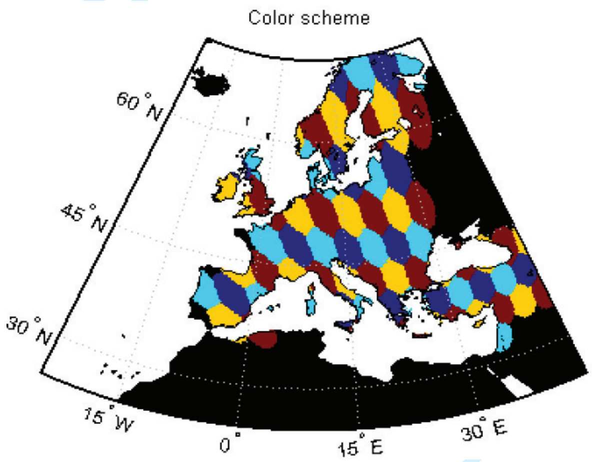

Fig. 3. User frequency plans for the scenario with 71 beams and frequency reuse pattern equal to four (copyright European Telecommunications Standards Institute, 2015; further use, modification, copy and/or distribution are strictly prohibited).

With the aim of lowering the cost per transmitted bit and increasing the spectral efficiency or the available system bandwidth, new systems aim at reusing more aggressively the available spectrum among the spot beams. This aggressive frequency reuse may not be applied to all the satellite beams, but only to those that are experiencing a higher traffic demand; thus, the performance indicator in this case is the area traffic capacity. Nevertheless, increasing the frequency reuse leads to a further increase of intrasystem interference among the cochannel beams, which shifts the classical noise-limited link budget analysis towards an interference-dominated situation. The side lobes of the beam radiation patterns create interference leakage among beams, and the carrier-to-interference ratio (CIR) can be severely degraded. In order to successfully implement high frequency-reuse, interference management has to be implemented at the GW, the satellite, the UT or, some combination of these. It follows that the CIR mostly depends on the position of the UT, the cross-over level, and the antenna radiation pattern. Hence, the most favorable case corresponds to the situation in which the UT is in the center of the beam, while the worst case is when the UT is located at the beam-edge area. We note that for a frequency reuse pattern equal to one (i.e., $f_r = 1$) the median of the CIR in dB is around 0 dB and ranges from -5 to 15 dB, for $f_r = 2$ it is 8 dB and goes from 0 to 25 dB, for $f_r = 3$ it is 25 dB and ranges from 10 to 30 dB, and for $f_r = 4$ it is 30 dB and goes from 10 to 40 dB. In any case, the interference power that comes from the high frequency-reuse adds to the one originating from the nonlinear distortion of the HPA. Unfortunately, the traditional approach to diminish interference by using power control is insufficient, and, hence, novel signal processing alternatives that exploit the structure of the cochannel interference are needed.

We note that the final performance of the V/HTS system depends not only on the capabilities of the applied SP, but also on many system choices. Complex design trade-offs and practical aspects need to be respected, as detailed in references like [7]. For example, if hundreds of beams are available in the system, high frequency-reuse schemes can stress the payload resources of the satellite in terms of mass, power, and thermal dissipation. Another important consequence of increasing the frequency reuse is that the bandwidth of the feeder link should increase accordingly. As this is not straightforward, different alternatives should be studied, such as, for instance, employing multiple gateways in the feeder link. Provisioning SP on board the satellite (i.e. onboard processing) can provide additional degrees of freedom towards mitigating the interference and enabling seamless multiple GW operation; choice of onboard processing depends on mission limitations and commercial viability.

Finally, it is important to note that V/HTS systems require the most advanced transmission standards. Currently, DVB-S2X in [6] is the standard of both forward broadcast and broadband satellite networks. We can say that they have been designed to best fit the features of the GEO satellite channel (i.e. its delay and transmission power limitations). Using high efficiency modulation and coding schemes (MODCODs) up to 256 Amplitude Phase Shift Keying (APSK), combined with advanced interference management techniques, enable aggressive and flexible frequency reuse. DVB-S2X incorporates the novel superframing structure that enables the use of SP techniques that have never been used before in the satellite context, such as precoding and multiuser detection at the UT. Among other things, this standard introduces orthogonal Walsh-Hadamard sequences as reference/training sequences, allowing simultaneous estimation of the channel state information of multiple beams. The superframe concept is designed to maximize the efficiency of the channel coding scheme by encapsulating the information intended to several UTs using the same MODCOD. The novel superframe specification has opened a new scope of research related to advance interference management techniques for aggressive frequency reuse scenarios. Namely, it is a multicast transmission because in the framing structure of DVB-S2X the same data are transmitted to multiple receivers. This precludes the calculation of a precoding matrix on a receiver-by-receiver basis. Differently, a multicast precoder has to be designed for a group of receivers. Further details on DVB-S2X standard, which have a beneficial impact on precoding and multiuser detection, can be found in Annex E of [6]. Although DVB-S2X is the most popular satellite standard nowadays, due to its design for broadband and broadcast communications, the other most used standard is BGAN. This is for mobile interactive communications and it also uses a similar superframe concept. Next, we address spatial precoding techniques, which are supported by these standards in order to mitigate the interbeam interference.

IV. Precoding in Multibeam Satellite Systems

A. Architecture and communication peculiarities

With the aim of increasing the offered data rates of a given satellite, both operators and manufacturers are investigating a variety of alternatives. One main approach is to consider satellite communication links at extremely high frequencies such as the W-band. However, large investments are required for implementing the communication subsystems in these bands because no previous commercial satellite communication systems have been operating at these frequencies. In addition, new challenges with regard to channel impairments appear. As a result, spectral efficient alternatives that exploit the current frequency bands are of great interest.

This is the case of precoding techniques that allow a high frequency-reuse factor among the different beams in the user link. With the aid of precoding, a satellite UT can obtain a sufficiently large signal-to-interference-plus-noise-ratio (SINR) even though the carrier bandwidth is reused by adjacent beams. This is because in order to maintain a certain SINR value, the precoder uses the channel knowledge to mitigate the interference towards the UTs. Resorting to the system architecture that is schematically depicted in Figure 1, the precoding matrix is, in general, computed at the satellite GW. After that, the beam signals are precoded and transmitted through the feeder link using a frequency division multiplexing scheme. Then, the satellite payload performs a frequency shift and routes the resulting radio signal over an antenna array that transmits the precoded data over a larger geographical area that is served by the multiple beams in the user link. The block diagram is summarized in Figure 4.

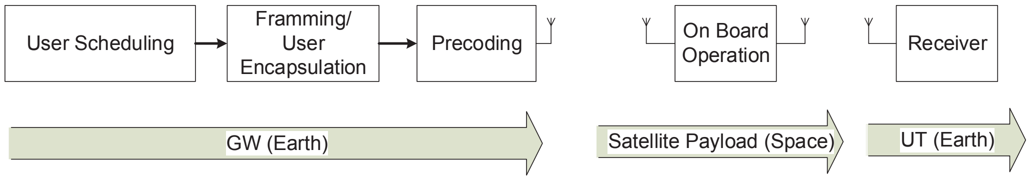

Fig. 4. Block diagram with the multibeam satellite precoding at the GW.

Low-complexity linear precoding techniques are of great interest and this is a problem that deserves further attention and research. Indeed, the computational complexity that is required to implement multibeam satellite precoding techniques, gains even a larger importance compared to the cellular systems as the dimensions of multibeam satellite systems are extremely high. For instance, the forthcoming Viasat-3 system is expected to utilize nearly 1000 beams per color to serve the coverage area as it is presented in Figure 5. As a result, the on ground equipment should be prepared to update a precoding matrix to cope with 1000 users on a per-frame basis. This complicates the precoding implementation due to the extremely large size of the precoding matrix that must be calculated (i.e. 1000x1000 matrix). Further details are explained in this section after the introduction of the mathematical system model (i.e. Section IV-C).

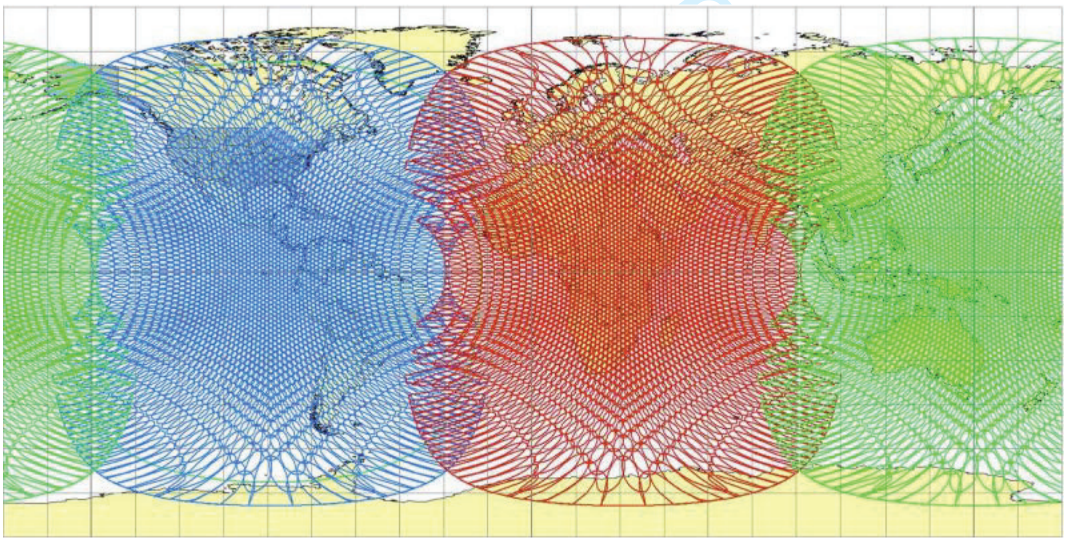

Fig. 5. Viasat 3 beampattern footprints. Each of the colors corresponds to three different satellite coverage areas. There are 1000 spots per color (source: Viasat).

In order to create all these beams, at least an equal number of antenna elements is needed on board the satellite. Bearing in mind that, with these large dimensions, we could draw an analogy with a **massive multiple-input-multiple-output (MIMO)** system [8], [9]. However, the actual multibeam satellites performance does not scale with the number of antennas and UTs as terrestrial systems do. The main reasons can be summarized as follows:

- *The multibeam satellite system is in general user overloaded.* Current satellite payloads are equipped with N antennas (generally coined as feed elements) that have to serve a total amount of UTs that is much larger than N. This fact clearly contrasts with massive MIMO, where the number of base stations antennas grows larger than that of UTs. Therefore, as N grows in massive MIMO the angular resolution of the antenna array increases so users can be separated more effectively by the appropriate antenna weights, which does not happen in V/HTS SatCom.
- *The co-channel interference power does not decrease as the number of beams increases.* The favourable propagation in massive MIMO mentioned in [10] does not occur in multibeam satellite systems. That is, in a scattered terrestrial channel environment, the off-diagonal elements of the channel covariance matrix tend to zero as the number of antennas grows, leading to an ideal interference-free scenario. In other words, the fast fading channel from the different users become almost orthogonal as the number of antenna in each base station grows large while keeping the number of UTs fixed. On the contrary, due to the low scatter in the GEO satellite channel (i.e. line-of-sight channel), there is always strong cochannel interference among the beams.
- *There is no pilot contamination.* Massive MIMO in multicell scenarios entails difficulties in the channel estimation operation as base stations or users located in adjacent cells might inject interference into the estimation process, in the downlink and in the uplink, respectively. This is because orthogonal pilot sequences have to be reused from cell to cell, as in massive MIMO systems each cell has a high number of antennas and users active in the same time/frequency resource. Ideally one orthogonal sequence per transmitting antenna would be needed for each base station in the downlink and one per user transmitting antenna in the uplink, for time division multiplex, and this is not possible. Note that in frequency division multiplex, the users send their corresponding estimated channel to the base station. If this is done simultaneously by a large amount of users, signal contamination may also arise. In the multibeam satellite case, this is not the usual case, since in the downlink each UT has to estimate the channel from each beam, and therefore, only one pilot signal per beam is needed. Due to the high directivity of the beams, only the adjacent beams are the ones that create a relevant interference. This

is a limited number of beams and, therefore, the pilot signals of adjacent beams can be orthogonal (i.e. note that DVB-S2X considers a set of 32 Walsh-Hadamard sequences that can be used as simultaneously sent pilots for the estimation of up to 32 channels in the downlink). For the same reason, in the uplink, either in frequency or in time division multiplex, the number of simultaneous transmitted signals for channel state information acquisition purposes is limited. Also, the satellite channel is, in general, non-frequency-selective and preserves the orthogonality at the UT. Pilot contamination could only appear in the case of having multiple satellites with independent processing of their signals.

In any case, although V/HTS cannot be considered a massive MIMO system, it could benefit in future from the technology advances in terrestrial massive MIMO. This can be the case, for instance, of the use of larger antenna apertures to reduce the size of the spot beams in order to improve capacity, and with capability of flexible formation of beams. This would contrast with the actual spot beams with fixed footprints (e.g. Figure 3).

Another interesting aspect is that precoding in multibeam satellite systems presents certain similarities with respect to the **cloud radio access network** (C-RAN) architecture (e.g. see [11]). Referring to the description in Figure 6, in both, C-RAN and multibeam satellite systems, the baseband processing is performed by a centralized entity, which is coined in C-RAN as baseband unit (BBU). In the satellite, the feeder link plays the role of the optical fiber links from the BBU to the remote antenna. Therefore, we can say that a multibeam satellite is similar to a C-RAN cellular architecture with a wireless fronthaul in the sense that the baseband processing and the RF elements are placed in separated locations. However, in contrast to C-RAN, multibeam satellites do not perform in general a digitalization of the baseband signal. In other words, they behave as an amplify-and-forward relay. In light of the above discussion, current multibeam satellite systems do not present the same peculiarities as a C-RAN. There exist substantial differences between these approaches, which refer mainly to implementation details that reflect the nature of the underlying systems. Nevertheless, the progress on SP on board the satellite, and on the softwarization of the satellite networks, makes the convergence of both systems almost inevitable in the future. Thus, we identify C-RAN architectures that include the satellite component as a new direction worth investigating. Next, let us introduce the system model in order to understand the peculiarities of the SP precoding for high throughput satellites. The challenge is to cope with a frequency reuse pattern equal to one.

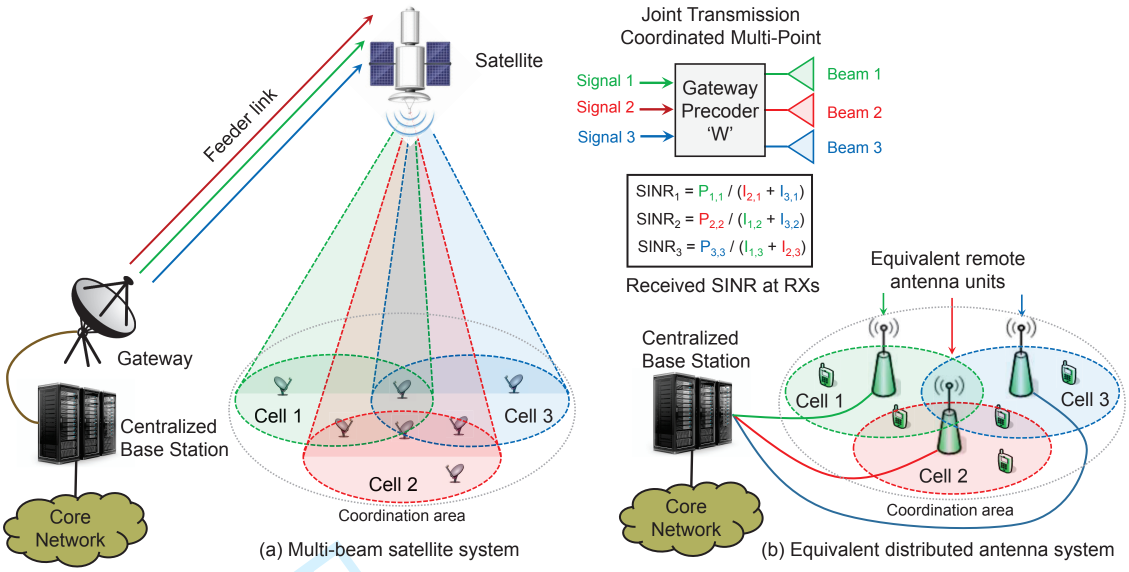

Fig. 6. Multibeam satellite precoding system versus a C-RAN cellular system.

B. System Model

Let us consider a multibeam satellite system where each UT has only one antenna and the satellite is equipped with N antennas, which are combined to generate a beam radiation pattern composed of K fix beams [12] [2]. For each beam, we assume that a total number of N_u users are simultaneously served per beam (i.e., the total number of served users by the satellite is KN_u).

Considering that all beams radiate in the same frequency band (i.e., $f_r = 1$). Single carrier modulated signals are transmitted, which is the most usual case in SatCom due to the degradation that the non-linearities of the high power amplifier create on multicarrier signals (e.g. DVB-S2X uses single carrier modulations). The discrete equivalent baseband model for the received signal at the i-th user terminal of each of the K beams is written as

$$\mathbf{y}^{[i]} = \mathbf{H}^{[i]}\mathbf{D}\left(\mathbf{x}+\mathbf{z}\right)+\mathbf{n}^{[i]}, \quad i=1,\ldots,N_u, \tag{1}$$

where vector $\mathbf{y}^{[i]} \in \mathbb{C}^{K\times 1}$ is the vector containing the received signals of the i-th UT (i.e., the value $\left[\mathbf{y}^{[i]}\right]_k$ refers to the received signal of the i-th UT at the k-th beam), whereas vector $\mathbf{n}^{[i]} \in \mathbb{C}^{K\times 1}$ contains the noise terms of each i-th UT. The entries of $\mathbf{n}^{[i]}$ are assumed to be independent and Gaussian distributed with zero mean and variance equal to σ_n^2 (i.e.,

[2]In this section, we use the term *beam* when we refer to a satellite transmission of N_u UTs over a certain coverage area. Despite the term might generate confusion to some readers and its original meaning is fuzzy in systems with precoding, we would like to keep it as it is generally employed by the satellite engineering community.

$E\left[\mathbf{n}^{[i]}\mathbf{n}^{[i]H}\right] = \sigma_n^2\mathbf{I}_K \quad i = 1,\ldots,N_u$). Vector $\mathbf{x} \in \mathbb{C}^{N\times 1}$ contains all the transmitted signals (i.e. one per transmitting feed). The term $\mathbf{z} \in \mathbb{C}^{N\times 1}$ corresponds to the noise term of the feeder link transmission assumed to be Gaussian zero mean with variance equal to σ_f^2.

Matrix $\mathbf{H}^{[i]} \in \mathbb{C}^{K\times N}$ is the channel matrix, whose k-th row denoted by $\mathbf{h}_k^{[i]}$ is the channel vector of the i-th user located at the k-th beam. This vector contains the channel coefficients from each antenna element $n = 1...N$ to the user. In order to clarify this system model, Figure 7 describes how $\mathbf{H}^{[i]}$ is constructed. As it can be observed in the example with $K = 3$ and $N_u = 2$, the different rows of these channel matrices $\mathbf{H}^{[1]} \in \mathbb{C}^{K\times N}$ and $\mathbf{H}^{[2]} \in \mathbb{C}^{K\times N}$ are formed by UTs channel vectors located at each of the beams.

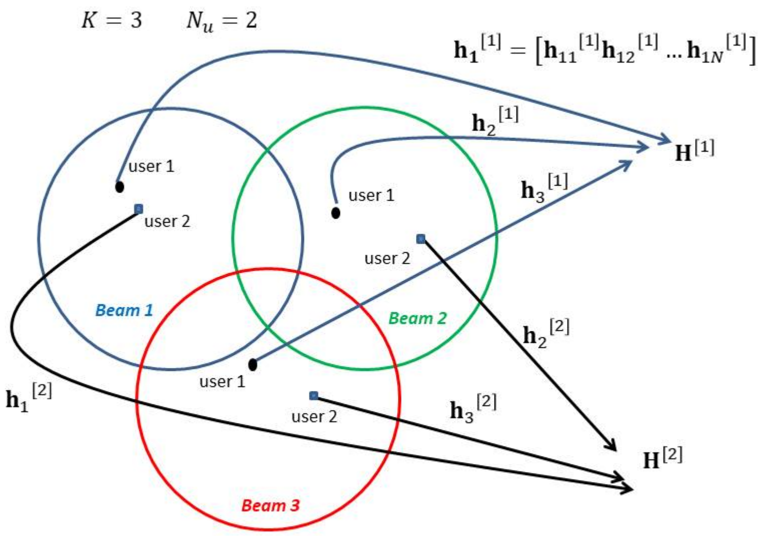

Fig. 7. System model example with $K = 3$ and $N_u = 2$. The channel matrices $\mathbf{H}^{[1]}$ and $\mathbf{H}^{[2]}$ are formed by the UTs represented by squares and the circles, respectively.

Note that it is a flat-fading channel, which is the realistic assumption in the use cases for Ku/Ka-band with the 500 MHz user link bandwidth, even for mobile satellite systems [13]. This is due to the high directivity of the beams that eliminates any possible multipath. Matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ consolidates the feeder link and payload operation. Note that in this section we do not consider the nonlinearity onboard operation effects, which are analysed in the following section. We write matrix $\mathbf{D}$ as follows

$$\mathbf{D} = \text{diag}\left(m_1,\ldots,m_N\right), \tag{2}$$

where m_n denotes the resulting amplitude variation of the n-th feed signal. Assuming perfect onboard automatic gain control mechanisms, we can establish that $m_n = 1$ for all $n = 1, \dots, N$, rendering $\mathbf{D}$ to be an identity matrix.

Feeder links are supported by parabolic antennas of 7-9 m in diameter at the GW and a high gain antenna located at the payload. The feed signals are multiplexed in frequency over the feeder link available bandwidth. In this context, it is always the case that the feeder link SNR, which is defined as

$$\mathrm{SNR}_{\text{feeder}} = \frac{\mathrm{E}\left[|\mathbf{x}|^2\right]}{\sigma_f^2}, \tag{3}$$

leads to extremely high values and; thus, the feeder link is considered noiseless.

With this, the system model in (1) can be simplified as

$$\mathbf{y}^{[i]} = \mathbf{H}^{[i]}\mathbf{x} + \mathbf{n}^{[i]}, \quad i = 1, \dots, N_u. \tag{4}$$

Curiously, in spite of the signal model in (1) presenting strong similarities with the multiuser MIMO amplify-and-forward relay channel [14], it is evident that after the high feeder SNR assumption, the system model in (4) does not consider the relay processing effect.

Getting more into the detail of $\mathbf{H}^{[i]}$, this matrix is usually assumed to follow:

$$\mathbf{H}^{[i]} = \mathbf{F}^{[i]} \circ \overline{\mathbf{H}}^{[i]}, \quad i = 1, \dots, N_u, \tag{5}$$

where $\overline{\mathbf{H}}^{[i]} \in \mathbb{C}^{K \times N}$ is a full column-rank matrix containing the antenna radiation pattern and path loss and $\mathbf{F}^{[i]} \in \mathbb{C}^{K \times N}$ denotes a matrix of random entries modeling the propagation effects. The operator $\circ$ denotes the Hadamard product. Despite the similarities with the terrestrial multiuser MIMO case, both channel models are quite different, as we are going to describe next. The (k, n)-th entry of matrix $\overline{\mathbf{H}}^{[i]}$ is

$$\left[\overline{\mathbf{H}}^{[i]}\right]_{k,n} = \frac{G_R a_{kn}^{[i]} e^{j\psi_{k,n}^{[i]}}}{4\pi \frac{d_k^{[i]}}{\lambda}} \; k = 1, \dots, K; n = 1, \dots, N; i = 1, \dots, N_u. \tag{6}$$

with $d_k^{[i]}$ as the distance between the i-th UT at the k-th beam and the satellite. λ is the carrier wavelength, G_R^2 is the UT receive antenna gain. The term $a_{kn}^{[i]}$ refers to the gain from the n-th feed to the i-th user at the k-th beam. Due to the high directivity of each feed, these gains are very low for users in beams that are far away from the illuminated area by each feed.

The time varying phase due to beam radiation pattern and the radio wave propagation is represented by $\psi_{k,n}^{[i]}$. The phase value, $\psi_{k,n}^{[i]}$, comprises different contributions. In particular,

$$\psi_{k,n}^{[i]} = \theta_{\text{RF},k}^{[i]} + \theta_{\text{LNB},k}^{[i]} + \theta_{\text{PL},n}, \tag{7}$$

where $\theta_{\text{RF},k}^{[i]} = \frac{2\pi}{\lambda} d_k^{[i]}$ is the phase rotation due to the RF signal propagation which depends on the UT distance to the satellite, $\theta_{\text{LNB},k}^{[i]}$ is the phase contribution of the receiver low noise block downconverters assumed to be Gaussian with zero mean and standard deviation of 0.24 degrees and $\theta_{\mathrm{PL},n}$, which are the payload oscillator phase offsets that are assumed to be Gaussian with zero mean and a standard deviation of around 2 degrees with ultra-stable oscillators [15].

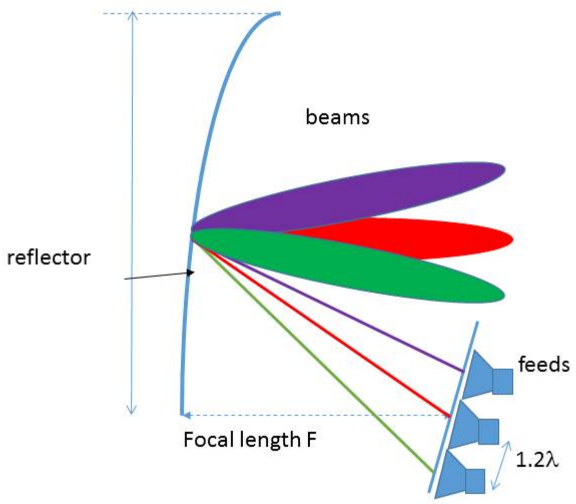

Fig. 8. Array-fed reflector antenna and the beam generation process.

We note that the values of $a_{kn}^{[i]}$ depend on the onboard multiantenna solution that is employed. If the satellite employs direct radiating elements (i.e. a phased array aperture), then $a_{kn}^{[i]}$, with $n = 1...N$, is the steering vector of user k; thus conferring a Vandermonde structure to matrix $\overline{\mathbf{H}}^{[i]}$, whenever the phase $\psi_{k,n}^{[i]}$ is compensated. For instance, in L-band satellite systems, the onboard beamforming is done via direct radiating elements (this is the case of Inmarsat and Globalstar systems). On the contrary, at Ku/Ka-bands the satellite generally employs a reflector antenna with multiple feeds (i.e. array-fed reflectors) as the one depicted in Figure 8, which presents the required satellite gains at these mm-Wave frequencies.

Also, in (4), we consider that all feed signals impinge on the UT antenna at the same time instant (i.e there is no time misalignment between simultaneously transmitted frames). This aspect, together with the assumption of perfect channel state information at the transmitter, transmitter, enables the precoding techniques that are proposed in next section. If perfect channel state information is not available, the same strategies as in terrestrial wireless communications have to be designed. In this aspect, the satellite scenario does not bring any new

SP problem, and, therefore they are not covered in the paper.

To finalize the description of (5), we note that the atmospheric fading coefficients from one user to the different antenna elements are the same. This is caused by the relatively small separation of the antenna elements compared with the satellite altitude. Therefore, $\left[\mathbf{F}^{[i]}\right]_{k,n} = \mu_k^{[i]} e^{j\theta_k^{[i]}}$, where each fading coefficient is independent of the transmission feed. Notably, there is no multipath and a strong line-of-sight is present in frequencies above 10 GHz (i.e., above the Ku band). Whenever there is no blockage, $\mu_k^{[i]}$ can be assumed to be Ricean distributed; therefore, (4) models a line-of-sight channel).

C. Precoding Techniques

In order to mitigate the cochannel interference due to the high frequency reuse factor, precoding is performed, and the transmitted signal vector per beam is given by $\mathbf{x} = \mathbf{W}\mathbf{s}$, where $\mathbf{s} \in \mathbb{C}^{K\times 1}$ is the vector that contains the transmitted symbols per UT, which we assume are uncorrelated and with power equal to one $\left(E\left[\mathbf{s}\mathbf{s}^H\right] = \mathbf{I}_K\right)$. Matrix $\mathbf{W} \in \mathbb{C}^{N\times K}$ is the linear precoding matrix to be designed. As mentioned previously, each frame contains information intended for multiple users in order to attain a large channel coding gain. In this context, every UT with index $i = 1, \ldots, N_u$ at the k-th beam shall detect the same information $[\mathbf{s}]_k$, leading to the so-called multigroup multicast transmission [16]. While most of the works on multigroup multicast precoding have only considered sum-power constraints at transmission, in satellite communications a per-antenna power constraint needs to be taken into account, due to the intrinsic per-antenna power limitation on the satellite.

An important figure of merit in order to design the system is the so-called system sum-rate, which is the aggregate information rate of all the beams. Specifically, it is defined as $\mathcal{SR} = \sum_{k=1}^{K} \min_{i=1,\ldots,N_u} \log_2\left(1 + \text{SINR}_k^{[i]}\right)$, where $\text{SINR}_k^{[i]}$ is the signal-to-interference-plus-noise-ratio of the i-th user at the k-th beam. This is defined as the ratio between the power of the desired received signal and the power of the noise and interference. It reads as:

$$\text{SINR}_k^{[i]} = \frac{|\mathbf{h}_k^{[i],H}\mathbf{w}_k|^2}{\sum_{j\neq k}|\mathbf{h}_k^{[i],H}\mathbf{w}_j|^2 + \sigma_n^2}, \tag{8}$$

where $\mathbf{w}_k$ is k-th column of $\mathbf{W}$. Note that since we are considering a multicast transmission, the achievable data rate at each beam is determined by the data rate that the UT with the lowest SINR can achieve. In this way, the selected MODCOD for transmission is decodable by all UTs in the frame. The modulation and coding selection can be made in practice thanks to the feedback that each user makes of its SINR, as it is indicated in reference [13].

As a matter of fact, the solution to the following optimization problem:

$$\begin{aligned} \mathcal{P}_1: \quad & \underset{\mathbf{W}}{\text{maximize}} \quad \mathcal{SR} \\ & \text{subject to} \\ & \left[\mathbf{W}\mathbf{W}^H\right]_{nn} \leq P \quad n = 1, \ldots, N, \end{aligned} \tag{9}$$

provides the maximum sum-rate that a multicast system can attain in terms of sum rate. We note that the objective function in problem $\mathcal{P}_1$ is nonconvex. Also, in $\mathcal{P}_1$ a matrix of around 10,000 complex elements shall be optimized over hundreds of per-antenna power constraints. The work in [17] considers the optimization of $\mathcal{P}_1$ via an alternating projection technique with a semidefinite relaxation procedure, which is adequate from small to medium coverage areas (i.e. small to medium number of beams). In case a notably larger number of beams and/or users are targeted, current nonconvex optimization alternatives might fail due to the immense computational complexity; thus, opening potential avenues for future research. This is the case of the optimization framework of [18], which tackles nonconvex objective functions over nonconvex constraints.

It is important to remark that, before tackling any precoder design, the scheduling process needs to be considered as it plays a key role in obtaining relevant sum-rate values in multicast precoding (see [17]and [19]); as it is crucial to select the most convenient users to be served in each satellite frame. This scheduling process could just simply consider the geographical position of each UT. In this way, information from UTs that are geographically close can be embedded into the same frame in order to yield efficient data rates. Note that geoclustering would not work in terrestrial context since geographical proximity does not imply similarity in channels. On the other hand, the GW can rely on the user channel vector and group users with similar vectors. This is the case of the minimum Euclidean norm scheduling technique (minEuclidean) presented in [13]. This technique randomly selects one user at each beam and; afterwards, it finds $N_u - 1$ users whose channel vector has the lowest Euclidean norm with respect to the randomly selected one. Although suboptimal, note that this scheduling process is done distributively for each beam with an affordable complexity. Further techniques for grouping users to be served over the same frame are an unexplored field of investigation. Using a graph to represent the UTs that are covered by the satellite, with their different features (i.e. channel quality, traffic, ...), allows to use spectral clustering as a possible alternative; which may render better performance when the optimal clusters are not convex [20].

To alleviate the complexity of $\mathcal{P}_1$, suboptimal precoding alternatives with lower complexity have been reported in the literature. One of the most adopted techniques, which gives a good performance versus computational complexity trade-off is the spatial minimum-mean-squared error (MMSE) precoder, which can be formulated for unit variance noise as $\mathbf{W}_{\text{MMSE}} = \beta_{\text{MMSE}} \left(\widehat{\mathbf{H}}^H \widehat{\mathbf{H}} + \frac{1}{P}\mathbf{I}_N \right)^{-1} \widehat{\mathbf{H}}^H$, where β_{MMSE} controls the transmit power to fulfill the per-feed power constraints and $\widehat{\mathbf{H}} = \frac{1}{N_u}\sum_{i=1}^{N_u} \mathbf{H}^{[i]}$. In other words, this design consists of the MMSE precoding over the average channel matrix of all users simultaneously served at each beam. This is one important peculiarity of the multicast problem when designing a low complexity precoder: to find out the best channel metric that represents N_u channels. Averaging is not suitable for all wireless channels, but it works well for the satellite channel, which has a strong line of sight and no phase variation across the user channel vector. This precoder is the so-called 'UpConst Multicast MMSE'. A comprehensive study of different linear precoding techniques for the general multigroup multicast problem can be found in [21]. As an example, Figure 9 shows the average beam data rate defined as

$$\mathcal{SR}\frac{B}{K}, \tag{10}$$

where B is the user bandwidth. The figure depicts the performance of both 'UpConst' multicast minimum mean square error (UpConst Multicast MMSE) and the block singular value decomposition (block-SVD) technique presented in [12]. The technique block-SVD is inspired in the seminal work in multiuser MIMO precoding of [22]. In particular, this technique aims at mitigating the interbeam interference by employing the subspace null projection.

For obtaining the results, we consider a beampattern with 245 beams and a maximum per-antenna power constraint of 55 W. The results have been obtained over 1000 Monte Carlo runs and unit variance noise. Figure 9 also depicts the average beam data rate whenever the GW performs a certain scheduling or not (i.e. random scheduling). In particular, we opt to use the minEuclidean scheduling mechanism.

Clearly, the larger N_u, the lower are the attainable rates obtained by both block-SVD and UpConst Multicast MMSE. In all cases, Block-SVD leads to larger data rates compared to the UpConst Multicast MMSE. In both cases the impact of scheduling is remarkable as key enabling component for yielding to a capacity increase with respect to the benchmark case (i.e. the four-color scenario depicted with the black curve in the figure).

Nevertheless, the computational complexity of UpConst Multicast MMSE is much lower than Block-SVD and does not grow notably when the number of N_u users per frame increases. On the contrary, block-SVD requires more computational time to compute the precoding matrix as the number of UT grows. In any case, despite its low computational complexity, UpConst Multicast MMSE still presents implementation challenges when serving large coverage areas (i.e., the computation of the matrix inverse becomes a computationally demanding

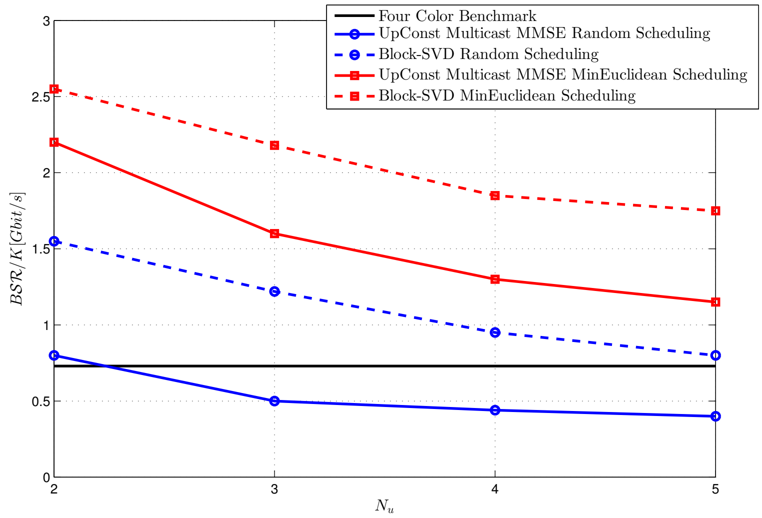

Fig. 9. Average beam data rate of two precoding techniques UpConst Multicast MMSE and block-SVD.

operation as K grows). Consequently, the study of alternative precoding designs is of extraordinary interest for both academia and industry. These alternative designs shall require only a limited number of operations, when computing its precoding matrices, while they provide large data rates.

Recently, the authors in [23] have proposed a linear low complexity multicast precoding alternative for the massive MIMO context based on a cascaded precoding design with a certain pilot allocation optimization. However, the presented work in [23] cannot be directly mimicked in the multibeam satellite context since the proposed mechanisms assumes a Rayleigh channel distribution and it takes advantage of the presence of pilot contamination, which does not occur in satellite communications.

Whenever higher layers are considered, the precoding design should be able to guarantee certain QoS to the UTs. In contrast to cellular systems, satellite operators offer their clients service level agreements (SLA) that involve a minimum data rate over a certain percentage of the channel access attempts. In this case, the fulfilment of the SLA contracts by precoding is done by optimizing the following problem:

$$\mathcal{P}_2: \quad \underset{\mathbf{W}}{\text{minimize}} \quad ||\mathbf{W}||^2$$

subject to

$$\left[\mathbf{W}\mathbf{W}^H\right]_{nn} \leq P \quad n = 1, \ldots, N,$$

$$\text{SINR}_k^{[i]} > \gamma_k \quad k = 1, \ldots, K \quad i = 1, \ldots, N_u.$$

The optimization problem $\mathcal{P}_2$ is a nonconvex quadratically constrained quadratic problem

(QCQP), which limits its applicability in large-scale coverage areas. This problem can be tackled via semidefinite relaxation (SDR) approximation methods. Bearing this in mind, efficient parallel implementation of the nonconvex QCQP optimization tools can be a good alternative for solving $\mathcal{P}_2$ in real multibeam satellite systems. This is the case for the work simultaneously done in [24] and [25], which promotes the use of the alternating direction method of multipliers (ADMM) to solve the nonconvex QCQP in $\mathcal{P}_2$. Still, it is open topic for research to find out good low-complexity alternatives.

The increase in the capacity of the user link requires a corresponding increase in the capacity of the feeder link. In principle, the exploitation of higher frequency bands by the wireless link could address this issue. However, often this approach is not feasible in practice. In the following, we propose different system architectures able to tackle the feeder link bandwidth limitation aspects.

D. Multiple gateways

Multibeam precoding over multiple GWs consists of transmitting the precoding signals over geographically separated GWs that are usually interconnected. In this way, the equivalent feeder link can aggregate the bandwidth of the feeder links of the different GWs and can accommodate the bandwidth increase that is needed when frequency reuse increases. Furthermore, in case one of the GWs fails or it has very adverse fading the traffic can be rerouted in order to maintain the system capacity. The configuration is depicted in Figure 10.

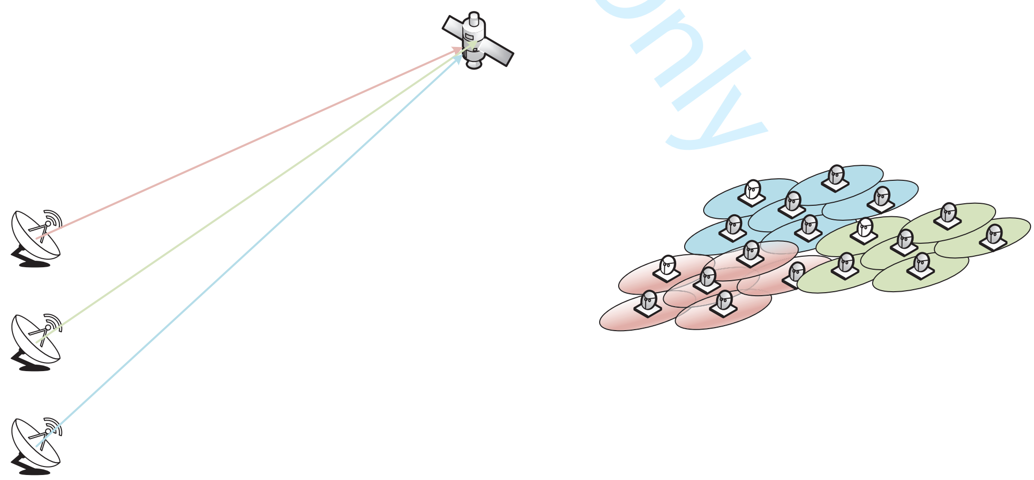

Fig. 10. Multigateway architecture configuration.

Again one can find a strong relationship between the multiple gateway mutlibeam scenario and a multicell C-RAN system, where each cell fronthaul has a limited bandwidth capacity as each GW has. Remarkably, challenges of both systems are still to be tackled as we describe in the following. In contrast to the single gateway scheme, multiple-GW precoding presents two main challenges. First, the original precoding matrix $\mathbf{W}$ becomes block-diagonal such that

$$\mathbf{W} = \text{block-diag}\left\{\mathbf{W}_1, \dots, \mathbf{W}_l, \dots, \mathbf{W}_L\right\}, \tag{11}$$

where $\mathbf{W}_l \in \mathbb{C}^{K_l \times N_l}$ is the precoding matrix associated with the l-th gateway $(l = 1, \dots, L)$. Note that for multiple-GW precoding $N = \sum_{l=1}^{L} N_l$, and $K = \sum_{l=1}^{L} K_l$. In other words, each GW can only use a subset of the N feed signals for performing the interference mitigation. This fact limits the overall system performance as it reduces the available degrees of freedom. On the other hand, each of the GW feeder link bandwidth requirements is reduced. Indeed, the l-th gateway only transmits $K_l N_l$ precoded signals instead of the KN signals that were transmitted in the single-GW scenario.

The second main challenge is the channel state information acquisition. Each GW can only access the feedback information from their served users, but each GW needs the channel state information of the adjacent beams to reduce the generated interference. Therefore, a set of matrices must be exchanged by the different GWs, leading to a large communication overhead [26], [27]. This exchange of channel state information (CSI) is the same sharing challenge as in the terrestrial multicell scenario.

Perfect connectivity between gateways might not be possible in real deployments. In this context, the multiagent optimization of $\{\mathbf{W}_l\}_{l=1}^{L}$ may be of interest to implement assuming certain Quality of Service (QoS) requirements between the different GW connections. This impacts not only on the tentative optimization, but also on the design of the compression algorithms for exchanging information from the different GWs. Finally, the precoding structure in (11) is similar to the group sparse beamforming presented in [28]. In light of this, promoting group sparsity in both $\mathcal{P}_1$ and $\mathcal{P}_2$ might result in an efficient multiple-GW precoding design.

E. Hybrid on ground-onboard precoding

Another alternative to reduce the bandwidth requirements in the feeder link is to carry out certain part of the precoding onboard the satellite. This architecture is coined as hybrid on ground-onboard and it offers a substantial feeder link bandwidth reduction [29].

Precisely, the original true on ground precoding matrix is transformed to

$$\mathbf{W} \rightarrow \mathbf{U}\mathbf{V}, \tag{12}$$

where $\mathbf{U} \in \mathbb{C}^{N \times K}$ is the beamforming operation done at the satellite and $\mathbf{V} \in \mathbb{C}^{K \times K}$ is the on ground beamforming implemented at the GW. In this context, instead of the original precoded signal $\mathbf{Ws}$ of NK user signals, with this hybrid onboard on ground architecture, $\mathbf{Vs}$ shall be transmitted instead and it is formed by K^2 user signals. Bearing this in mind, the feeder link bandwidth requirement is reduced by a factor $\frac{N}{K}$, which can be substantially high in multiple-feed-per-beam architectures. The price to pay for this *data compression* is a reduction in the final achieved throughput. As [30] studies, this reduction can be up to 20% and it is because the interference mitigation process acts on the K user beams directly instead of on all the N radiating elements; thus, reducing the available degrees of freedom for those cases for which the number of radiating elements is higher than the number of user beams.

In any case, we note that there are several design choices regarding the partitioning of the processing load between the ground and the space segment. Thus, the hybrid onboard-on ground precoding architecture presents additional challenges apart from the ones introduced in the previous section. Specifically, the onboard beamforming network does not have the processing capabilities as the one that is purely implemented on ground. Indeed, current payloads allow beamforming reconfigurability on a monthly or daily basis. This restriction translates into a difference on the transmitter channel state information: while $\mathbf{V}$ can rely on instantaneous channel state information, the onboard $\mathbf{U}$ shall rely on a forecast ergodic value of the channel matrix. Specifically, in [31] the authors study the capacity loss that comes from the fact that the onboard beamforming is fixed or it can only be updated in a very slow fashion.

Notably, the proposed scheme in (12) presents certain similarities to current works in mm-Wave precoding techniques, where hybrid digital-analog beamforming is under study in order to reduce the complexity that comes up when working with large apertures (i.e. with high number of elements). These large apertures are needed to achieve high gains at the mm-Wave. Remarkably, SatCom has been dealing with large scale apertures at mm-Waves for a long time; thus, developing low complexity architectures and implementations that were able to manage them. That is why the concept of beamspace communications (i.e. $\mathbf{U}$ is then the so-called beamspace matrix) arose. The beamspace is implemented with analog technology on the satellite, while the digital part of the precoder is implemented at the GW (see [32] and [33] for more details). In fact, beamspace communications are now being reconsidered in mm-Wave, but with phased-array technology (e.g. see [34]) instead of the usual array-fed reflector in SatCom.

The very inspiring recent work of [35] introduces the idea that $\mathbf{U}$ is a sparse matrix. Indeed, considering large coverage areas, beams that are sufficiently separated do not interfere to each other. Therefore, onboard beamforming can reduce its implementation costs and the feeder link can limit the number of transmitted precoded signals if $\mathbf{U}$ is designed to be sparse. This

TABLE I. THE BENEFITS AND SIGNAL PROCESSING CHALLENGES OF DIFFERENT MULTIBEAM SATELLITE PRECODING FUNCTIONALITY.

System Architecture	Benefits	Signal Processing Challenges	References
Single GW Precoding	System throughput increase.	User clustering, large-scale precoding optimization.	[12], [17]
Multiple GW Precoding	Per GW feeder link bandwidth reduction, gateway redundancy (outage reduction).	limited CSI GW sharing , large-scale distributed optimization, precoding under per feeder link bandwidth constraints.	[26], [27]
Precoding with on board beamforming	Feeder link bandwidth reduction, on board antenna cost reduction.	On board beamforming (beamspace MIMO) design with payload antenna restrictions.	[31], [35]

idea can be introduced for the multicast scenario in $\mathcal{P}_2$ by incorporating in the optimization the matrices $\mathbf{U}$ and $\mathbf{V}$ separately. This is a challenging problem due to the nonconvex constraints.

Finally, recent preliminary investigations show that the mass and cost of a satellite payload can be reduced severely if the onboard beam generation is done via sparse arrays. This restricts the $\mathbf{U}$ to have 0 or 1 entries, resulting in a very challenging variation of $\mathcal{P}_2$ optimization problem.

F. Summary of Precoding in Multibeam Satellite Systems

The following list summarizes the main peculiarities of precoding in multibeam satellite systems:

- antenna systems of extremely large scale dimensions
- multicast transmission nature
- per-antenna power constraints instead of sum-power constraint have to be considered
- more than one GW may serve one satellite in order to alleviate the capacity limitations of the feeder link
- the satellite system architecture and complexity imposes constraints that are different to the ones in the terrestrial segment, motivating hybrid onboard-on ground precoding architectures.

Finally, Table I summarizes the aspects that we have considered in this section, as well as their current signal processing challenges. Next section deals with SP on board the satellite, which introduces additional degrees of processing and performance improvement when compared to the traditional satellite approach that uses a transparent payload.

V. Onboard Signal Processing

The V/HTSs do not offer the possibility of hardware replacement once launched into the orbit. This motivated the traditional payload designs where the minimum necessary processing was performed onboard the satellite. Further, such processing was performed in analog domain due to their technological maturity in supporting high bandwidth while sustaining the constraints of satellite platforms, including power limitations, heat dissipation, and radiation. This resulted in the *transparent payload architecture* where the satellite is a passive relay performing only frequency translations (feeder to user link and vice-versa), channelization and amplification onboard.

With improvements in space hardened digital circuitry, a few satellites with digital onboard processing (OBP) are being designed [36]. In these satellites, the OBP is used to enable digital synthesis of narrowband user/beam specific carriers from the incoming wideband stream using filter banks. OBP further enables programmable routing of such carriers to end users. The aforementioned OBP functionalities suffice for most of the current transparent satellites based on link budget design. However, with the gradual design shift from link budget based to interference-limited paradigm, there is a need for considering advanced digital signal processing/communication algorithms for interference management. Similar to the advantages offered by relay processing in terrestrial networks, OBP offers additional degrees of freedom for implementing interference management algorithms efficiently. This motivates the discussion on OBP pursued in this section.

A. *State-of-the-Art in OBP*

Providing limited digital processing onboard the satellite is not a new concept and has been discussed in the last decades [37]. The key OBP paradigms observed from these developments can be categorized as follows:

- **Regenerative processing** is the straightforward way to OBP; it involves generating the digital baseband data on board after waveform digitization, demodulation, and decoding. This is similar to the *decode and forward* paradigm in relay systems and is considered for multiplexing different streams, switching, and routing [38]. Clearly, regenerative processing provides better noise reduction and flexibility. However, its complexity is rather large for V/HTSs due to the high bandwidths used. Needless to say, such processing needs to be reconfigurable to accommodate evolutions in the physical layer.
- A simpler approach to OBP is **digital transparent processing (DTP)**, which operates only on the samples of the input waveform. The *amplify and forward* architecture in

relaying is a simple DTP. Since neither demodulation nor decoding are implemented [37], DTP processing results in payloads that are agnostic to air-interface evolutions. Typical applications include digital beamforming, broadcasting/multicasting based on single channel copies, RF sensing and path calibration [37].

- An interesting **hybrid processing** paradigm involves digitizing the entire waveform, but regenerating only a part for exploitation. As a case in point, the header packet is regenerated to allow for onboard routing [39].

B. New Drivers for OBP: Signal Processing Algorithms

As discussed in Section III, satellite systems are evolving from link budget limited to interference-limited designs. Advanced signal processing like interference mitigation for such evolving scenarios are being accommodated currently through onground implementation at the GW. However, OBP provides additional degrees of freedom that can be exploited to yield an implementation of these signal processing algorithms more efficient than those implemented onground. The various advantages offered by relays with processing capabilities in terms of diversity, rate etc., are well explored in the terrestrial literature. However, OBP in satellite communications offers the following particular features that necessitates novel signal processing applications:

- *Latency Reduction*: Due to long round-trip delays between the terminals (terminal-satellite-terminal) on the ground, there is a large latency (250 ms) before the effect of onground processing at one communication terminal is discovered at the other. The delay can be reduced by half through OBP, thereby enhancing the efficiency of the underlying techniques. This feature is particularly useful for precoding to mobile terminals where round-trip delay severely affects fidelity of CSI at the transmitter; onboard precoding can reduce the degradation due to perturbed CSI.
- *Accessibility to Information*: Since the satellite aggregates information from multiple GWs or UTs before appropriate channelization, it has more information than the individual GWs or UTs. This enables joint processing onboard without the additional cost of sharing information across GWs/UTs. This is particularly useful in multiple GW scenarios where inter-GW interference can be easily mitigated on board.
- *Support to Evolved and Emerging Techniques*: Since many of the challenges and constraints arise on board the satellite, OBP possesses the wherewithal to address them. OBP extends support to predistortion and emerging techniques like full-duplex operations and antijamming techniques. For example, full-duplex relaying by satellite requires cancelling self-interference on board the satellite both in analog and digital domains

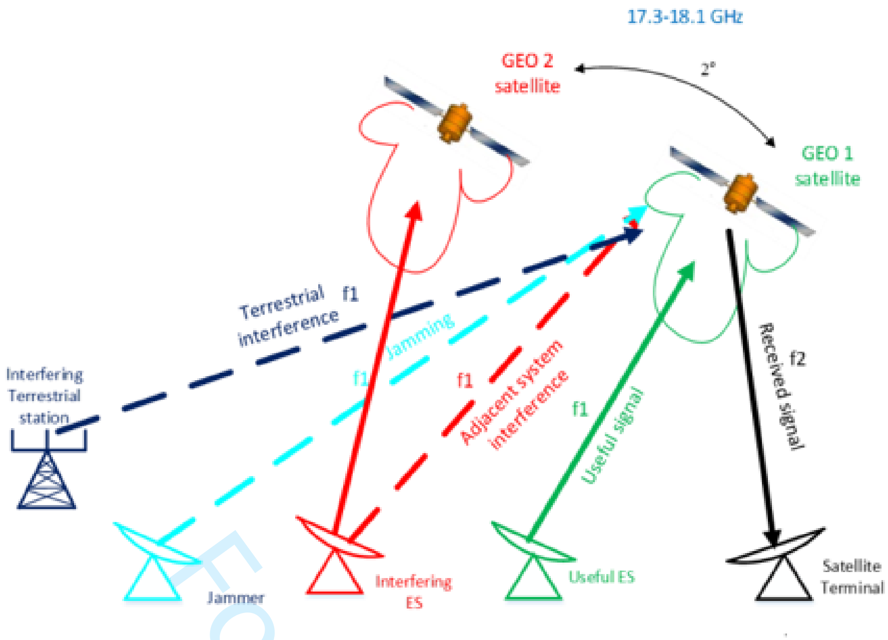

Fig. 11. Interference onboard the satellite.

[40].

The OBP techniques used thus far have focused on networking such as onboard switching, traffic routing, and multiplexing data/multimedia [38], with limited signal processing applications *per se*. However, novel signal processing techniques, as for instance the ones in Table I would greatly benefit from the additional degrees of freedom offered by OBP. Hence the emerging signal processing algorithms serve as drivers for further proliferation of OBP. The benefits of OBP are illustrated next through a simple signal processing application in interference detection.

Interference Detection: Exploiting Different Flavors of OBP As an illustrative example, we consider the detection of interference at the satellite generated from on ground terminals, either maliciously or due to improper installation. The scenario is depicted in Figure 11. Such an intersystem interference appears on the uplink; however, these unwanted transmissions corrupt the desired signal being relayed, thereby reducing the end-user SINR and impacting the operations significantly. Currently, such uplink RF interference is detected and localized onground from perturbed downlink transmission. Subsequently, the interference is mitigated by satellite operators using standard manual procedures. However, this methodology suffers from longer reaction times and degradation in performance. The latter arises from the fact that additional noise (receiver front end, propagation, etc...) and distortions from the satellite transponder corrupt the downlink signal that is used for onground detection [41].

On the other hand, onboard interference detection and localization can be undertaken by introducing a dedicated spectrum monitoring unit within the satellite payload that can take advantage of the emergent OBP capabilities. This provides for a faster reaction time and enhances detection capability since it avoids additional downlink noise and payload induced perturbations. For the sake of exposition, we consider the detection exercise herein and illustrate how the degrees of freedom offered by OBP can be leveraged.

We consider a generic scenario in which the satellite, the desired GW, and the interferer are equipped with a single antenna each. We further assume ideal sampling and digitization onboard. Under these conditions, detection of the uplink RF interference can be formulated as the following binary hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0 :&\quad \widetilde{x}_k(n) = hs_k(n) + \eta_k(n),\ \ n = 1, 2, \ldots, T, \\ \mathcal{H}_1 :&\quad \widetilde{x}_k(n) = hs_k(n) + \eta_k(n) + p_k(n),\ \ n = 1, 2, \ldots, T, \end{aligned} \tag{13}$$

where T is the number of samples used for detection, and h denotes the scalar flat fading channel from the desired GW to the satellite. Further, let $s_k(n)$ be the sample of the intended signal transmitted by the desired GW on the kth channel (or stream) at instance n; similarly, let $p_k(n)$ be the interfering signal onboard and $\eta_k(n)$ be the noise receiver modeled as a realization of independent and identically distributed (i.i.d.) complex Gaussian variable with zero mean and unit variance.

For such a problem, several interference detection techniques can be implemented onboard.

- *The conventional energy detector* (CED) technique works on samples $\widetilde{x}_k(n)$ directly and chooses a hypothesis based on $\sum_{n=1}^{T} |\widetilde{x}_k(n)|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_1$ where γ_1 is a threshold. [42]. Such a detector can be easily implemented with a DTP. While CED is shown to be effective for strong interference, it is rather susceptible to variations in the noise power.
- *Energy Detector with signal cancellation on pilots* (EDSCP) exploits the frame structure in the transmitted waveform to estimate the channel h on pilot symbols (known $s_k(n)$) [43]. Subsequently, the EDSCP detector performs the detection test, $\sum_{n \in \text{Pilots}} |\widetilde{x}_k(n) - \hat{h}s_k(n)|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_P$ for some threshold γ_P. Here $\hat{h}$ is the channel estimate obtained on the pilots through standard processing and $\widetilde{x}_k(n) - \hat{h}s_k(n)$ is an estimate of the interference $p_k(n)$ at the pilot locations. This is an example of *hybrid processing* where only the frame header is decoded to determine the location of pilots.
- *Energy Detector with signal cancellation on data* (EDSCD), initially proposed in [44] and further developed in [45], first detects $\{s_k(n)\}$ and subsequently removes its contribution from $\widetilde{x}_k(n)$. This facilitates an estimation of the interference, not only on pilots but also on data. The EDSCD detector takes the form, as $\sum_{n=1}^{T} |\widetilde{x}_k(n) - \hat{h}\hat{s_k}(n)|^2 \underset{H_0}{\overset{H_1}{\gtrless}} \gamma_D$

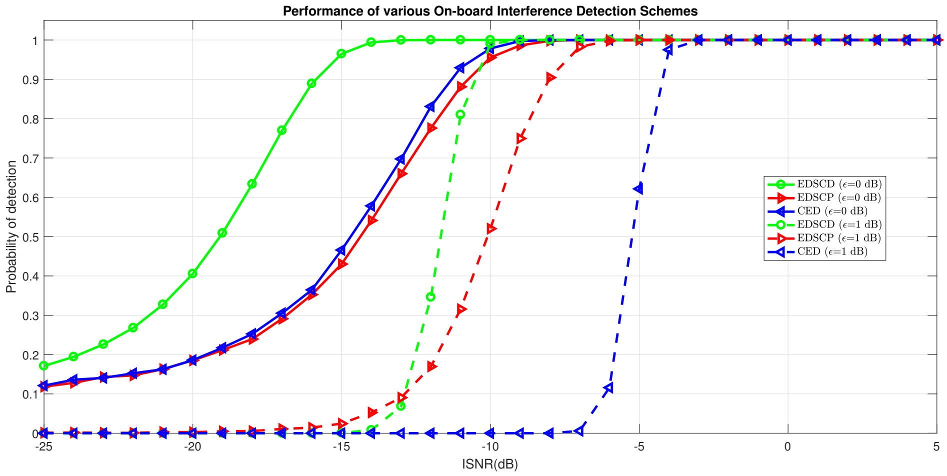

Fig. 12. Probability of detection versus the ISNR, QPSK modulation for $s_k(n)$, $T = 516, SNR = 6dB$.

for some threshold γ_D, where $\hat{h}$ is the channel estimate obtained on the pilots and $\{\hat{s_k}(n)\}_{n \notin \text{Pilots}}$ are the detected symbols. EDSCD requires regenerative processing.

Figure 12 presents the probability of detection as a function of the received interference-to-signal-plus-noise-ratio (ISNR) comparing the following detection schemes: i) CED, ii) EDSCP, and iii) EDSCD. ISNR is chosen as a performance metric to reflect the fact that interference is the signal of interest in this application while the other quantities are unwanted. In practice, there is an uncertainty of $0-1$ dB in the variance of $\eta_k(n)$ in (13); this uncertainty is represented as ϵ in the legend in Figure 12. For this set-up, we consider $T = 516$ modulated symbols comprising $T_d = 460$ data symbols and $T_p = 56$ pilots, representing a realistic waveform according to the DVB-RCS2 standard. It is observed that the interference detection performance decreases with uncertainty. The latter may lead to the ISNR wall phenomenon, where the detectors cannot detect the interference robustly beyond a certain ISNR value. Furthermore, we see that the EDSCP and EDSCD schemes perform considerably better than CED with uncertainty, improving the ISNR wall by more than 5 dB. It should be noted that all the algorithms can be implemented on ground; however, their performance would be penalized by additional payload impairments and downlink noise.

Thus, classical interference detection problems can be dealt with via different onboard architectures, with sophisticated processing providing additional performance benefits.

C. Architecture of an Onboard Processor

Having reviewed the state of the art and demonstrated the usefulness of OBP with an illustrative example, we now proceed to detail the OBP block of Figure 4. Such a system description will lead to modelling of key OBP blocks, enabling further understanding of the system and design of processing algorithms. The details are given in Figure 13, which presents a payload transponder employing digital OBP. Standard analog front-end receiver processing is carried out prior to the digital processing. These include filtering, low noise amplification, mixer and automatic gain control; these are used in down-converting the input RF signal to an appropriate intermediate frequency (IF).

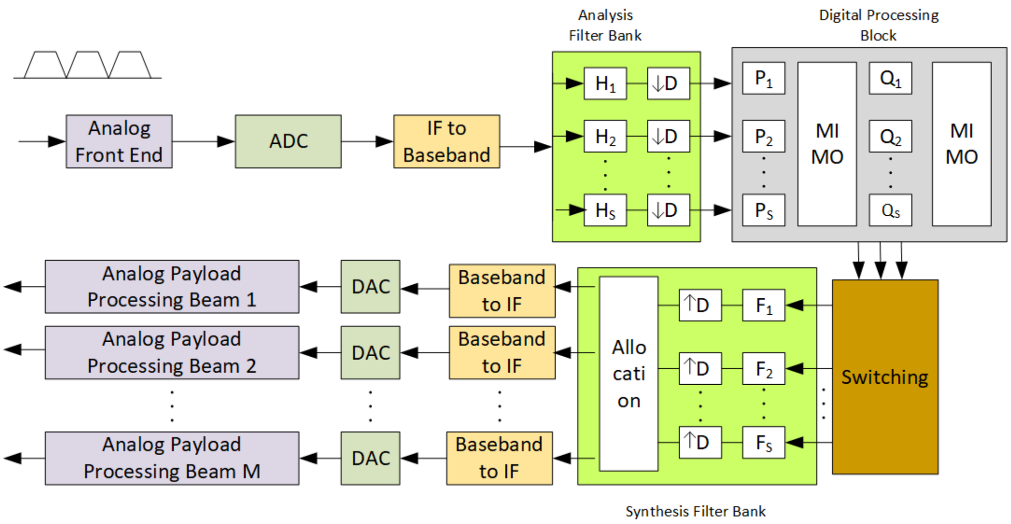

Fig. 13. Generic Architecture of Onboard Processor.

The key components in OBP are detailed below:

a) High-Speed Analog ADC and Baseband/IF Conversion: Assuming the IF signal with maximum bandwidth of $2f_c$ centered around f_c, ADCs sample at frequency $F_s \geq 4f_c$ to avoid aliasing. Subsequently, the resulting samples are converted to baseband (I/Q channels) using appropriate filtering [46].

b) Analysis Filter Banks: The baseband/IF signal is a multiplex of carriers of varying bandwidth, each carrier broadcast over a beam or addressed to a particular user. A first step towards dedicated processing is to demultiplex the digitized stream into user/ beam allocations. Towards this, the baseband/ IF input is spectrally decomposed using a filter bank, where the output of each filter corresponds to the smallest quantum of user bandwidth. Typically, non-critically sampled implementation of the analysis filter bank is considered and a polyphase structure is used. With reference to Figure **??**, the input signal is decomposed into S sub-bands using filters $\{H_k\}_{k=1}^S$, each having the smallest quantum of user bandwidth and the sampling factor is D with $D \leq S$. Further, the filters are modulated versions of a primary filter, leading to a fast Fourier transform (FFT) based polyphase matrix implementation.

c) Digital Processing Block: This generic block subsumes both transparent and regenerative architectures. Since the bandwidth allocation to users/ beams varies, the processing block is designed to work on the smallest carrier bandwidth used. It implements processing of individual streams (e.g., blocks P_i, Q_i) including predistortion, modulation/ demodulation and encoding/ decoding. The MIMO blocks impart joint processing capability among blocks P_i, Q_i, e.g., precoding. In the case of a transparent architecture, these blocks can implement waveform manipulation techniques on one or more outputs of the filter bank; typical examples include a look-up table (LUT) for predistortion, beamforming, precoding, and spectral analysis. Regenerative processing, in addition, mandates operations defined by the air-interface protocol for generating the information bits and the baseband signal.

d) Switching: The outputs of all transparent/ regenerative processing chains are input to a switch matrix that effects routing in spatial (e.g., from one beam to another), temporal (e.g., store and forward), and spectral (e.g., frequency hopping) domains. The switching block is implemented through controlled memory reads and writes.

e) Synthesis Filter Bank and DAC: A key component of the payload is the HPA, which typically amplifies RF signals from one or two beams. Towards generating the HPA input signals, the synthesis filter bank and DAC implement the process of converting the digital samples in baseband to IF and finally to the RF domain. Such an implementation is similar to their counterparts – ADC and analysis filter banks. With reference to Figure **??**, the synthesis filter comprises S filters $\{F_k\}_{k=1}^{S}$ and an *Allocation* block, which multiplexes relevant outputs to cater to the case when beams/users are allocated an integer multiple of the smallest quantum of bandwidth. Note that the filter banks without additional processing are designed to be near-perfect reconstructing pairs [47].

Critical to the implementation and operation of onboard processors is the accuracy of the processing chain. Due to the link budget constraints of satellite systems, mildly inaccurate processing chains can severely impact availability and performance. Hence, it is imperative to study various imperfections induced by the digital processing thoroughly and incorporate them in the design of onboard techniques. The non-idealities are listed below, and additional details can be obtained from [48].

- Quantization errors induced by the ADC conversion
- Non-idealities in filter implementation and use of fixed-point operations
- Impairments due to phase noise and carrier offsets

D. Signal Model and OBP Design Challenges

For ease of comprehension, we focus here on a DTP since the modelling of regenerative payloads is rather intractable. Let $x_k(t)$ be the analog signal corresponding to the kth frequency sub-band after the analysis bank. The signal $x_k(t)$ can be routed to any beam after appropriate switching. Assuming ideal filtering (i.e., rejection of out-of-band interference and adjacent channel interference, which would otherwise appear as in-band noise when using practical filters), a DTP would provide the designer access to the samples $\widetilde{x}_k(n)$ (at the input of the digital processing block in Figure **??**) where,

$$\widetilde{x}_k(n) \approx e^{j(2\pi[\triangle f]n+\theta+\omega_{n,k})} x_k(nT_s + \tau_{n,k}) + \eta_k(n), \tag{14}$$

where T_s is the effective sampling rate (after the downsampling), $[\triangle f]$ and θ are the frequency and phase offsets of the carrier assumed to be independent of k, $\tau_{n,k}$ is the sampling jitter, $\omega_{n,k}$ considers phase noise, and $\eta_k(\cdot)$ represents the noise before the processing block. It should be noted that $\eta_k(n)$ possesses a flat spectral density over the sub-band k only when the analysis filters are ideal. In addition, Doppler frequencies exist depending on the orbit; these can be included in $\triangle f$ as appropriate.

Let $g_{D,k}(\cdot)$ be the functional equivalent of the OBP (digital processing and filtering) for the kth stream. In order to include the joint processing and intersymbol interference introduced by filtering, $g_{D,k}(\cdot)$ can be typically modelled as a multiple input single output function with memory. Particularly, $g_{D,k}(\cdot)$ can take as input samples from different streams and time instances, i.e., $\{\widetilde{x}_l(m)\}_{l,m}$. Further, let $g_{P,l}(\cdot)$ denote the equivalent transfer function of the analog payload processing block for the lth payload. Note that, due to switching and allocation, the kth digital stream is processed by π_kth payload. In this work, we assume $\pi_k = k$. Hence the output of the kth sub-band analog payload processing takes the form

$$y_k(t) = g_{P,k}\left(g_{D,k}(\{\widetilde{x}_l(m)\}_{l,m}) + \widehat{\eta}_k(t)\right) + \widetilde{\eta}_k(t), \tag{15}$$

where $\widehat{\eta}_k(t)$ is the perturbation arising from the noise contributions from synthesis filter banks and DAC. Further, $\widetilde{\eta}_k(t)$ refers to the noise in the analog processing part of the kth sub-band. The presence of $\widetilde{\eta}_k(t)$ in (15) caters to the situation where the feeder link is non-ideal.

Due to the high carrier frequency, large bandwidths and adverse operating conditions, the perturbations introduced by the processing chain are significantly large and highly varying compared to those encountered in current terrestrial systems. This coupled with the limited link margins used in satellites necessitates that any OBP algorithm operating on the digitized samples must take cognizance of:

- Perturbations in the input waveform as presented in (14)
- Impact of processing on noise at the output of the payload as presented in (15)

In addition, due to limited power on board, the OBP algorithms need to have low complexity efficient implementations. These warrant modification of existing onground algorithms and make the OBP algorithm design and implementation challenging. In the ensuing section, a low-complexity OBP algorithm design to minimize impairments will be discussed.

E. Impairment Cognizant onboard Predistortion using DTP

As an illustrative example, we consider the implementation of predistortion on board the satellite. This is motivated by the significant effort expended in the SatCom community on predistortion techniques towards countering the nonlinear impairments introduced by the onboard HPA and achieving higher power and spectral efficiencies [49].

Signal predistortion (SPD), where the waveform is non-linearly transformed to mitigate the HPA non-linearities, has been considered in the literature (see references in [50] for a detailed list). These techniques have been traditionally implemented onground. By virtue of being a non-linear operation, SPD results in a bandwidth expansion; when implemented onground, this can have serious consequences in SatCom systems where transmitted signals have to satisfy tight emission masks. Alternatively, additional constraints need to be imposed on SPD to satisfy the mask, which leads to diminished gains. However, such limitations do not appear in onboard SPD due to its colocation with HPA. Moreover, the LUT approach provides for an attractive alternative for implementation onboard rather than functional evaluation.

1) Modeling Onboard SPD: We consider a simple payload model, where the analog processing prior to DTP is denoted by an input multiplexing (IMUX) filter used to remove out-of-band signals from the signal being processed. Further, the analog processing subsequent to the DTP comprises HPA and an OMUX filter for mitigating out-of-band emissions and naturally serves to regulate the output spectrum. With reference to the DTP architecture in Figure **??**, the block P_k implements SPD for the kth stream. Focusing on a generic kth stream and omitting the filter bank for ease of analysis, the resulting payload architecture can be simplified to the schematic shown in Figure14.

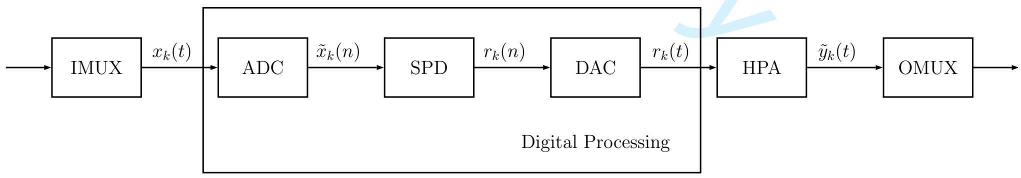

Fig. 14. Block scheme of the considered transponder with DTP.

In this setup, let $x_k(t)$ be the analog signal at the input of the ADC. An ideal ADC would perform the digitization, giving as output the sampled signal $x_k(n) = x_k(n/F_s)$, where F_s is the sampling frequency. However, this is seldom the case in practice; high-speed ADCs suffer from clock jitter caused by phase noise affecting the clock oscillator of the ADC. The jitter generates a non-uniform sampling of the input signal, leading to a degradation of the SNR inside the transponder [51]. Under some mild assumptions [52], the input of the SPD can be described as

$$\widetilde{x}_k(n) \cong x_k(n) + e_k(n)\dot{x}_k(n) + \eta_k(n) \tag{16}$$

where $\dot{x}_k(n)$ is the sampled first-order derivative of $x_k(t)$ [52] and $e_k(n)$ is the jitter error signal. The quantization noise and ambient thermal noise are included in $\eta_k(n)$. A typical SPD model involves a third-order memoryless polynomial function taking the form

$$r_k(n) = \gamma_k \widetilde{x}_k(n) + \delta_k \left|\widetilde{x}_k(n)\right|^2 \widetilde{x}_k(n). \tag{17}$$

On the other hand, the HPA can be modeled as a nonlinear non-invertible memoryless function approximated by a memoryless third-order Volterra series expansion as

$$\widetilde{y}_k(t) = \alpha_k r_k(t) + \beta_k \left|r_k(t)\right|^2 r_k(t), \tag{18}$$

where $r_k(n)$ is the sampled version of $r_k(t)$ using an appropriate sampling rate. Relating (17), (18) in the context of (15) (ignoring filterbanks), we can immediately recognize that $g_{D,k}(\cdot)$ is a single-input single-output memoryless function taking the form $g_{D,k}(\widetilde{x}_k(n)) = \gamma_k \widetilde{x}_k(n) + \delta_k \left|\widetilde{x}_k(n)\right|^2 \widetilde{x}_k(n)$ and $g_{PL,k}(r_k(t)) = \alpha_k r_k(t) + \beta_k \left|r_k(t)\right|^2 r_k(t)$. Further, $\widehat{\eta_k}(t), \widetilde{\eta_k}(t)$ are assumed to be zero. The aim is to determine the implementation gains of onboard SPD; central to such an implementation is the estimation of the predistorter coefficients minimizing a meaningful metric. This exercise is described below.

2) Parameter Estimation and SPD Implementation: As a first step, the HPA model parameters α_k, β_k need to be determined. They are jointly estimated by minimizing $\alpha, \beta = \text{argmin}_{\mathbb{C}^2} E_{r_k}\left\{|\widetilde{y}_k(t) - y_k(t)|^2\right\}$, where $y_k(t)$ is the output of the actual HPA and $E_{r_k}\{.\}$ denotes the expectation with respect to the signal $r_k(t)$. This is a least-squares minimization problem, where the error function is linear in the coefficients α_k and β_k. We choose the least-squares cost function since (i) it adequately represents the minimization of modelling errors and (ii) it is tractable leading to an elegant solution without the need for additional information.

The SPD parameters, γ_k and δ_k, can be found by resorting to a similar minimization problem, i.e., $\{\gamma_k, \delta_k\} = \text{argmin}_{\mathbb{C}^2} E\left\{|y_k(t) - x_k(t)|^2\right\}$, where $E\{.\}$ denotes the expectation with respect to the signal $x_k(t)$, the jitter $e_k(n)$, and the noise $\eta_1(n)$. Herein, we again choose the least-squares cost function since it represents the MSE between transmitted and received symbols and does not require additional information. However, now the error function is nonlinear in the coefficients γ_k and δ_k because of (18), and the method used to estimate the Volterra coefficients of the HPA can no longer be used. Several methods have been proposed in the literature, including direct learning [50], [53], which is based on least mean squares (LMS)/ recursive least squares (RLS), or divide and concur algorithm, which is based on message passing (see reference [54]).

While the SPD output $r_k(n)$ can be generated using (17), LUTs provide for a low complexity solution. Since the dynamics of the input signal, $\widetilde{x}_k(n)$ is typically known (from calibration tests for DTP), a LUT catering to this dynamic range can be calculated. Such a LUT implementation follows the memory-performance trade-off.

3) Results: Figure 15 compares the performance of the onboard SPD with its onground counterpart, where each has a similar structure [54]. The figure also illustrates the performance of onboard SPD optimized with and without the cognizance of jitter. Figure 15 shows that the onboard SPD outperforms its onground counterpart both in terms of SINR (providing a 0.4 dB gain at output back-off, i.e. OBO = 4 dB) and in terms of OBO (providing a 2 dB gain at SINR = 9.5 dB). The on-ground SPD cannot compensate for the distortions of IMUX due to its memoryless nature, a shortcoming not encountered by OBP. Figure 15 also shows gains when jitter statistics are accounted for during the optimization of the SPD coefficients. To put the SINR gains in perspective, it should be noted that the DVB-S2X standard allows the use of a MODCOD with higher spectral efficiency [6] for small SINR changes. Further, the OBO gain is of interest to satellite operators because it translates into a power-efficient amplification.

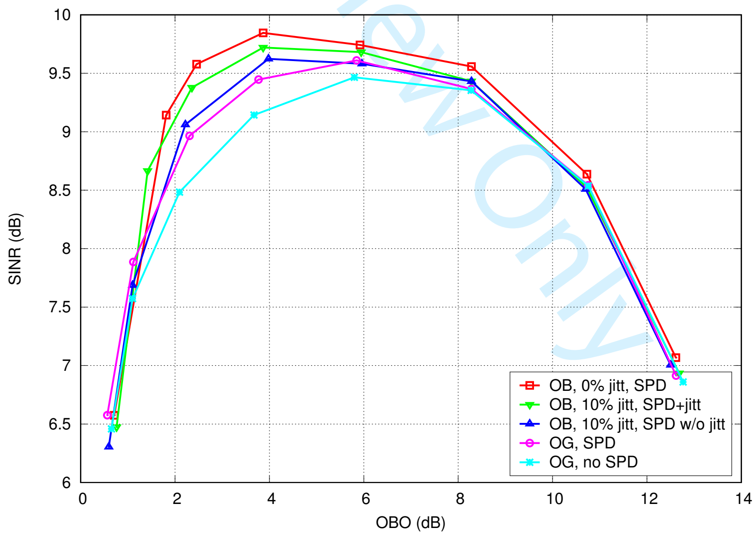

Fig. 15. SINR versus OBO for the jitter cognizant onboard SPD, compared to its on ground counterpart.

This example shows how OBP provides a platform for signal processing algorithms to exploit, providing gains over onground processing. This opens up avenues for further investigation including those on sophisticated multi-stream interference mitigation techniques like precoding.

VI. Conclusions and Future Prospects

V/HTSs offer far more advantages than their predecessors. This article covered the main signal processing challenges and tools that can be used to boost the spectral efficiency that these satellites can offer. We first discussed high frequency reuse in the multiple beams of these satellites and the role of on ground processing, at the GW. Next, with most of the traffic carried over being digital, signal processing onboard the satellite comes as a natural consequence; this overcomes some drawbacks or impairments of on ground processing. The focus of the article has been FSS. However, SatCom refers to a wide range of systems operating in various frequencies and providing different types of services. Future prospects for this work include the application and adaption of the explained signal processing techniques to the MSS. There are also high expectations in the so-called mega-constellations of MEO and LEO satellites. In these cases, the signal processing concerns are Doppler compensation, and inter-satellite communications, to mention just two important aspects. Non-GEO satellites require scanning user terminals that are more complex than the GEO ones. Finally, note that the feasibility of the proposed solutions depends very much not only on their complexity level but also on their compatibility with the legacy systems, full integration with the wireless terrestrial communications networks, user terminal complexity, and final adopted policy.

References

[1] G. Corazza, *Digital Satellite Communications*. Springer, 2007.

[2] J. N. Pelton, *Satellite Communications*. SpringerBriefs in Space Development, 2012.

[3] T. de Cola, S. Covaci, C. Ravishankar, and A. Perez-Neira, "Integration of satellite and 5g networks," *IEEE Network*, vol. 32, no. 5, September-October 2018.

[4] "Study on Scenarios and Requirements for Next Generation Access Technologies; (Release 14) 3GPP TR 38.913," June 2017.

[5] "Study on New Radio (NR) to support non terrestrial networks (Release 15) 3GPP TS 38.811," December 2017.

[6] ETSI EN 302 307-2 Digital Video Broadcasting (DVB), Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications, Part II: S2-Extensions (DVB-S2X), Available on ETSI web site (http://www.etsi.org).

[7] H. Fenech, A. Tomatis, S. Amos, V. Soumpholphakdy, and D. Serrano-Velarde, "Future high throughput satellite systems," in *2012 IEEE First AESS European Conference on Satellite Telecommunications (ESTEL)*, October 2012.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[10] E. Bjornson, E. G. Larsson, and T. L. Marzetta, “Massive MIMO: ten myths and one critical question,” *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[11] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, “Cloud ran for mobile networksa technology overview,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[12] M. Á. Vázquez, A. Perez-Neira, D. Christopoulos, S. Chatzinotas, B. Ottersten, P. D. Arapoglou, A. Ginesi, and G. Taricco, “Precoding in Multibeam Satellite Communications: Present and Future Challenges,” *IEEE Wireless Communications*, vol. 23, no. 6, pp. 88–95, December 2016.

[13] M. A. Vázquez, M. R. B. Shankar, C. Kourogiorgas, P. D. Arapoglou, V. Icolari, S. Chatzinotas, A. D. Panagopoulos, and A. I. Pérez-Neira, “Precoding, Scheduling and Link Adaptation in Mobile Interactive Multibeam Satellite Systems,” *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2018.

[14] C. B. Chae, T. Tang, R. W. Heath, and S. Cho, “MIMO Relaying With Linear Processing for Multiuser Transmission in Fixed Relay Networks,” *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 727–738, Feb 2008.

[15] G. Taricco, “Linear Precoding Methods for Multi-Beam Broadband Satellite Systems,” in *Proceedings of 20th European Wireless Conference*, May 2014.

[16] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, “Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups,” *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1268–1279, 2008.

[17] D. Christopoulos, S. Chatzinotas, and B. Ottersten, “Multicast Multigroup Precoding and User Scheduling for Frame-Based Satellite Communications,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 4695–4707, Sept 2015.

[18] G. Scutari, F. Facchinei, and L. Lampariello, “Parallel and Distributed Methods for Constrained Nonconvex Optimization Part I: Theory,” *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 1929–1944, April 2017.

[19] A. Guidotti, A. Vanelli-Coralli, G. Taricco, and G. Montorsi, “Clustering Strategies for Multicast Precoding in Multi-Beam Satellite Systems,” *arxiv.org/abs/1804.03891*, 2018.

[20] U. Luxburg, “A Tutorial on Spectral Clustering,” *Statistics and Computing*, vol. 17, no. 4, p. 395416, December 2007.

[21] Y. C. B. Silva and A. Klein, “Linear transmit beamforming techniques for the multigroup multicast scenario,” *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4353–4367, October 2009.

[22] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb 2004.

[23] M. Sadeghi, E. Bjornson, E. G. Larsson, C. Yuen, and T. L. Marzetta, “Max-Min Fair Transmit Precoding for Multi-Group Multicasting in Massive MIMO,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1358–1373, Feb 2018.

[24] M. Á. Vázquez, A. Konar, L. Blanco, N. D. Sidiropoulos, and A. I. Perez-Neira, “Non-convex consensus admm for satellite precoder design,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6279–6283.

[25] E. Chen and M. Tao, “ADMM-Based Fast Algorithm for Multi-Group Multicast Beamforming in Large-

Scale Wireless Systems," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2685–2698, June 2017.

[26] D. Christopoulos, H. Pennanen, S. Chatzinotas, and B. Ottersten, "Multicast multigroup precoding for frame-based multi-gateway satellite communications," in *2016 8th Advanced Satellite Multimedia Systems Conference and the 14th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, Sept 2016, pp. 1–6.

[27] V. Joroughi, M. Á. Vázquez, and A. I. Perez-Neira, "Generalized Multicast Multibeam Precoding for Satellite Communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 952–966, February 2017.

[28] Y. Shi, J. Zhang, and K. B. Letaief, "Group Sparse Beamforming for Green Cloud-RAN," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[29] J. Tronc, P. Angeletti, N. Song, M. Haardt, J. Arendt, and G. Gallinaro, "Overview and comparison of on-ground and on-board beamforming techniques in mobile satellite service applications," *International Journal of Satellite Communications and Networking*, vol. 32, no. 4, pp. 291–308, 2014. [Online]. Available: http://dx.doi.org/10.1002/sat.1049

[30] J. Arnau, B. Devillers, C. Mosquera, and A. Perez-Neira, "Performance study of multiuser interference mitigation schemes for hybrid broadband multibeam satellite architectures," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 132, Apr 2012. [Online]. Available: https://doi.org/10.1186/1687-1499-2012-132

[31] V. Joroughi, M. Á. Vázquez, A. I. Perez-Neira, and B. Devillers, "Onboard beam generation for multibeam satellite systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3714–3726, June 2017.

[32] A. Jacomb-Hood and E. Lier, "Multibeam active phased arrays for communications satellites," *IEEE Microwave Magazine*, vol. 1, no. 4, pp. 40–47, Dec 2000.

[33] P. Angeletti, N. Alagha, and S. D'Addio, "Space/ground beamforming techniques for satellite communications," in *2010 IEEE Antennas and Propagation Society International Symposium*, July 2010, pp. 1–4.

[34] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid Beamforming for Massive MIMO: A Survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[35] N. Song, T. Yang, and M. Haardt, "Efficient hybrid space-ground precoding techniques for multi-beam satellite systems," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 6284–6288.

[36] T. R. SES, "The changing DNA of satellites prompts a landslide of new products and services [Online]," 2016.

[37] P. Angeletti, R. de Gaudenzi, and M. Lisi, "From bent pipes to software defined payloads: evolution and trends of satellite communications systems," in *26th International Communications Satellite Systems Conference (ICSSC)*, 2008.

[38] T. M. Braun, *Satellite Communications : Payload and System*. Wiley-IEEE Press, 2012.

[39] P. Angeletti and N. Alagha, "Space/ground beamforming techniques for emerging hybrid satellite terrestrial networks," in *27th International Communications Satellite Systems Conference (ICSSC)*, 2009.

[40] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional mimo: Achievable rates under limited dynamic range," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3702–3713, July 2012.

[41] C. Politis, S. Maleki, S. Chatzinotas, and B. Ottersten, "Harmful interference threshold and energy detector for on-board interference detection," in *22nd Ka band and Broadband Communications Conference*, 2016.

[42] S. Atapattu, C. Tellambura, and H. Jiang, "Performance of an energy detector over channels with both multipath fading and shadowing," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3662–3670, December 2010.

[43] C. Politis, S. Maleki, C. Tsinos, S. Chatzinotas, and B. Ottersten, "On-board the satellite interference detection with imperfect signal cancellation," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2016, pp. 1–5.

[44] S. Stotas and A. Nallanathan, "On the throughput and spectrum sensing enhancement of opportunistic spectrum access cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 97–107, January 2012.

[45] C. Politis, S. Maleki, C. Tsinos, S. Chatzinotas, and B. Ottersten, "Weak interference detection with signal cancellation in satellite communications," in *IEEE ICASSP*, 2017.

[46] V. Sulli, D. Giancristofaro, F. Santucci, and M. Faccio, "Computing the hardware complexity of digital transparent satellite processors on the basis of performance requirements," in *IEEE ICC*, 2017.

[47] P. P. Vaidyanathan, *Multirate systems and filter banks*. Prentice Hall, Englewood Cliffs, 1993.

[48] V. Sulli, D. Giancristofaro, F. Santucci, and M. Faccio, "An analytical method for performance evaluation of digital transparent satellite processors," in *IEEE Globecom*, 2016.

[49] B. Beidas and R. Seshadri, "Analysis and compensation for nonlinear interference of two high-order modulation carriers over satellite link," *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1824–1833, June 2010.

[50] R. Piazza, M. R. B. Shankar, and B. Ottersten., "Data predistortion for multicarrier satellite channels based on direct learning," *IEEE Trans. Signal Processing*, vol. 62, no. 22, pp. 5868 – 5880, November 2014.

[51] B. Razavi, *RF Microelectronics*, 2nd ed., 2011.

[52] Z. Towfic, S.-K. Ting, and A. Sayed, "Clock jitter compensation in high-rate adc circuits," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5738–5753, November 2012.

[53] D. Zhou and V. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *IEEE Trans. Signal Processing*, vol. 55, no. 1, pp. 120–133, January 2007.

[54] N. Mazzali, M. R. B. Shankar, and B. Ottersten, "On-board signal predistortion for digital transparent satellites," in *IEEE SPAWC*, 2015.

Dear EiC, AE and Reviewers,

Please, find below (in blue) our explanation on how we have addressed your concerns and comments in the new version of the paper that we have submitted. Thank you for your time and considerations. The modifications have also been highlighted in blue in the new version of the article. Note that the number of references has increased from 50 to 54 in order to answer some of the reviewers' comments.

Reviewer: 1

Comments to the Author
The paper is well written and it is now better structured and focused. It is a little bit quite long, but it provides the readers a good introduction into the relevant signal processing related challenges for high-throughput satellites. This is an area that has not been addressed that much in the IEEE SP Community. I recommend this paper for acceptance.

Thank you.

Reviewer: 2

Comments to the Author
This submission appears to be a full paper, not a proposal. The paper appears to be a reasonable response to the earlier criticisms, as noted in the "reply-to-reviewers." However, I am not an expert in this area, and do not wish to fully review the current paper submission. A quick scan suggests to me that it looks OK for SPM, e.g., is not overly mathematical.

Thank you.

Reviewer: 3

Comments to the Author
I think the readability of the resubmitted paper has been improved. My comments for further enhancing readability are:

Thank you very much for your careful reading of the paper.

-"In this review, new avenues for research are identified": "review" to "article"; "avenues" to directions
Corrected.

-"In all these cases, the satellite acts as a very particular wireless relaying node, whose specifics lead to": what is "specifics"?
We have rewritten the sentence. By "specifics" we mean the channel, communication protocols, and complexity constraints of the satellite system, which create a set of particularities and that are explained in Section II.

- I think it is good to provide more references particularly in Sections I and II so that readers can refer to them if they want more information in this research field.
Due to your comment and that of reviewer 4, the references from [1] to [5] have been incorporated. Thanks.

-"For its part, VHTS systems": "V/HTS"
Corrected.

-"Fig. 1. Scheme of a multibeam satellite system. The forward link goes from the GW to the UTs via de satellite": what is "de"?
It was a typo, we mean “the” and it has been corrected.

-"User frequency plans for the scenario with 71 beams and frequency reuse of four": do you miss something after "four"?
We mean frequency reuse pattern equal to four and we have corrected it.

-"same data is transmitted": "is" to "are"
Corrected.

- Are "DVB-S2X" and "DVBS2/S2X" same?
DVB-S2X is the evolution of the standard for Digital Video Broadcasting with Satellite (DVB-S2). As the standard that takes advantage of all the features for V/HTS is DVB-S2X, we have unified the notation and refer only to "DVB-S2X". We have also introduced a reference to this standard.

-"extremely large size of the precoding matrix that must be calculated, (source: Viasat)": can you elaborate "extremely large size"? It is not clear how large it is.
The precoding matrix is 1000x1000 and to obtain it, the inverse of a matrix or singular value decomposition has to be computed. The details are explained later on in this section (e.g. Section IV.c). We have indicated this in the text.

-"there are served Nu > 1 UTs": this is difficult to understand
Corrected. We have simplified this paragraph in order to avoid introducing Nu. We introduce Nu later on in the Sub-section of System Model.

-"as the number of antenna in each base station grows large": "antenna" to "antennas"
Corrected.

-"precoding in multibeam satellite systems present certain similarities with respect to the cloud radio access network (C-RAN) architecture": "present" to "presents"
Corrected.

"Attending to the description in Figure 6": "Attending" to "Referring"
Corrected.

-"radiofrequency elements": "radiofrequency" to "RF"
Corrected and used the acronym RF in the rest of the article.

-"of 0.24 degrees and _PL;n which are the payload oscillator phase offsets which are assumed to be Gaussian with zero mean and standard deviation that is usually around 2 degrees": Can you provide reference(s) for "0.24 degrees" and "2 degrees"
The original source of these quantifications/measurements is a project conducted for the European Space Agency. However, as it is not a public document, we have preferred to incorporate reference [15] by G. Taricco.

-What is "feads" in Fig.8?
We mean “feeds”, now it is corrected.

-"transmitter, enable the precoding techniques that are proposed in next section": "enable" to "enables"
Corrected.

-"In this respect": "respect" to "aspect"
Corrected.

-As matrix F[i] is diagonal, do you mean "k;k" instead of "k:n"?
There was a mistake; F[i] is not diagonal. The property of this matrix is that all the elements in one of its columns are the same. It has been corrected and better explained in the new section.

-"de system" - What is "de"?
It was a typo, we mean "the" and it has been corrected.

-"signal to interference and noise ratio (SINR)" to "signal-to-interference-plus-noise-ratio (SINR)"
Corrected.

-"whose channel vector have the lowest Euclidean": "have" to "has"
Corrected.

-"which may render efficient an": not a complete sentence
We have corrected the sentence, which now reads as follows: "… which may render better performance when the optimal clusters are not convex.

-"performance vs. computational": "vs." to "versus"
Corrected.

-"quality of Service (QoS) requirements between the different GW connections. This": "quality" to "Quality"
Corrected.

-"straight-forward" to "straightforward"
Corrected.

-"on-board" or "onboard"
Corrected.

-"The details are given in Figure ??": What is "??"
Corrected, it is Figure 13.

-Add vertical "..." in Fig. 13
Corrected. As per the suggestions of the reviewer, vertical dots have been included in Fig 13 appropriately.

-"the bandwidth allocation to users/ beams vary": "vary" to "varies"
Corrected.

Reviewer: 4

Comments to the Author
This magazine paper describes the role of signal processing in satellite communications and also draws interesting connections between satellite methods and similar terrestrial methods. The paper is generally well written, and I only have some minor comments:

Thank you very much for your careful reading of the paper and also for your thoughtful comments, which have helped improving the paper, indeed.

Page 1: “providing connectivity anywhere” – isn’t this an overstatement? One of the main issues with satellite communications is the poor indoor converge, and the majority of the wireless data traffic is generated by indoor users.
You are right, we have rewritten the sentence to indicate its usefulness in unserved or under-served areas.

Page 2: I agree that SatCom can be viewed as a green communication technology if solar panels are being used, but that is not the same thing as having high “energy efficiency” as claimed in the paper. Energy efficiency is rather something that is measured in bit/J and it ignores if the energy is generated from renewable or non-renewable sources. I would guess that, due to the long distances and propagation losses, SatCom is not particularly energy efficient if measured in bit/J, so I suggest that you rephrase this part.
Thanks for the note, this is rephrased in the revised version.

Page 2: I thought the IMT 2020 requirements are the ones guiding what 5G will be and it does not contain anything about SatCom. I understand that terrestrial and satellite communications might be integrated better in the future, but is it really correct so say that it will be a part of 5G?
Thank you, you have a point. The actual standardization works are only study items (Sis), and since they have not ended up in work items (WIs) yet, it is hard to say that SatCom is definitely considered as part of 5G. Maybe it would be better to say xG, and x can be 5, 6,… G. Nevertheless, Satellite Systems can be a specific implementation of Rel 15 5G standards with its own benefits, and we our paper provides insights in facilitating that.

In order to clarify this aspect we have incorporate references [3-5] and rewritten a part of the introduction as follows: “While such a diversification of satellite-only services is foreseen to bear fruit, maximum benefits are envisaged by integrating satellite and terrestrial communications in the future generations of communications. For instance, the roles and benefits of satellites in the fifth generation (5G) [3] have begun to be studied in 3GPP Release 14, leading to the specific requirement to support satellite access being captured in TS22.261 - “Service requirements for next generation new services and markets; Stage 1”, recognizing the potential added value that satellite coverage brings, as part of the mix of access technologies for 5G [4-5].”

Page 2: You say things like “providing superior capacity” and “lower cost per bit” which are rather vague marketing-like statements. I think you should sharpen these statements, by clarifying what you compare against and provide some indicative numbers of the capacity and cost per bit that you are considering (or at least how large the improvements are in absolute or relative terms).

We have clarified this by stating that the spectral efficiency of V/HTS “(number of b/s/Hz) will be multiplied by two to three times with respect to the current non-high throughput satellites. In total, the expected aggregate high throughput will vary according to application served and satellite, but it is anticipated to be around 100 Gb/s. For very high throughput satellites the expected aggregate rate is within the range of Tb/s. This superior performance does not necessarily require higher bandwidth or increasing the weight of the payload. Therefore, a superior aggregate rate is provided at a lower cost per bit.

Page 5: “to get through the satellite” should perhaps be “to get through to the satellite”
Corrected.

Page 7: The coverage area of each beam in Figure 3 is very large, so it seems that the area throughput (bit/s/km^2) is very, very low in satellite communications as compared to terrestrial communications, even if the throughput per beam is high. With this in mind, it seems that satellite communications can never become more than a niche technology that can be used for broadcasting of television and other data streams that are of interest to many users, and to fill in coverage holes in terrestrial cellular networks. I suppose it cannot be continuously relied upon by massive numbers of users that request different data. Is this correct? I think the paper should better explain what role that satellite communications can really play in a 5G system, given its low area throughput.

Thank you, for rising this point. Let us comment that V/HTS, with their new technology that enables spot beams, can deliver different broadband data to different users. This improves the area data traffic they can support, compared to those linguistic beam satellites, which had a much wider coverage area per beam. It is true that the area throughput, measured in b/s/km^2, is still lower than that offered by terrestrial communications. In any case:

- SatCom has its own traditional markets: resilient overlay communications and disaster relief, governmental services, traffic off-loading and remote cellular backhaul provisioning (particularly for integrated access and backhauling which is a cornerstone of upcoming mobile communications generations), multicast services, and SCADA (supervisory control and data acquisition) for tele-supervision of industrial processes.
- With V/VHTS these markets can be enlarged to include specific mobile scenarios and content delivery; thus, not limiting to TV broadcast, but incorporating broadband.
- With 5G, which implies a new infrastructure concept that embraces cellular, fiber and satellite, SatCom is attractive to fill in coverage holes in terrestrial cellular networks, as you say. As 5G is not only about broadband mobile users, but also machine type communications (or Internet of Things), this opens new possible markets for SatCom. New business models have to be found, as the days of the TV broadcast cash cow are over.

we have rewritten some parts of the Section Introduction in order to clarify all these points.

Page 9: Please explain why “large investments are required for implementing…” Is it because no previous systems have been operating at these frequencies, so all the hardware needs to be designed from scratch?
You are right, we have clarified the sentence in the paper as follows: “However, large investments are required for implementing the communication subsystems in these bands because no previous commercial satellite communication systems have been operating at these frequencies.”

Page 12: The discussion around pilot contamination seems to compare massive MIMO with TDD operation and satellite communications with FDD operation. Isn't this the key difference and reason for not having pilot contamination?

Not really. The main reason is the presence of beams (i.e. due to the very directive gains in equation (6)), which changes the scenario for SatCom. We have rewritten this paragraph on pilot contamination to clarify this: "Massive MIMO in multicell scenarios entails difficulties in the channel estimation operation as base stations or users located in adjacent cells might inject interference into the estimation process, in the downlink and in the uplink, respectively. This is because orthogonal pilot sequences have to be reused from cell to cell, as in massive MIMO systems each cell has a high number of antennas and users active in the same time/frequency resource. Ideally one orthogonal sequence per transmitting antenna would be needed for each base station in the downlink and one per user transmitting antenna in the uplink, for time division multiplex, and this is not possible. Note that in frequency division multiplex, the users send their corresponding estimated channel to the base station. If this is done simultaneously by a large amount of users, signal contamination may also arise. In the multibeam satellite case, this is not the usual case, since in the downlink each UT has to estimate the channel from each beam, and therefore, only one pilot signal per beam is needed. Due to the high directivity of the beams, only the adjacent beams are the ones that create a relevant interference. This is a limited number of beams and, therefore, the pilot signals of adjacent beams can be orthogonal (i.e. note that DVB-S2X considers a set of 32 Walsh-Hadamard sequences that can be used as simultaneously sent pilots for the estimation of up to 32 channels in the downlink). For the same reason, in the uplink, either in frequency or in time division multiplex, the number of simultaneous transmitted signals for channel state information acquisition purposes is limited. Also, the satellite channel is, in general, non-frequency-selective and preserves the orthogonality at the UT. Pilot contamination could only appear in the case of having multiple satellites with independent processing of their signals." We have modified the text in the article in order to incorporate this paragraph.

Page 12: "(e.g. see [6]." There is a missing parenthesis here.
Corrected.

Page 13: Please elaborate on the channel model. It seems to be a flat-fading channel. Is that realistic in the typical use cases, even if the bandwidth is large? Is it single-carrier modulation of OFDM that is of interest?
You are right; it is a flat-fading model. The modulation is single-carrier. We have clarified this in the text as follows: "Single carrier modulated signals are transmitted, which is the most usual case in SatCom due to the degradation that the non-linearities of the high power amplifier create on multicarrier signals (e.g. DVB-S2X uses single carrier modulations)."

Page 13: Footnote 1: "Section" should be written with lower-case s since it is not a name.
Corrected.

Page 14: I would recommend you not to normalize the noise variance to one. This just makes it difficult to understand the channel model in (6) since it also contains the noise variance.
Corrected.

Page 15: Standard deviations of 0.24 degrees and 2 degrees are mentioned on this page. Do you have any sources for these numbers?
The original source of these quantifications/measurements is a project conducted for the European Space Agency. However, as it is not a public document, we have preferred to incorporate reference [15] by G. Taricco.

Page 16: What do you mean with "robust precoding strategies". Is this worst-case and probabilistic robustness, using S-lemma and such things. These are practically questionable frameworks since deterministic channels can be estimated to any accuracy with a negligible overhead and random channels can be treated by coding over the channel variations, rather than using outage-like robustness methods. I would omit this part of the text to not cause any confusion.
The part is omitted, thank you.

Page 17: The SR expression assumes an ideal modulation and coding scheme, so that the rate assigned to beam k is adapted to the weakest user's channel in that beam. How does the system acquire information about the SINR in (8) so that such modulation and coding selection can be made in practice?
Each user sends back its SNR as it is indicated in reference [13]. We have clarified this in the actual version.

Page 18: The second paragraph ends in the middle of a sentence.
Corrected.

Page 18: I'm unsure if the proposed MMSE precoder is appropriate for multi-casting or not. Since you compute the average over N_u users, it is the users with the strongest channel that will dominate in \hat{H}. But if you want all the users to be able to decode the signals, don't you want to do the opposite – to let the users with the weakest channels determine the beam direction? The strongest users will be able to decode the signal even if they don't get a beamforming gain.
In principle you are right and there exist some studies on that, which, for instance, carry out a weighted channel average. However, as we comment in the article, in the GEO satellite case, the users that are clustered for the multicast present very similar geographical location, therefore, they present a similar line-of-sight channel. Reference [12] comments more in detail on this matter. In any case, the problem is still open for further research.

Page 29: There is a missing ?? reference.
Corrected.

Page 31: Is it reasonable to assume that out-of-band interference can be filtered out? Some of this interference will leak into the band of interest and act as colored noise. That seems impossible to filter out.
Indeed, under practical filtering scenarios with realizable filters, these interferences perturb the in-band signaling. To emphasize this, we have updated the first paragraph of Section IV-D as, " Assuming ideal filtering (i.e., rejection of out-of-band interference and adjacent channel interference, which would otherwise appear as in-band noise when using practical filters), a DTP would provide the designer ….

Page 34: Several expectations on this page involve the norm of scalar quantities. Why are you using norms and not absolute values?

Indeed, it suffices to use absolute values as the concerned variables are scalars. Accordingly, two instances of Expectations involving norms has been replaced by absolute values.

General: The terms “broadcast” and “multicast” are both used in the manuscript. In some contexts, these things have the same meaning, while in other context they have not. Hence, I suggest that you clarify if broadcast means multicasting in your paper.
Thank you, in order to answer your question in Section Introduction we have said that “In both broadcast and multicast multiple users receive the same information using the same radio resources. While broadcasting refers to same information for the entire coverage area, multicasting refers to a beam or a set of beams. In this article we try to stick to this meaning, however, these terms are often used interchangeably in the SatCom community and broadcast may be sometimes used to encompass both.”

We have reviewed the use of these terms along the article in order to differentiate them accordingly, whenever possible. For instance DVB-S2X stands for “digital video broadcasting”, although it allows multicasting services.