Cross-layer design applied to small satellites for data collection

Vicente Almonacid Zamora

Cross-layer Design Applied to Small Satellites for Data Collection

Vicente Almonacid Zamora
CSU, Université de Montpellier

October 6, 2017

To Seban and Milovan.

Abstract

With the introduction of the CubeSat standard, the number of small-satellite missions has increased dramatically over the last two decades. Initially developed by universities and research centres for technology validation and academic experiments, these low-cost platforms currently allow to perform a variety of advanced, novel applications. In this thesis we are interested in the use of small satellites for global data collection and, more generally, for Internet of Things (IoT) and machine-to-machine (M2M) applications. Since both the space and ground segments are subject to stringent constraints in terms of size and mass, the overall capacity of the communications channel is highly limited, specially that of the uplink, which is a multi-access channel. These systems are also characterised by bursty, short messages, meaning that any protocol overhead may have a significant impact on the bandwidth efficiency. Hence, a random access approach is usually adopted for the uplink. Facing these challenges requires to optimize the communication system by taking an holistic approach. In particular, a joint design of both the physical (PHY) and Medium Access Control (MAC) layers is needed. The main contributions of this thesis are related to the study of Time– and Frequency–Asynchronous ALOHA (TFAA), a random access approach adopted in terrestrial ultra narrowband (UNB) networks. By trading data rate for communication range or transmission power, TFAA is particularly attractive in power constrained applications such as low power wide area networks and M2M over satellite. First, we evaluate its MAC performance (*i.e.*, its throughput and packet error rate) under three different reception models: the collision channel, the capture channel and a more detailed model that takes into account the PHY layer design. Then, we study the impact of PHY layer parameters, such as forward error correction (FEC), pulse shaping filter and modulation order, on the MAC performance. We show that, due to the characteristics of the multiple access interference, significant improvements can be obtained by applying low-rate FEC. To further improve TFAA's performance, we propose Contention Resolution Time– and Frequency–Asynchronous ALOHA (CR-TFAA), a more advanced design which is in line with recent developments such as Asynchronous Contention Resolution Diversity ALOHA (ACRDA). Under the same set of hypothesis, we see that CR-TFAA provides similar and even better performance than ACRDA, with a decrease in the packet error rate of at least one order of magnitude. Finally, we study the benefits that can be obtained by trading delay for MAC performance and energy efficiency, using simple techniques such as transmission control and packet-layer erasure coding.

Résumé

Avec l'introduction des plate-formes CubeSat, le nombre de petits satellites lancés dans l'espace a grandi de manière importante pendant les deux dernières décenies. Étant développés initiallement par des universités et des centres de recherche pour des simples tests technologiques ou des expériences académiques, ces plate-formes aujourd'hui permettent d'envisager de nouvelles applications et services. Dans cette thèse, nous nous intéressons à l'usage de petits satellites à défilement pour des réseaux globaux de collecte de données et, plus généralement, pour des applications de type *machine-to-machine* (M2M). En raison des contraintes existentes tant au segment sol comme au segment spatial, la capacité du canal de transmission est fortement limitée—notamment celle du lien montant, qui correspond à un canal à accès multiple. Ces réseaux sont aussi caractérisés par des très petits messages arrivant au système de manière imprévisible, ce qui implique que toute redondance liée au protocole a un impact important sur l'éfficacité spectrale. Ainsi, des méthodes d'accès aléatoires sont souvent préféres pour le lien montant. Relever ces défis nécessite d'aborder l'optimisation de la transmission de manière holistique. Plus spécifiquement, la conception des couches physiques (PHY) et de contrôle d'accès au support (MAC, de l'anglais *Media Access Control*) doit être menée de manière conjointe. Les principales contributions de cette thèse portent sur l'étude du protocole *Time– and Frequency–Asynchronous ALOHA* (TFAA), une technique d'accès aléatoire utilisée dans des réseaux terrestres à modulation de bande étroite. En réduisant significativement le débit binaire de transmission, TFAA permet notamment d'établir des liaisons à longue portée et/ou à faible consommation énergétique, dont des systèmes M2M par satellite sont un exemple. D'abord, nous évaluons les performances au niveau MAC (*i.e.*, le taux d'utilisation de canal et la probabilité d'erreur de packet) sous trois différents modèles de réception: le modèle de collisions, le modèle de *capture* et un modèle plus détaillé qui prend en compte les paramètres de la couche PHY. À partir de ce dernier modèle, nous étudions ensuite l'impact de certains paramètres de la couche PHY sur les performances au niveau MAC. Afin d'améliorer la performance de TFAA, nous proposons *Contention Resolution Time– and Frequency–Asynchronous ALOHA (CR-TFAA)*, une solution plus sophistiquée intégrant des techniques de suppressions successives d'interférences. Enfin, nous étudions les bénéfices obtenus en exploitant le compromis "performance–délai de bout-en-bout" en utilisant des techniques simples telles qu'un système de controle de transmission et le codage au niveau packet.

Contents

List of Figures

"I cannot define the real problem, therefore I suspect there's no real problem, but I'm not sure there's no real problem."

Richard Feynman (1918 – 1988)

1 Introduction

The number of small satellites being launched into space each year has experienced an exponential growth in the last two decades. This has been possible at a great extent thanks to the introduction of the CubeSat standard, which has helped to lower the costs for space access. Thus, many CubeSat-based, low-budget space missions have been successfully developed across the globe, in particular by universities. This trend represents an important paradigm shift, as traditionally space missions were exclusively carried by government space agencies—usually involving sizeable investments.

While the practical functions performed by the former CubeSats were limited to technology validation and academic experiments, the number of opportunities for new applications and services with enhanced capabilities is currently increasing fast. Indeed, the success of the CubeSat approach for space access has given support for the development of a rich technological ecosystem involving start-ups, universities, research centres and space agencies. In the longer term, we expect to see CubeSats (or perhaps a novel class of miniaturized platforms) playing active roles in the space sector.

This thesis takes part of research and development efforts of the *Centre Spatial Universitaire Montpellier-Nîmes* aiming at conceiving novel nanosatellite applications. More specifically, the work presented here is motivated by one prospective application: global data collection from small, unmanned terminals. Although store-and-forward microsatellites in the low Earth orbit (LEO) have been developed for similar applications during the 1990s and early 2000s, the growing demand for machine-type data communications is creating a momentum for a new generation of

data collection networks based on low-cost nanosatellite platforms. From a broader perspective, it is likely that such networks will become an important component of the Internet of Things (IoT) and Machine-to-Machine (M2M) communications.

While there are high expectations in the potential of small satellites such as CubeSats, there are still a number of challenges that must be overcome. In particular, mass and size limitations of these platforms affect the capabilities of every subsystem. Regarding the communication subsystem, this translates into reduced antenna gain, transmission power and signal processing capabilities. Note that, in the IoT/M2M context, similar constraints are observed in the ground segment. On the other hand, the LEO environment brings additional difficulties in the design of a robust communication link. Furthermore, since in a data collection network many terminal nodes can be simultaneously active within the coverage area, the terminal-to-satellite link (*i.e.*, the uplink) requires special attention. Current CubeSat-compatible communications solutions—which are typically based on commercial-of-the-shelf (COTS) components—do not fulfil with the requirements imposed by massive machine type communications. Hence, the development of new approaches to communications system design is encouraged.

The proliferation of Low Power Wide Area Networks (LPWAN) in the terrestrial communications sector has brought new design paradigms. Instead of maximising the capacity achieved by each user—as in conventional cellular systems—these networks trade data rate for communication range and energy-efficiency, opening a variety of new application possibilities. It is easy to see a connection between terrestrial LPWAN and machine type communications with LEO satellites, which suggests that recent designs adopted in the former could also be applied in the latter. In particular, the Ultra Narrow Band (UNB) approach has caught the attention of the author.

1.1 Problem Statement

Most CubeSats radios are based on the AX.25 link layer protocol and binary digital modulations (*e.g.*, frequency shift keying and phase shift keying), with data rates ranging from 1200 to 9600 [bps]. While this configuration may provide a relatively robust link for point-to-point transmissions, it is not optimal for the multiuser machine-type communications scenario described above. On the other hand, since the traffic in these applications is characterised by very short, unpredictable messages, it is desirable to adopt contention-based access schemes and to minimise the communication overhead. Under such conditions, providing both an spectral-efficient and a highly-reliable uplink channel, while keeping cost, complexity and energy consumption low, remains a fundamental problem intersecting communications engineering and information theory. The work developed in this thesis is a step forward towards this major objective.

Due to the stringent system constraints, the design of an appropriate communications system demands a systemic perspective, *i.e.*, a *cross-layer* design (CLD) approach. More specifically, a joint design of both the physical (PHY) and Medium Access Control (MAC) layers is needed.

Following these principles, the UNB technique loomed as a good candidate solution, as several advantages were claimed (*e.g.*, reduced terminal cost, long communication range, energy efficiency and asynchronous operation). A unique characteristic of UNB networks is that packets are transmitted freely (that is, without predefined slot boundaries) in both time and frequency domains. However, a complete analysis of its performance, in particular regarding its MAC throughput and packet error rate was not available in the literature. Thus, evaluating these figures of merit—which provide essential insight on the overall capacity—was one of the primary objectives of this thesis.

Despite of its advantages, as the access scheme in UNB is contention-based, an important fraction of the packets transmitted through the network is lost due to collisions. As a result, the network spectral efficiency is harmed. This leads to a more specific problem: that of optimising the MAC performance in networks where the packets are transmitted randomly in both time and frequency domains.

Traditionally, collision resolution in random access multiple access was achieved through MAC-level procedures. Examples of these are well-known algorithms such as the binary exponential back-off and the tree algorithms. While these solutions allow to recover packet collisions and to guarantee system stability (provided that the overall packet arrival rate is below the maximum system capacity), they do not achieve high-throughput performance. In addition, resolving collisions at the MAC level is usually not energy-efficient due to the relatively high amount of retransmissions and signalling overhead required to transmit a packet successfully. As a consequence, this solution is not convenient for machine type communications. The CLD approach takes relevance in this context, as novel techniques allowing to resolve packet collisions at the physical layer have been introduced recently. Evaluating the performance gains obtained through this approach in UNB networks is another problem that is addressed in this thesis.

1.2 Outline of Contributions

An important portion of the work developed in this thesis was devoted to the study and enhancement of the underlying access scheme used in UNB networks, referred to as time- and frequency-asynchronous ALOHA (TFAA). In particular, we evaluate the TFAA performance under three different scenarios: the collision channel model, a capture channel based on the signal-to-interference-plus-noise ratio and a more detailed channel considering the actual multiple access interference. The effects of PHY-layer parameters such as modulation, forward error correction (FEC) and pulse shaping filter are also evaluated through computer simulation. The main results obtained on this subject were published in [AF17c].

It is worth noting that, at the moment when the work on this subject was started, performance evaluation of UNB networks in the LPWAN context was only addressed in [DGG14a; DGG14b]; at least to the author's knowledge. However, the approach followed in those publications did not take into account the system dynamics and was based on the *Gaussian assumption* for modelling

the multiple access interference, which motivated further research. Interestingly enough, two other PhD thesis were started on the subject at about the same time. As a result, the analysis of the TFAA throughput over the collision channel model was done independently in [Ant+15], [GM16] and by the author himself in [AF17c].

A second contribution was the introduction of Contention Resolution Time- and Frequency-Asynchronous ALOHA (CR-TFAA), a high-throughput random access scheme that allows to resolve the packet collisions through the exploitation of time and frequency diversity, low-rate FEC and successive interference cancellation techniques. The performance of CR-TFAA was obtained through simulation and semi-analytical methods, and the results contrasted with a similar technique. In particular, it was found that CR-TFAA provides better packet error rate performance for practical packet arrival rates. These results were published in [AF17a], and a journal version is also being prepared at the moment these lines are being written.

In addition to the work focused on UNB/TFAA LPWANs, a discussion regarding some general aspects of data collection networks based on nanosatellite platforms was presented in [AF16], of which a journal version was subsequently published in [AF17b]. In particular, a delay tolerant approach to multiple access based on packet-layer coding is discussed, including some trade-offs relating throughput, delay and energy efficiency.

1.3 Structure of this Thesis

This research has been developed following a general-to-specific approach. It is the author intention to reflect this process in this thesis report. On the other hand, since this research has been carried on a multidisciplinary R&D environment, it is likely that it will serve system designers as well as non-expert readers. For this reason, an important effort has been done to provide a text that is as self-contained as possible, while compact at the same time. For the expert reader, Chapters 2 and 3 may be optional.

We start by introducing store-and-forward LEO satellite networks for data collection in Chapter 2. Historical aspects as well as perspectives for future applications such as IoT/M2M communications are considered. The use of low-cost CubeSats platforms for data collection is also discussed. This chapter also describes a generic network model and provides some details regarding the specificities of the LEO communications channel.

Chapter 3 provides general background related to communications engineering, including network architecture, link budget and multiuser protocol design. A generic reference design is discussed and some protocols proposed in the literature are briefly described. A discussion on the overall network capacity shows the importance of the uplink design.

Chapter 4 introduces random access protocols and provides basic definitions and concepts that will be employed throughout the rest of the thesis. In addition, a review of the state of the art is provided, focusing on solutions designed for satellite communications.

In Chapter 5 we propose a delay-tolerant approach to random access. Here, the idea is to evaluate the gains that can be obtained by trading delay for MAC performance and energy-efficiency. Moreover, the use of end-to-end higher-layer coding, which is completely transparent to the satellite, is discussed.

Chapter 6 provides a general analysis of TFAA focusing on the throughput and packet error rate performances. As mentioned, three different channel models are considered. The modelling considerations presented here are then used for the study of CR-TFAA. There is also an analysis of the multiple access interference characteristics at the symbol level.

Chapter 7 presents simulation results showing the impact of PHY-layer parameters (*i.e.*, modulation order, FEC coding rate and pulse shaping filter roll-off factor) on the TFAA MAC performance. These results provide some insight regarding an optimal PHY layer design, and show that TFAA benefits greatly from FEC.

Chapter 8 describes CR-TFAA and presents the performance results obtained through semi-analytical and simulation results.

Finally, Chapter 9 summarizes this thesis and provides insights for future work.

Part I

Background

" I can't change the laws of physics, captain!"

Lt. Commander Montgomery Scott

2

Store-and-Forward and Data Collection Small Satellite Systems

Contents

The use of small, low Earth orbiting store-and-forward satellites to provide global communications and data collection services involves many interesting engineering challenges. Some of these will be introduced in the following. In addition, we provide general background information that is relevant to describe the context within which this research has been conducted.

2.1 Introduction

Satellites, or any object orbiting around Earth, can be classified according to their orbit. An orbit whose altitude range is between 160 [km] and 2000 [km] is called a *low Earth orbit (LEO)* and satellites in this orbit are typically referred to as *LEO satellites*. In practice, most LEO satellites are launched at altitudes below 1000 [km], since at higher altitudes the Van Allen radiation belts compromise the satellite payloads. Among the most typical orbits we find also geostationary orbits (GEO) and medium Earth orbits (MEO). Satellites in GEO are placed at an altitude of 35786 km and are typically employed to provide broadband telecommunication services, as well as broadcast services such as television and radio. Satellites in MEO are placed in between LEO and GEO, and are mostly employed for navigation (*e.g.*, the Global Positioning System) and communications services. Compared to MEOs and GEOs, LEOs have the advantage of involving shorter propagation delays—which make them particularly attractive for real-time mobile communications—and the disadvantage of a more limited coverage or *footprint*. As a consequence, many LEO satellites are required to cover the entire Earth continuously. Several networks involving large constellations of LEO satellites have been proposed to provide global, real-time mobile communications, usually involving important investments, though some of them have failed as they proved uneconomical.

Alternatively, LEO satellites have also been widely used to provide non-real-time data communication services. In contrast to real-time LEO networks, non-real-time LEO networks usually consist of just one or a few *small*[1] satellites—which is why they cannot provide instantaneous global coverage. Consequently, they are considerably less expensive. In order to link two arbitrary points on Earth, these satellites are equipped with *store-and-forward (S&F)* communication payloads. Thus, a source node willing to transmit a message to a given destination node must first wait until a satellite visits the area and a link can be established. The message is then transmitted on the Earth-to-space link (called the *uplink*) and stored by the satellite until the destination node is reached and the message is forwarded on the space-to-Earth link (called the *downlink*). Note that the uplink and downlink are fundamentally different: the former is a *multi-user* or *multiple access* channel, while the latter is a *broadcast channel*.

Since a satellite on a high-inclination LEO orbit is able to visit any point on Earth one or several times a day, a single satellite suffices to provide global S&F communications services. For

[1] In general, satellites whose mass is below 500 kg are referred to as small satellites.

instance, a satellite in a 800 [km] circular polar[2] orbit with a period of 100 minutes visits any point near the equator approximately once a day. By contrast, it visits any point near the poles about 14 times a day.

A *Data Collection* satellite system is a network in which satellites are used to gather data from a set of nodes (also referred to as *terminals*) located anywhere on Earth. These terminals are typically low-cost, unmanned sensors equipped with small transmitters. Unlike generic S&F networks, which allow to establish an intermittent, two-way channel between any pair of nodes in the network, data collection networks may only allow to establish unidirectional channels between the sensor nodes and one or several central Earth stations where the data is gathered.

In the following subsections, we will review historical as well as current and future scenarios in S&F Data Collection satellite systems. New trends in the market, in particular the convergence towards miniaturized platforms will also be presented. Section 2.2 provides a general model of a S&F Data Collection satellite network and highlights some of its particularities. Then, some of the main characteristics of the LEO satellite channel are briefly presented in Section 2.3. Finally, Section 2.4 summarizes this chapter.

2.1.1 An Historical Perspective

Launched in 1958 as a response to the Russian satellites Sputnik 1 and 2, SCORE was the first American satellite and the first satellite capable of providing communication functions. SCORE carried both a real-time communications payload—by which it was possible to link two nodes within the satellite footprint—and a S&F communications payload. Soon after, William Brandon proposed the idea of using LEO satellites to provide *global* communications services. His concept was materialised through the construction of the Courier satellite, launched in 1960. The project was particularly successful in demonstrating that a single, small LEO satellite was sufficient to provide world-wide coverage, provided that a high-inclination orbit is adopted. However, at the time there was not enough demand for such services and no further developments occurred during some decades. Research and development was mainly devoted to big GEO satellite platforms.

In 1973, Brandon publishes a design proposal [Bra73] for a digital S&F satellite communications system, which he called the Data Courier Satellite System. His paper identifies key issues and lays the conceptual foundations for the actual systems deployed over a decade later. In particular, Brandon suggested that a S&F satellite network using low-cost microsatellites could be an attractive, commercially viable enterprise.

Launched on 1984, UoSAT-2, also known as OSCAR-11, was the first satellite carrying a digital S&F communications payload and was developed by the University of Surrey. UoSAT-2 was placed in a polar LEO at an altitude of 700 [km], and surprisingly, has remained active for over thirty years after its launch [Osc]. Along with the S&F communications payload, UoSAT-2 carried a

[2] A polar orbit has an inclination around 90° with respect to the equator.

few experiments to gather data regarding the impact of the LEO environment in digital electronic components. A comprehensive review of the UoSAT-2 mission is provided in [Swe87].

After UoSAT-2, the University of Surrey and its spin-off company Surrey Satellite Technology Ltd (SSTL) developed and launched several S&F payloads on microsatellite platforms. Most of these satellites used continuous phase frequency shift keying (CPFSK) at rates ranging from 9600 to 38400 baud/s, and operated over narrowband channels in the VHF and UHF frequency bands. While the first systems were meant to interconnect relatively expensive ground terminals equipped with tracking antennas, it was soon demonstrated [All95] that much cheaper terminals equipped only with an omnidirectional antenna could also be employed with acceptable quality of service.

During the 1990s, two main trends were distinguishable within the LEO satellite industry. On the one hand, an important part of the industry and academia was focused on developing technology for highly-capable big LEO satellites. This trend was reinforced by important investments in large LEO constellations, as for instance, the Iridium constellation. On the other hand, SSTL and other groups were oriented to low-cost platforms offering data collection and messaging services.

With the success and rapid expansion of second-generation terrestrial cellular networks, the development of S&F LEO satellites for communications services started to decline. By the early 2000s, SSTL, who was highly involved in this market, changed its strategy and started to focus on remote sensing applications.

Satellite data collection systems have been in operation since 1978, after the creation of the Argos system [Cla89]. Argos was jointly developed by the National Aeronautics and Space Administration (NASA) and the Centre National d'Études Spatiales (CNES). In contrast to S&F communications systems, the demand for data collection has been growing slowly but steadily over the past decades. In addition to Argos, several low-cost solutions were proposed in the early 1990s, as for instance OHB's Satellite for Information Relay (SAFIR) [GL92] and Sierracom's Global Environmental Data Distribution System (GEDDS) [Wal91]. All these systems target scientific environmental data collection from fixed and mobile terminals located in remote areas. Argos, for its part, reached a saturation point by the end of the 2000s, which motivated research for capacity optimization techniques [Far11]. Currently, Argos' fourth-generation system is being deployed [Sar15].

2.1.2 From Microsatellites to Nano and Picosatellites: The Rise of CubeSats

While the former S&F systems were installed on microsatellites platforms weighing around 50 kg and involving an overall cost of several million USD (in today's dollars), it is currently possible to develop similar applications on even smaller platforms like nanosatellites (with a mass between 1 and 10 kg) and picosatellites (with a mass between 0.1 and 1 kg). Furthermore, the introduction of the *CubeSat* design standard [PSTA01] in the early 2000s has significantly reduced the barriers

for space access allowing to develop extremely low-cost applications.

CubeSats are designed upon building blocks or *units* of 10 cm^3 and a maximum payload mass around 1 kg. This is the smallest standard size and is referred to as 1U. Typical sizes range between 1U and 3U, though other bigger platforms have also been proposed. Naturally, a limited mass and size translates into reduced capabilities. A typical 1U CubeSat, for instance, cannot carry deployable solar panels and its overall useful power generation capacity is in the order of 1W. These stringent energy constraints limit communications capacity as well as computing and signal processing power. Also, it is difficult to implement active attitude determination and control systems (ADCS) over such small platforms, which makes the use of directive antennas ineffective. For these reasons, early CubeSats projects were limited to academic applications and space technology validation. However, from 3U+, it is already possible to include high-efficiency deployable solar panels. For instance, a 4-face solar panel using triple junction cells allows to achieve a maximum power of about 30 W. Thus, the range of applications that we can currently envision is becoming increasingly larger.

There are three main reasons why CubeSats can be developed, launched and operated at reduced cost. First, most CubeSat projects employ commercial off-the-shelf (COTS) components, which are usually non-space-qualified. This allows to avoid developing custom subsystems but comes at the cost of limited performance and robustness. Second, non-commercial CubeSat missions usually employ amateur frequency bands, which can be obtained for free. Finally, and more importantly, they are launched as secondary payloads and without strict requirements in terms of scheduling and orbital parameters. In addition, most CubeSats projects are developed by university groups in relatively short periods of time (typically between 6-12 months) with relaxed reliability constraints.

Currently, there are no doubts about the success of the CubeSat design concept. As of march 2017, more than six hundred CubeSats have been launched into space [Nan].

2.1.3 State-of-the-Art on CubeSat Communications Engineering

Most of the early CubeSat communication systems were based on COTS transceivers employing amateur radio technology. In particular, the AX.25 link layer protocol and binary FSK transceivers have been widely employed [KL13]. These systems operated in amateur VHF/UHF narrowband channels at rates between 1200-9600 bps. As the experimental data generated by CubeSat missions became increasingly larger, it was soon observed that more communication capacity was required. As a consequence, universities and research institutions, as well as industrial suppliers, have been focusing on finding alternatives to improve CubeSat communication capabilities.

Current trends in CubeSat communication system engineering include ([Jon13; Pal+14; Cla+09; MV10; SAM13]):

- Moving from AX.25 to more efficient link layer protocols specially designed for bulk data

transfer in *delay* or *disruption tolerant networks (DTN)*, as for instance the Consultative Committee for Space Data Systems (CCSDS) File Delivery Protocol (CFDP) [Ccs], the Licklider Transmission Protocol (LTP) [Lic] and Saratoga [Sar]. These protocols allow to optimize the system *goodput*[3] in conditions of intermittent connectivity, low signal-to-noise ratio and long round-trip delays.

- Shifting towards higher frequency bands as the L-, S- and even the X-bands. While transmitting signals at higher frequencies entails a more important attenuation due to free-space loss (see Section 2.3.3), it also allows to exploit high-diameter dishes at the Earth stations and hence, to improve the link budget. Note that increasing the transmission frequency also increases the Doppler effect (see Section 2.3.2), which would require some compensation measures.

- The use of commercial (*i.e.*, non-amateur) frequency bands. Currently, most amateur VHF/UHF bands are overcrowded and significantly affected by interference. By acquiring a dedicated channel it is therefore possible to improve the signal-to-noise-plus-interference ratio and increase the data rate. On the other hand, larger bandwidths can be allocated in non-amateur bands.

- The use of near-Shannon capacity forward error correction (FEC). Several FEC codes allowing to achieve near-Shannon capacity over additive white Gaussian noise channels have been proposed (*e.g.*, turbo codes and LDPC codes). The high coding gain given by these codes comes at the cost of bandwidth and decoding complexity. Since the complexity of the encoder is minimal with respect to that of the decoder, these codes can be integrated in the downlink channel without involving significant processing power at the satellite end.

- Adopting flexible radio payloads. As we will see later on, the LEO satellite channel is characterized by a time-varying signal envelope and a severe propagation environment. In these conditions, the use of static modulation and coding schemes is highly inefficient. In [Sie13], an adaptive coding and modulation CubeSat protocol capable of dynamically adjusting the transmission parameters according to the channel conditions is proposed. Another interesting trend related to this is the adoption of software defined radio (SDR) platforms (see for instance [Oli+12; Kne+12]), which provides additional flexibility in terms of protocol configuration.

Another constraint on the amount of data that can be collected from CubeSats missions is due to the fact that there are usually no many access opportunities since the number of ground stations

[3]In this context, we define goodput as the effective average bit rate that is achieved at the application layer during a communication session.

available is limited. Indeed, there are around 2-4 access opportunities per day from a given ground station location to a CubeSat in LEO. In order to cope with this problem, two approaches have been proposed: first, the creation of global networks of ground stations, called federated ground station networks (FGSNs) [Spa+12]. Examples of these are the Global Educational Network for Satellite Operations (GENSO) [LPST07] and the SatNet project [Sat]. Note that, in order to exploit FGSNs, it is required to use standard radio transceivers that are compatible with the ground station hardware available. FGSNs are therefore best suited for typical CubeSats using amateur radio technology. An interesting low-cost solution to further improve the performance of nanosatellite projects exploiting FGSNs was recently proposed by the European Space Agency (ESA). The idea was studied in [Var+15] and consists on using a peer-to-peer BitTorrent-like protocol to distribute space data between the ground nodes in a FGSN. Simulation experiments showed that the proposed protocol allows to improve the reliability and latency of the system.

As an alternative approach to deal with limited communication availability, it is possible to exploit network interoperability with other satellite systems, such as, for instance, the Iridium [RBB16] and the Globalstar constellation [Bab+13]. The advantage of this approach is that almost near-time connectivity can be achieved. However, there are still several technological issues that remain to be tackled (*e.g.*, antenna design, tracking systems, availability of CubeSat-compatible inter-satellite radios, etc.).

It is worth noting that all the aforementioned strategies are focused on increasing the downlink communication capabilities. In fact, most CubeSats generate data from on-board instrumentation and the uplink is mainly used to transmit commands to the spacecraft. Hence, the amount of data that is transmitted on the uplink is usually low and there is no need for high data rates.

2.1.4 Towards M2M/IoT over Low-cost CubeSat Platforms

The Internet of Things (IoT) and machine-to-machine (M2M) communications currently represent the next evolution of the Internet ecosystem. A recent forecast by ABI Research [Abi] indicates that by 2020 75% of the total number of connected wireless devices—which is expected to be near 50 billion—will correspond to small sensor nodes or other types of communicating objects. In this context, satellites are expected to play an important role as enablers of ubiquitous communication services and as a means to further improve reliability where terrestrial networks are already present.

IoT applications in which the terminal nodes are located in remote areas without terrestrial network coverage and that rely only on space-based communications services have been called Internet of Remote Things (IoRT) [De +16]. In particular, the authors in [De +16] discuss three main IoRT application scenarios in which satellite systems are fundamental: smart grid, emergency management and environmental data collection. The first two are characterized by more stringent requirements in terms of reliability and latency and are therefore best served by traditional satellite network constellations. On the other hand, we note that, in most cases, environmental data

collection does not require real-time nor high-rate communications and is therefore one prospective application scenario in which S&F nanosatellite systems loom as a cost-effective solution. Furthermore, there is a growing demand for environmental data collection services, specially for climate change research and monitoring.

With today's nanosatellite technology, it is possible to develop S&F communication and data collection systems similar to those developed during the 1990s over microsatellite platforms at even lower costs. Several papers describing design proposals and actual projects based on nanosatellite platforms are available in the literature. In [AKZ07], for instance, a S&F data collection and tracking system using a single 1U CubeSat in LEO is proposed. The communications payload is based on the Automatic Packet Reporting System (APRS) and operates in the VHF (uplink) and UHF (downlink) bands. The uplink radio interface employs FSK and the AX.25 protocol, with a data rate of 1200 [b/s]. In [AKZ08] the authors study a similar system, focusing on the channel access and automatic repeat request (ARQ) retransmission schemes. It is found that a maximum of $\sim$ 150 APRS packets can be transmitted to the satellite during a single passage of about 10 minutes, while ensuring stable operation. A more sophisticated architecture is considered in [Bed13], where a full DTN protocol stack is proposed, including a novel multiple access convergence layer called ALOHAGP (ALOHA with gateway priority).

Among actual implementations we may mention HumSat [TP+14], a nanosatellite program developed by several universities and designed to provide communications services for emergency and humanitarian applications, as well as environmental data collection services for climate change research. The first operational satellite, HumSat-D, was a 1U CubeSat launched in 2013. After HumSat-D, two additional nanosatellites carrying HumSat S&F payloads have been launched. In order to improve its capacity and reduce latency, HumSat exploits the SatNet FGSN.

Currently, several commercial IoT/M2M systems based on nanosatellite platforms are being developed. AstroCast[4], a system designed by the startup ELSE, will employ a constellation of 64 3U CubeSats and will be operational by the end of 2018. Along with ELSE, many organizations are planning to develop similar systems in the short term, among which we may mention Kepler Communications[5], Terran Orbital[6], Magnitude Space[7] and Helios Wire[8], all of which will offer low data rate, low-power IoT/M2M services.

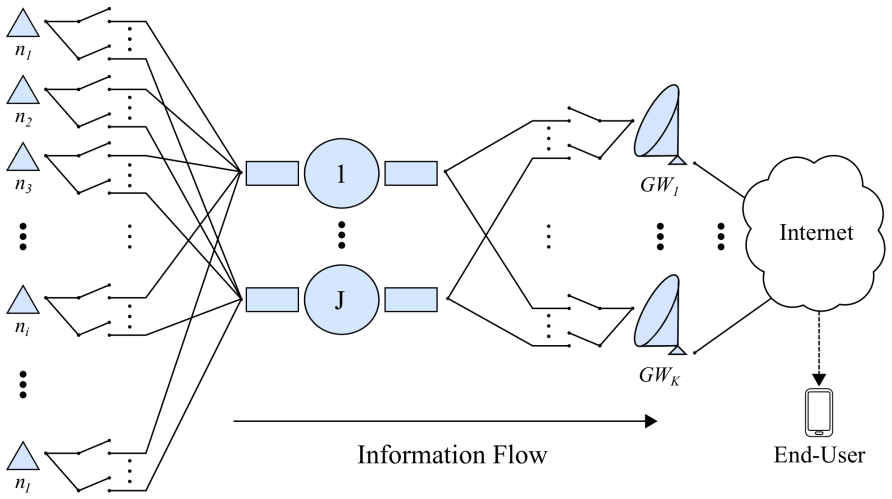


Figure 2.1 – Network topology.

2.2 System Overview

A typical S&F data collection satellite network is composed by a potentially large number of small ground terminals (which could be fixed or mobile), one or a few satellites in LEO and one or more central Earth stations, or gateways, where the data is gathered. The network topology is shown in Figure 2.1. The ground terminals can be located anywhere on Earth, though they are more likely installed in remote, non-urban areas. Also, in many applications, terminals are unmanned and autonomous (*i.e.*, battery-powered), and therefore power and energy constrained. When a terminal has packets to transmit, it waits until any of the satellites is visible and then transmits its packets on the uplink, following some multiple access scheme[9]. Packets that are successfully transmitted on the uplink are stored in a buffer and then forwarded to the first gateway visited by the satellite. At this point, we may assume that the information can be immediately retrieved by the end-users, given that the gateways are connected to the Internet.

Due to size limitations, a typical ground terminal can only be equipped with an omnidirectional antenna. Similarly, it is difficult to implement high-gain antennas on-board the satellites, specially

[4] `http://www.astrocast.net/`
[5] `http://www.keplercommunications.com/`
[6] `http://terranorbital.com/`
[7] `https://magnitudespace.com`
[8] `http://helioswire.com/`
[9] The choice of this scheme has important implications on the system capacity, which is why it will be the main subject of this thesis. The multiple access problem will be treated in the next chapter.

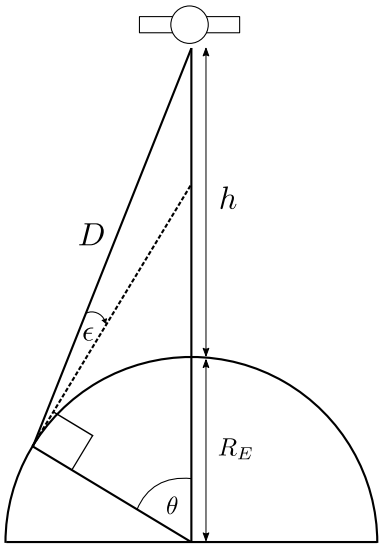


Figure 2.2 – Geometrical model of a LEO satellite system.

if the system operates over low radio frequency bands (*i.e.*, VHF/UHF bands) or if the satellite is too small and a stabilization system cannot be implemented. Hence, omnidirectional antennas are typically employed also. With bigger platforms (*e.g.* microsatellites) and higher frequencies, the satellite receiving antennas can be designed in order to cope with the important variations of free-space loss (see Section 2.3.3) experimented by terminals at different locations within the satellite footprint [War93].

The satellite visibility from a given point on earth is limited to a time interval T_w that we call the *visibility window*, and we define it as the time-window over which the satellite is seen with an elevation angle ϵ that is above some threshold ϵ_{min}. Due to irregularities in the Earth surface (*e.g.* buildings, mountains), a communications link is typically established above $\epsilon_{min} = 10°$. T_w also depends on the maximum elevation angle ϵ_{max}, which varies across different satellite passages according to the orbital parameters. Given ϵ_{min} and ϵ_{max}, T_w can be approximated as [AADH99]

$$T_w \approx \frac{2}{(\omega_s - \omega_E \cos i)} \cos^{-1} \left[\frac{\cos\left(\cos^{-1}\left(\frac{R_E}{R_E+h}\cos\epsilon_{min}\right) - \epsilon_{min}\right)}{\cos\left(\cos^{-1}\left(\frac{R_E}{R_E+h}\cos\epsilon_{max}\right) - \epsilon_{max}\right)} \right] \tag{2.1}$$

,

where ω_s and ω_E are the satellite and the Earth angular velocities in the Earth-centered inertial coordinate frame, i is the satellite's orbit inclination, R_E the Earth's radius and h the satellite altitude, as shown in Figure 2.2. As an example, if we consider an orbit inclination $i = 45°$ and a satellite altitude $h = 650$ [km], the visibility window for a typical passage with maximum elevation

angle $\epsilon_{max} = 60^\circ$ is about 9 minutes. Note that, since ϵ_{max} varies from one passage to the next, some access opportunities may be much shorter.

From Figure 2.2 we may derive the satellite–terminal distance D as a function of ϵ using simple trigonometric relations:

$$D(\epsilon) = \frac{R_E + h}{\cos\epsilon} \times \sin\theta, \tag{2.2}$$

where θ, called the Earth central angle, is given by

$$\theta = \cos^{-1}\left(\frac{R_E \cos\epsilon}{R_E + h}\right) - \epsilon. \tag{2.3}$$

The *coverage area* (also referred to as the *access area*) is the set of points on Earth from which the satellite is visible. By contrast, the *footprint* is the area that can be actually seen by the satellite antenna [Wer01]. The coverage area A can be estimated by first computing the central angle θ setting $\epsilon = 0$ in Eq. (2.3), and then using the well-known formula for the surface of a spherical cap:

$$A = 2\pi R_E^2(1 - \cos\theta). \tag{2.4}$$

For instance, if we take again the same LEO configuration considered above ($h = 650$ [km]), we get $\theta \approx 25^\circ$, which represents a coverage area $A \approx 24 \times 10^6$ [km^2]. This is in contrast to what we observe for a GEO satellite, in which case $h = 35786$ [km], $\theta \approx 81^\circ$ and $A \approx 2.17 \times 10^8$ [km^2]. However, since GEO satellites are fixed in the Equatorial plane, they cannot cover latitudes above/below $\theta \approx \pm 81^\circ$.

2.3 The LEO Satellite Channel

Radio signals transmitted between a satellite and a ground station are subject to several sources of impairments. On the one hand, there are some effects that are specific to the Earth–space propagation environment which are caused by the interaction between the information-bearing electromagnetic wave and different particles in the atmosphere. These effects include depolarization, absorption, refraction, scattering, rain attenuation—among others—, and depend strongly on the operating frequency band. On the other hand, since LEO satellites move at high velocities with respect to ground stations and mobile terminals, the LEO satellite channel is also characterised by time-varying Doppler shift. Similarly, as the transmitter–receiver path length and the elevation angle change significantly during the visibility window, there is also an important variation in the received signal strength due to the combined effect of path loss and differences in both transmitter and receiver antenna gain. As a result, the phase, amplitude and frequency of the received waveforms are distorted, which compromises the reliability of the communication link.

We will review some of these issues in more detail in the following subsections, as well as some modelling frameworks provided in the literature.

2.3.1 Atmospheric Propagation Effects

There are two layers in the atmosphere that have a significant impact on radio wave transmission: the troposphere, which lies at 0-20 [km] above the Earth surface, and the ionosphere, at about 75-1000 [km]. The effects from other layers are negligible. As mentioned, atmospheric effects also depend on the frequency band. For frequencies over 3 GHz, tropospheric effects are predominant. These include rain attenuation and depolarization, principally. However, since most small satellite communications systems do not operate at frequencies beyond the S-band, tropospheric effects can be ignored. On the contrary, below 3 GHz the effects of the ionosphere become more important. A framework to quantify these impairments is provided by the International Telecommunication Union (ITU) in its ITU-R P.531-12 recommendation [Itu].

Ionospheric effects manifest in two ways: background ionization—caused by the electron content cumulated along the signal path—and ionospheric irregularities—which correspond to random, localized changes in the propagation medium. The background ionization is characterized by the total electron content (TEC), defined as:

$$N_T = \int_S n_e(s) dS \tag{2.5}$$

where s is the propagation path in meters and n_e is the electron density measured in e/m^3. Background ionization is related with Faraday rotation and group delay. Both effects depend on N_T and are inversely proportional to the square of the frequency. Faraday rotation changes the signal polarization and is often mitigated by the use of circularly polarized antennas. Group delay may result in dispersion, as different frequency components of the signal propagate at different speeds. For wideband systems, this is an effect to take into account.

Ionospheric irregularities, on the other hand, are related with scintillation events caused by changes of the refractive index of the propagation medium. These effects can only be described statistically and are usually quantified by peak-to-peak fluctuations. As depicted in [Itu], scintillation events depend on several factors, including geographic location, time of the day, season and solar activity, showing maximums at the vernal equinox and appearing typically at sunset. Measurement data shows that, in general, 1 dB peak-to-peak fluctuations do not exceed 5% of time.

2.3.2 Doppler Effect

Doppler Shift

The instantaneous Doppler shift can be expressed as:

$$\Delta f_D(t) = \frac{|\mathbf{v}(t)| f_c}{c}, \tag{2.6}$$

where f_c is the signal frequency, c the speed of light and $\mathbf{v}(t)$ the relative velocity between the satellite and the ground terminal. In typical narrowband systems operating at relatively low frequency bands (*e.g.* VHF), Doppler shift may be neglected (at VHF, its maximum is about 2 kHz). However, for higher transmission frequencies it is sometimes required to implement some kind of compensation technique, either by adjusting the transmitter or the receiver frequency. A closed-form expression for the Doppler shift as a function of time, the orbital parameters and the maximum elevation angle is derived in [AADH98].

Doppler Rate

During the visibility window, $\mathbf{v}(t)$ changes continuously. Its maximum magnitude $|\mathbf{v}(t)|$ is observed at the start and at the end of the visibility window—when the elevation angle ϵ reaches ϵ_{min}—, while its minimum $|\mathbf{v}(t)| = 0$ occurs when ϵ reaches ϵ_{max} and the satellite is at its closest distance to the terminal. The Doppler shift changes accordingly. Thus, the maximum and the minimum values occur at $\epsilon = \epsilon_{min}$ and $\epsilon = \epsilon_{max}$, respectively. The Doppler rate is the rate at which the frequency changes, as observed by the receiver. For a typical LEO satellite link operating in the UHF band, the maximum Doppler rate is about 100 [Hz/s], while it can exceed 700 [Hz/s] in the L band.

2.3.3 Free-space Loss Variation

As an electromagnetic wave propagates through from transmitter to receiver, its energy tends to scatter through space, even in the absence of obstacles. If we take as a reference an isotropic radiator, the energy received at a point located at a distance D is attenuated by a factor

$$L_{FS} = \left(\frac{4\pi D}{\lambda}\right)^2, \tag{2.7}$$

where λ is the signal wavelength. This is known as *free-space loss (FSL)*.

As in the case of Doppler shift, the signal attenuation due to FSL varies across the visibility window according to the satellite-terminal distance. If we express this distance as a function of ϵ,

the difference between the maximum and the minimum value of the free-space loss, ΔL_{FS}, can be computed using Eqs. (2.2) and (2.3):

$$\begin{aligned}\Delta L_{FS} &= 20\log_{10}\frac{D(\epsilon_{min})}{D(\epsilon_{max})} \\ &= 20\log_{10}\left[\frac{\cos\epsilon_{max}\sin\left(\cos^{-1}\left(\frac{R_E}{R_E+h}\cos\epsilon_{min}\right)-\epsilon_{min}\right)}{\cos\epsilon_{min}\sin\left(\cos^{-1}\left(\frac{R_E}{R_E+h}\cos\epsilon_{max}\right)-\epsilon_{max}\right)}\right]\end{aligned} \tag{2.8}$$

For our $h = 650$ [km] LEO reference system and a maximum elevation angle near 60°, the FSL variation is about 9 dB. Figure 2.3 depicts the evolution of both the FSL and the elevation angle as a function of time obtained through computer simulation. This relatively high FSL range suggests that some kind of mitigation technique should be considered. One possibility is to implement an adaptive power control system at the transmitter side. Another possibility to partially compensate for FSL variation is the use of antenna beamforming [War93]. This technique allows to generate radiation patterns such that the antenna gain is higher for low elevation angles. However, in general it is not possible to obtain ideal radiation patterns and therefore some important power variations will be still observed for certain angles.

2.3.4 Multipath fading and Shadowing

The presence of obstacles surrounding the signal propagation path combined with a non-null satellite–terminal relative velocity results in time-varying attenuation or *fading*. Two types of fading can be distinguished. *Multipath fading* occurs when some components of the signal are reflected and are observed at the receiver at different phases and times of arrival. As a consequence, the multiple components interfere each other and the resulting signal is distorted. *Shadowing*, on the other hand, occurs when the main signal component (*i.e.*, the LOS component) is directly obstructed by an obstacle.

The lower the elevation angle at which a signal is transmitted, the higher the probability for the signal to experience multipath fading and shadowing. This is an aspect to take into account since, as noted by [WS00], about 76% of the footprint area corresponds to elevation angles below 20° (for a 775-km LEO orbit). For the applications considered here—where transmitters are located in remote areas—it is reasonable to assume LOS conditions for $\epsilon \geq 20°$. In this case, the main variations of the signal strength are originated by time-varying FSL.

Some experimental campaigns devoted to study signal fading in the LMS channel have been performed. For instance, [VG88] presents measurement data on signal fading caused by multipath in mountainous terrain, concluding that tree shadowing has a more significant impact when compared to multipath fading.

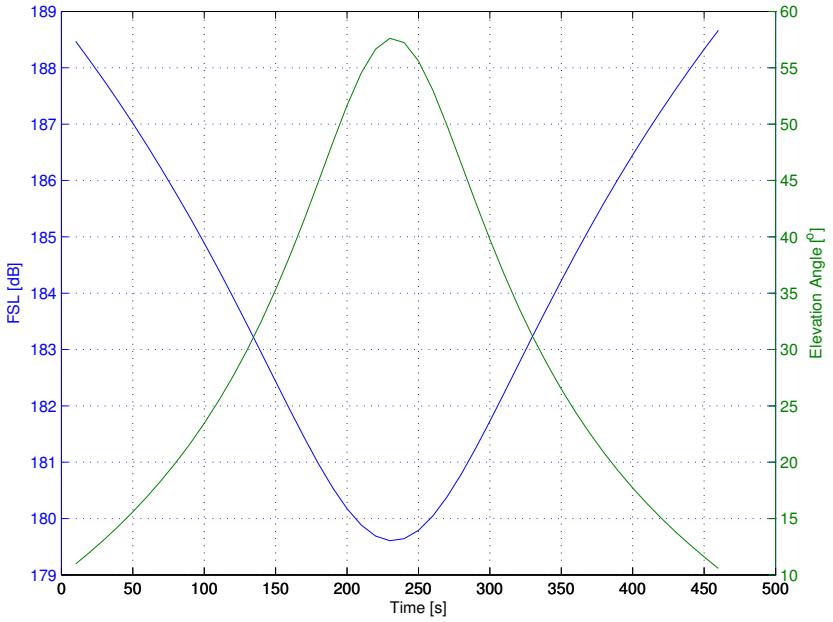


Figure 2.3 – FSL (V-shaped curve) and elevation angle (inverted V-shaped curve) variation for a typical LEO passage

2.3.5 Reference Models: The LMS Channel

The communication channel formed between a satellite and a mobile terminal has been generally referred to as the *land mobile satellite (LMS) channel*. The LEO satellite channel is therefore a special case. Several LMS modelling frameworks have been proposed in the literature (see for instance [CV94; Loo85; Fon+01; Lut+91]). In general, these models are built by first assuming some fade distribution function *a priori* (typically a combination of Rice and log-normal components) and then adjusting the set of parameters using empirical data. The data is usually obtained through experimental campaigns in which signals are transmitted from stratospheric balloons, helicopters or satellites and measured by a mobile terminal, in most cases using an antenna located on a car roof. NASA, the European Space Agency (ESA) and some universities have published several experiments of this kind (see [Sfo93; Gol97], for example). The measurements are typically classified according to the propagation environment (rural, suburban or urban), the elevation angle, frequency band and the mobile velocity.

It is in general difficult to get an accurate description of the LEO channel, specially for the uplink. On the one hand, further empirical data is required since most experimental campaigns only consider a few elevation angles and there is little information about the interference statistics. On the other hand, additional system-dependent parameters must be taken into account, such as

transmitter/receiver antenna radiation pattern. In [WS00], for instance, the LMS model proposed by Corazza [CV94] is used for the uplink of a data collection LEO satellite system. The model considers a predominant line-of-sight (LOS) component, which is assumed to be Rice distributed, but also log-normal shadowing on both LOS and multipath components of the signal. In addition, the model takes into account the satellite antenna radiation pattern. A more simpler approach has been followed in [DD12] to model the packet power distribution in a S-band M2M wideband satellite system using an opportunistic channel access approach. In particular, it is shown that the resulting *a posteriori* power levels follow a log-normal distribution[10].

2.4 Summary

This chapter introduced S&F LEO satellite networks and presented some of its applications and market trends. In particular, it was noted that the use of low-cost CubeSat platforms offers interesting advantages for the development of delay-tolerant, IoT/M2M-oriented data collection services and is now becoming a prospective application area.

The use of small, low-cost satellites and ground terminals, however, involves stringent constraints in the design of the communications system, such as a highly limited transmission and computing power and a low (or null) transmitter/receiver antenna gain. Similarly, the dynamics and the propagation characteristics of LEO systems bring additional issues, among which we described high Doppler effects, intermittent connectivity—which translates into high end-to-end latency—, shadowing and multipath fading, variable FSL and atmospheric propagation effects.

All these aspects need to be taken into account in order to design an appropriate communications system, which is the subject of the next chapter.

[10]The use of an opportunistic channel access approach refers to a transmission control scheme in which packets are transmitted only if the channel conditions are "good". As a result, the channel statistics, as observed by the receiver, are different. The modified channel statistics are referred to as the *a posteriori* distribution.

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

Claude Shannon (1916 – 2001)

3

Communications System Engineering

Contents

In this chapter we overview some general concepts involved in communications system engineering. Taking into account the overall system constraints described in the previous chapter, here we identify issues and requirements for efficient protocol design.

3.1 Introduction

In the previous chapter, we provided a high-level description of store-and-forward (S&F) data collection systems based on low-cost satellite platforms. We also mentioned some of the main constraints affecting the system, such as the limited power and limited processing capabilities in both satellite and terminals, and the channel impairments. The objective of this chapter is two-fold: first, to briefly review basic communications systems engineering concepts—in particular those related to wireless link design—and second, to discuss how the main system constraints affect the overall design.

A detailed communications system design goes beyond the scope of this chapter; we limit ourselves to provide a general conceptual overview on the subject. We begin with a discussion on network architecture in Section 3.2, where a minimalistic design is introduced. The need for cross-layer design is also discussed. Then, in Section 3.3, the link budget for both the uplink and downlink is analysed and its main design parameters are identified. As mentioned in the previous chapter, the uplink corresponds to a multi-user channel. The multi-user communications problem, as well as some of its more common solutions, is reviewed in Section 3.4. In particular, the need for random access communications in the IoT/M2M context is discussed. In Section 3.5 we study the overall capacity and show how it is impacted by the uplink design. Finally, Section 3.6 concludes this chapter.

3.2 Network Architecture

The set of protocols through which nodes of a network transmit and receive data is called the *network architecture*. In this section, the principles as well as the main networking functions that should be included in a satellite data collection architecture are identified and discussed.

3.2.1 Design Issues

In general, the network architecture depends on a number of aspects. From a high-level perspective, we must take into account application characteristics and requirements such as reliability, traffic patterns—including message arrival rate and message length statistics—and end-to-end delay tolerance, just to mention a few. IoT and machine-type satellite networks bring additional constraints. In particular, because of the stringent limitations in terms of energy and processing power for both terminals and satellite platforms, the design should be as compact and simple as possible. Further, the transmission of short packets stresses the need for a compact design since the additional protocol overhead is comparable with the payload size. As a consequence, the use of standard architectures such as the TCP/IP protocol stack is inefficient.

Among other network aspects that are relevant in our context and are directly related to protocol design we may mention security, energy efficiency and scalability (that is, the capacity of a network to support a growing number of nodes).

Thus, the major objective of network architecture design is to provide a virtual link through which nodes can communicate messages, and that guarantees (at least at some minimum degree) a set of properties as those mentioned above.

3.2.2 The Layered Approach

Traditionally, data communication networks have been designed following the so-called *layered architecture* [BG92]. The basic idea is to modularize the processes involved in data transmission in a hierarchical manner, so that, from a system-level perspective, each of these processes (or modules) can be abstracted as a black box that provides a given set of *services*. System integration is thus facilitated since different modules can be interconnected transparently, that is, without the need for implementation details of each individual module. Rather, a system designer only needs to be acquainted with the specifications regarding the layer interfaces. In order to guarantee these properties, a process from a given layer interacts either with processes within the same layer—eventually from different remote devices—or with processes at adjacent layers. The layered architecture is usually modelled according to the Open Systems Interconnection (OSI) model [KR12]. The main[1] layers defined by the OSI model are shown in Figure 3.1.

[1] In general, the Application layer is decomposed in three layers: Session, Presentation and Application.

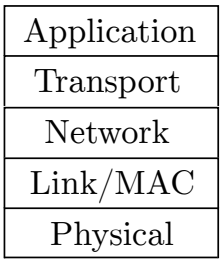


Figure 3.1 – The OSI layering model.

3.2.3 The Need for Cross-layer Design

A major problem with the layered architecture is that it is not optimal in disruptive environments such as wireless networks. Indeed, there is a trade-off in network protocol design regarding modularity and interoperability, on the one hand, and efficiency and performance, on the other hand. For instance, the layered architecture has pushed the development of the internet thanks to its capacity to integrate different networks and systems. However, the internet protocol architecture, specially its connection-oriented layer TCP/IP, performs poorly on channels with long round-trip delays, intermittent connectivity or time-varying conditions. As a consequence, cross-layer design (CLD) solutions have been introduced [SM05]. This approach breaks with the modularity principle of the layered architecture, since the overall architecture is designed by taking into account the interaction between the layers rather than by seeing each layer as an independent entity.

Although CLD allows to optimise the network performance in heterogeneous environments, it also involves some disadvantages. First, it is not evident to provide interconnectivity between traditional wired networks designed according to the layered architecture and wireless networks designed according to the CLD approach. Second, even though CLD might provide performance improvements, the interdependency among different layers also entails additional protocol complexity and signalling overhead. Finally, it has been argued that CLD may also lead to unintended cross-layer interactions and may also compromise the long-term evolution of wireless networks [KK05].

In the context of LEO satellite networks, a CLD approach is essential to exploit the channel efficiently due to the severe propagation environment and to the intermittent connectivity. Examples of CLD protocols are the DTN-oriented Bundle Protocol—for the transport layer—and CFDP, LTP and Saratoga—for the link layer (see also Section 2.1.3). A typical application wherein these protocols are employed is the transmission of raw image data from Earth observation LEO satellites In this case, the volume of data is important and the need for reliability is high. For data collection LEO satellite networks, these protocols are particularly suitable for the downlink. For the uplink, the adoption of a CLD approach is almost mandatory because—as we will see later on—the PHY design may also determine the design of the access layer.

3.2.4 A Generic Architecture

In the following we identify some of the most important tasks that should be considered when designing the network architecture.

As seen in Chapter 2, S&F LEO satellite networks were initially proposed to provide global communications services, in particular for users located in remote areas. The information flow in these networks was bidirectional and communication between two any arbitrary nodes was possible. There are two general approaches to achieve this. The first consists on using the satellites as an email-like server as proposed in the PACSAT protocol suite [WP91]. With this method, a list of all in-transit messages stored on board the satellite is constantly broadcasted through the downlink. Thus, nodes must check for new incoming messages at each satellite passage. Clearly, this solution is not optimal and shows scalability issues. The second approach consists on using an addressing mechanism. In this case, the satellite operates as a network switching node, *i.e.*, it stores and forwards the incoming messages according to their destination address. In order to do so, the satellite must keep a list wherein the physical address of each terminal node is mapped to some geographical region identifier. Note that this requires the satellite to be aware of its current geographical position, which can be easily achieved with a GPS receiver. Drawing an analogy with the standard TCP/IP protocol stack, the node physical address corresponds to the MAC address and the geographical region identifier corresponds to the subnet address. This addressing approach was proposed for the LEONET protocol [Che99].

We note that, since there is a single hop (*i.e.*, the satellite) to transmit a message between two nodes, the need for a *routing* protocol is avoided[2]. Thus, each message can be transmitted from end to end using only the addressing mechanism described above. According to the standard layered architecture, this can be considered a function of the link layer (in particular, the MAC sublayer).

An important function required in S&F LEO satellite networks is flow control. This can be a critical issue especially for the uplink channel, in that many nodes can transmit simultaneously to the satellite, which would result in channel overloading and packet errors. To avoid this issue, flow control signalling can be sent through the downlink broadcast channel. Note that this function operates at the link level: indeed, as the information volume transmitted by each individual node is typically very small, there is limited need for an end-to-end flow control mechanism.

The network architecture of a data collection satellite network may be simpler than that of a generic S&F satellite network. Indeed, as noted in the previous chapter, the information flow in a typical data collection application is unidirectional. Data is collected by sensor nodes, then transmitted to a satellite and finally collected by the gateways. At the gateways, an internet

[2]An exception is the case wherein several satellites can be visible simultaneously, or when a constellation of interconnected satellites is available. As seen in the previous chapter, this is not the case for typical low-cost, S&F LEO satellite networks.

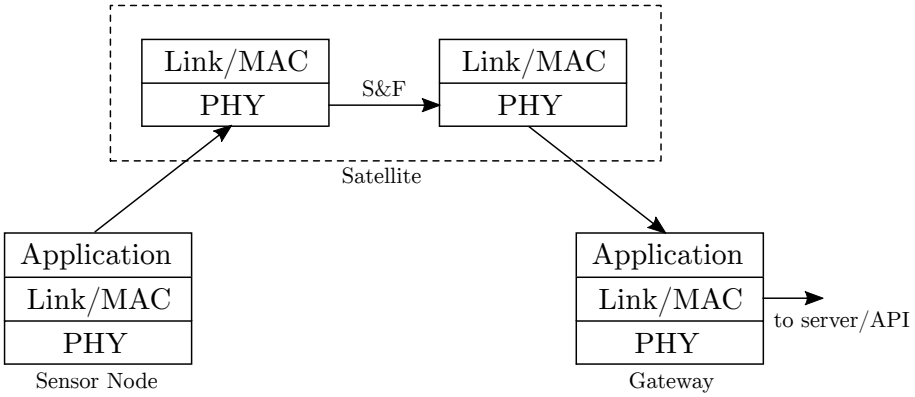


Figure 3.2 – Simplified network architecture.

interface (for instance, a web server or API[3]) can be implemented at the application layer to disseminate the collected data to the end users. In this case, the addressing mechanism can be further simplified since the gateways are the sole destination nodes.

Figure 3.2 depicts a simplified network architecture that implements the functions described above. The physical (PHY) layer performs standard functions: it maps sequences of bits coming from the link layer into electrical signals that are transmitted through the physical medium. In particular, it defines the modulation and coding schemes. Also, a preamble sequence is typically included for RF acquisition and channel estimation. At the link layer, application layer messages are formatted into *packets*. The source address (and eventually the destination address) is included in the link layer header. If an error detection scheme such as a cyclic redundancy check (CRC) code is considered, the check value is also added in the header. The MAC sublayer defines the mechanism by which the nodes transmit their packets through the uplink. This subject is discussed in Chapter 3.4. Note that, since routing is not required and flow control can be implemented at the link layer, there is no fundamental need for a network layer. Similarly, due to the nature of machine-type traffic, there is also no much need for a transport layer. Indeed, a transport layer makes more sense for relatively long communication sessions (*e.g.* video/voice communications, transmission of large files), which usually requires an additional end-to-end reliability layer, as well as end-to-end flow control. Finally, an application layer can be considered to perform tasks such as data compression and encryption.

[3]API stands for *application programming interface*.

3.3 Link Design

The objective of link design is to provide a communication channel that guarantees a given degree of reliability, which is frequently expressed as a maximum bit error rate (BER). Typically, the resulting *link budget* must be such that the expected BER should not exceed 10^{-5} or 10^{-6}. This minimum BER requirement is often considered as the *Quality of Service* (QoS) metric at the PHY layer. In practice, since the propagation conditions vary with time, the link QoS can only be guaranteed for some fraction of the time, so it is provided statistically. In order to simplify the analysis, in the following we assume that the link is only affected by free-space loss so that it can be characterised deterministically. Indeed, our objective is not to present a detailed link analysis, but rather to give a rough idea of the main issues that determine the link performance.

According to the modulation and coding schemes adopted at the physical (PHY) layer, the target BER translates into a minimum signal-to-noise (SNR) ratio per bit requirement, that is approximately

$$\frac{P_t G_t G_r}{L_{FS}(\epsilon) k T_s R} > \gamma_0, \tag{3.1}$$

where P_t is the transmitter power, $G_{t/r}$ the transmitter/receiver antenna gain, $L_{FS}(\epsilon)$ the free-space loss defined in Eq. (2.7), k the Boltzmann constant, T_s the system temperature and R the data rate. Here we are implicitly assuming that the filter bandwidth at the receiver's radio front-end is approximately equal to R, so that the total noise power is kT_sR. Also, since the path loss varies with ϵ, the link budget should be calculated considering a worst-case scenario, *i.e.*, $L_{FS}(\epsilon) = L_{FS}(\epsilon_{min})$.

With these considerations in mind, in the following we identify the main design parameters for both the uplink and downlink.

3.3.1 The Downlink

For the downlink, we observe that the satellite transmission power $P_t^{(\mathrm{sat})}$ as well as the antenna size (and consequently, its gain $G_t^{(\mathrm{sat})}$) are highly constrained. By contrast, it is physically possible to employ high-diameter dishes at the ground station. The main design parameters driving the downlink budget are given in the following relation:

$$\gamma_0^{(d)} \propto \frac{G_r^{(\mathrm{GW})}}{R^{(d)} L_{FS}(\epsilon_{min})}, \tag{3.2}$$

where $\gamma_0^{(d)}$ is the downlink SNR, $G_r^{(\mathrm{GW})}$ is the gateway antenna reception gain and $R^{(d)}$ is the downlink transmission rate. Recalling that $L_{FS}(\epsilon_{min}) = (4\pi D(\epsilon_{min})/\lambda)^2$ and that the gain of a parabolic dish is [Che89]

$$G = \frac{4\pi A_{eff}}{\lambda^2} e_A, \tag{3.3}$$

where A_{eff} is the antenna aperture or effective area and e_A the aperture efficiency, Eq. (3.2) becomes

$$\gamma_0^{(d)} \propto \left(\frac{1}{R^{(d)}}\right)\left(\frac{e_A A_{eff}}{4\pi D^2(\epsilon_{min})}\right). \tag{3.4}$$

Thus, in order to guarantee a minimum SNR, a system designer can

- adjust the data rate $R^{(d)}$;
- reduce the maximum communication range $D(\epsilon_{min})$;
- use a larger ground station antenna.

Note that, since our ultimate objective is to maximise the volume of data that can be collected by the ground station during the visibility window T_w, R should be as high as possible. This suggests to adopt frequency bands beyond the VHF/UHF bands because of the larger bandwidths available. On the other hand, if the communication range is reduced by increasing the minimum elevation angle, then T_w is also reduced, as seen in Section 2.2. An optimal implementation should instead adjust $R^{(d)}$ dynamically (*e.g.* using adaptive coding and modulation techniques).

3.3.2 The Uplink

For the uplink, the design approach is somewhat different: we are not interested in maximising the point-to-point capacity but rather the aggregate capacity achieved by all terminal nodes under the satellite footprint. In addition, it is desirable to minimise the transmission power in order to reduce the terminal size and increase its battery life span.

Since both the terminal and the satellite antennas cannot provide significant gain, the main design parameters driving the uplink budget are given by

$$\gamma_0^{(u)} \propto \frac{1}{R^{(u)} L_{FS}(\epsilon_{min})}, \tag{3.5}$$

where $R^{(u)}$ is the uplink data rate. In this case, we observe that the SNR is impacted by the operation frequency as the transmitter/receiver antennas do not provide significant directivity. As a consequence, the SNR can only be improved by:

- reducing the data rate;

- reducing the attenuation L_{FS} by either increasing the threshold ϵ_{min} or by adopting lower frequency bands.

The second approach requires the terminals to estimate ϵ as soon as a satellite is visible. This can be achieved by either estimating the Doppler shift first (as shown in [Ali+98]) or by simply measuring the signal strength of the satellite beacon. We will explore this idea in more detail in Chapter 5.

Considering the long communication range and the strict power limitations that characterise a satellite IoT/M2M system, we observe that in most application scenarios the achievable rate is small compared to the available channel bandwidth, *i.e.*:

$$R^{(u)} << W. \tag{3.6}$$

An example where the above relation appears is the ARGOS system, in which terminals transmit at data rates as low as 400 [b/s] and the total available bandwidth is 80 [kHz]. Broadly speaking, Eq. (3.6) characterises a type of wireless systems known as Low Power Wide Area Networks (LPWAN). In this scenario, there are two main PHY layer approaches that allow to close the link at an acceptable SNR: spread spectrum (SS) and narrowband/ultra narrowband (NB/UNB) techniques. While both SS and NB techniques have been subject to intensive research, the UNB approach has been less explored and provides a number of interesting advantages. We will see in Chapter 6 that the adoption of UNB as the PHY layer is typically accompanied by an asynchronous random access scheme. Similarly, if SS is adopted, the access scheme is typically based on CDMA (see Section 3.4.1). Thus, the PHY design has a direct impact on the election of the MAC scheme, which illustrates the need for a cross-layer perspective.

3.4 Multi-user Communications

A typical problem in communications engineering arises when two or more users share a single transmission channel. In this case, the channel is referred to as a *multiple access channel*. A *multiple access scheme* defines the way by which users share the channel. In the standard layered architecture, this logic is implemented at the media access control (MAC) sublayer and is considered as part of the link layer. The election of the MAC scheme is influenced by several aspects: traffic patterns, message length, delay constraints or network size. Furthermore, its performance can even be impacted (or improved) by the underlying PHY layer, and may show important variations in different transmission environments. To characterise the MAC performance in terms of channel utilisation, some authors resort to the *MAC efficiency* $\eta^{(\mathrm{MAC})}$ [MB09], defined as

$$\eta^{(\mathrm{MAC})} = C^{(\mathrm{MAC})}/C, \tag{3.7}$$

where $C^{(\mathrm{MAC})}$ is the overall channel throughput (considering all users) and C is the rate that can be achieved by a single user exploiting all the channel bandwidth[4].

3.4.1 Multiple Access Dimensions

There are three main dimensions in which a communication channel can be decomposed in order to serve multiple users: time, frequency and code. The most simple access schemes associated to each dimension are time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA). To examine them, we assume in the following that a single channel of bandwidth W must be shared by M users. Also, we assume that the user's transmitters are power limited to P. We denote by R the maximum rate at which a single user can transmit reliably using the whole bandwidth W and the maximum power P.

3.4.1.1 TDMA

In TDMA the channel is divided in time *slots* of equal length T_{slot}. Each user is allocated one out of M slots that form a TDMA *frame*. At the beginning of each slot, a user transmits a *burst* using all the available bandwidth W, and then releases the channel to the next user. In order for the receiver to be able to detect each burst, a guard interval and a synchronization sequence is added at the beginning and eventually at the end of each burst. Note that, as $M \to \infty$, $R \to 0$ and the delay experienced by each user between two consecutive transmissions tends to infinity. A disadvantage of TDMA in the LEO environment is that each terminal must accurately estimate the propagation delay so that its packets are received within the slot boundaries. Note that the propagation delay may vary significantly from terminal to terminal. An alternative is to oversize the slot length adding a guard interval equal to the maximum propagation delay, though this involves degradation of the MAC efficiency.

3.4.1.2 FDMA

In FDMA, the channel is divided into M non-overlapping sub-channels of bandwidth $B = W/M$. In contrast to TDMA, with FDMA all M users can transmit simultaneously and, therefore, continuously. In practice, to guarantee that the sub-channels do not interfere with each other (for instance, due to carrier frequency instabilities or filter leakage), a guard band must be included between adjacent sub-channels. As in TDMA, if we let $M \to \infty$—and assuming a limited spec-

[4]Note that this definition suggests implicitly that $C^{(\mathrm{MAC})} < C$ so that $0 < \eta^{(\mathrm{MAC})} < 1$. However, we know from information theory (see for instance [CT06]) that the MAC capacity is in fact greater than the single user capacity. In despite of this, the hypothesis that $C^{(\mathrm{MAC})} < C$ is valid for systems employing standard single-user detection.

tral efficiency[5]—then $R \to 0$ since each user is allocated a narrower sub-channel and the total available bandwidth is limited. Furthermore, the MAC efficiency of FDMA is severely affected since it becomes more difficult to isolate each sub-channel from adjacent-channel interference. A problem with FDMA in the LEO environment is that the Doppler effect must be compensated at the transmitter side, which adds some complexity in the design. This is an issue to take into account specially at high frequency bands or when the number of FDMA channels is high. For low Doppler shifts, the terminal design can be simplified by simply increasing the guard bands, which comes at the price of reduced MAC efficiency.

3.4.1.3 CDMA

A more elaborated multiple access method is CDMA, in which all users transmit continuously using the full available bandwidth W. In order for the receiver to be able to distinguish among different transmissions, the users transmit using *spreading* codes. Hence, CDMA is a spread spectrum technique, *i.e.* a technique in which the signal bandwidth is much larger than the information rate. If the spreading codes are orthogonal (that is, perfectly separable), then the M users can in theory transmit reliably. However, when the network size increases, it is difficult to guarantee orthogonality and the system performance is degraded. In general, two CDMA approaches can be adopted: in Direct-Sequence CDMA (DS-CDMA) the spreading code take the form of a binary pseudo-random sequence with which each transmitted symbol from a given user is modulated. The receiver then correlates the resulting overlapped signals with each sequence in order to separate the transmissions from the different users. An important disadvantage of DS-CDMA is that, if the signals from different users experience significantly different channel gains (resulting in some signals being much stronger than others), it is likely that only the stronger signals will be decoded—the weaker signals being perceived as background noise. This is know as the *near-far* effect. Since the LEO environment is characterised by varying free space attenuation loss, the use of DS-CDMA requires some power control scheme in order to avoid significant degradation in the MAC efficiency.

The second approach is called Frequency Hopping CDMA. Here, each user's data is transmitted at a lower rate following a predefined sequence of frequencies (hops). While this method allows to reduce the near-far effect, it involves some complexity issues since a dynamic, fast frequency synthesiser is required. Note that, as for FDMA, the Doppler effect may also impact the MAC efficiency since the hopping sequences from different users may overlap due to the frequency shift.

3.4.1.4 Other Dimensions and Hybrid Schemes

Among other multiple access methods recently proposed we may mention space division multiple access (SDMA), in which the users in a wireless network can be separated by advanced *beamforming*

[5]By a limited spectral efficiency we mean that the amount of information bits per transmitted symbol cannot be increased indefinitely as the signal-to-noise ratio $P/(N_0 B) \to \infty$.

techniques, *i.e.*, by adapting the receiver antennas' radiation patterns according to the user's locations. SDMA can be used in combination with other techniques such as TDMA, FDMA and CDMA.

In fact, most modern multiple access schemes actually combine some or all existing channel dimensions. Perhaps the most common approach, in particular in wireless networks, is to coordinate the transmissions following multi-frequency TDMA (MF-TDMA) as the basic scheme. In this way, several parallel TDMA frames can operate at different carrier frequencies, allowing to increase the network capacity without requiring the user's terminals to increase the transmission power. A different class of multiple access schemes based on advanced signal processing has also been introduced during the last decades. A well-known example is orthogonal frequency division multiple access (OFDMA), in which the transmitted signal is decomposed in several orthogonal subcarriers. This technique allows for a more flexible allocation of the transmission resources and is more resilient to multipath fading, which is why it is widely employed in current wireless communications standards.

3.4.2 Connection-oriented Multiple Access

In connection-oriented multiple access, a user is assigned a dedicated virtual channel, or circuit, until the communication session is ended. This virtual channel can be in the form of a time slot (TDMA), a carrier frequency (FDMA), a spreading code (CDMA) or a combination of these if a more sophisticated multiple access method is adopted. This approach is particularly suitable for applications with strict latency requirements (*e.g.* videoconferencing), but is highly inefficient for *bursty* traffic (that is, when the peak-to-average data rate is high). Consider for instance the transmission of a telephone call over a dedicated circuit. Since the circuit is reserved during the complete session, we will observe that a significant fraction of the channel resources will not be used during the silence periods.

In general, a connection can be established in two different ways. With a *fixed assignment* policy, every user terminal has a dedicated circuit reserved *a priori* and statically. This means that the connection is established permanently, even if the user has no data to transmit. Hence, it is a recommended strategy only when the traffic requirements are roughly constant or predictable. An advantage of this approach is its simplicity and that it involves no connection set-up overhead.

The second approach is *on-demand assignment*, in which the channel resources are allocated dynamically. To establish a connection, a user must first send a request through an independent sub-channel reserved to that end. Note that the access to this channel must be performed in a contention-oriented fashion (see Section 3.4.3 below). Clearly, on-demand assignment is in general much more efficient than fixed assignment, specially for large networks with unpredictable traffic patterns and varying message lengths. On-demand assignment is also referred to as demand assignment multiple access (DAMA).

3.4.3 Contention-based Multiple Access

In many applications, the amount of data that each node needs to transmit is so low that it can be comparable with the amount of signalling overhead required to set up a connection. In this case, a simpler and perhaps more efficient approach would be to accommodate the data in a single packet and send the packet through the channel directly without prior coordination. If the load on the channel is not too high, then it is likely that the packet will be transmitted successfully. Conversely, if there is significant traffic on the channel and other packets are transmitted simultaneously, then a collision occurs and, eventually, some or all packets will be corrupted and lost, resulting in reduced MAC efficiency. Even though transmitting packets in an uncoordinated fashion increases the risk of packet loss, it also helps to reduce the end-to-end delay, since the connection set-up stage is avoided. This is particularly appealing in networks with long round-trip delays such as satellite networks. This approach is known as *contention-based multiple access*, or simply as *random access*, and was popularized in packet-based radio networks starting form the early 1970s.

3.4.4 The Need for Random Access Communications in the Satellite and the IoT/M2M Context

Random access protocols have always played a fundamental role in satellite networks. Since the coverage area of a single satellite is much larger than that of a typical terrestrial network, the number of potential nodes that can be served is also much more important. As a consequence, adopting a fixed-assignment multiple access scheme would be highly inefficient. For instance, as we pointed out above, with static TDMA the delay increases linearly with the network size and a large fraction of the channel capacity would be wasted due to idle nodes. In addition, today's data communication networks are characterised by heterogeneous, highly-unpredictable traffic patterns, which is why more flexible access schemes such as DAMA are implemented in general. Note that with DAMA a random access channel is always required.

Satellite networks, in particular GEO-based broadcasting systems providing television and internet access, also employ random access in the *return channel*, that is, the channel that transports data from the user terminal to the gateway. This is because the traffic on this channel is characterised by short, bursty packets (*e.g.* a simple request for some specific content such as a website), and because at the same time a low end-to-end delay is desired.

IoT and M2M networks stress the need for random access even further. First, the number of nodes is significantly larger. Second, a typical machine-type message consists only of a few bytes of information, which makes any communication overhead more significant. Finally, some terminals operate asynchronously and do not include receiving capabilities (as a way to further reduce energy consumption and cost), meaning that dynamic allocation methods are not feasible. In this case it is also highly desirable to minimise the packet error probability because a retransmission scheme

cannot be implemented. Clearly, a low packet error probability is in general desirable since it limits the number of retransmissions required to transmit packets reliably, increasing both the MAC efficiency and the transmission energy efficiency. A more complete review of random access concepts and protocols is provided in the next chapter.

3.4.5 Some MAC Protocols Employed in S&F LEO Satellite Networks

Several MAC protocols specially designed for S&F LEO satellite networks have been proposed in the literature. In [HMR95], the authors propose a design based on the slotted ALOHA protocol[6], with binary exponential backoff (BEP) as the collision resolution algorithm. The slot length is designed in order to accommodate the maximum propagation delay, which involves some performance degradation as discussed in Section 3.4.1.1. The authors also propose a *stop-and-wait* automatic repeat request scheme. With this approach, after a terminal transmits a packet it must wait for an acknowledgement (ACK) before attempting further transmissions. If a negative ACK is received, the current terminal transmission probability is divided by 2 in accordance to the BEP algorithm. The ALOHA and the slotted ALOHA protocols have also been adopted in [AKZ07; Bed13]. A variation is proposed in [Had98], wherein selective reject (SREJ) ALOHA is adopted[7]. Compared to the slotted ALOHA with BEP adopted in [HMR95], the retransmission policy in SREJ ALOHA can be more efficient since a smaller amount of data is retransmitted in general. A more flexible, frame-based design is described in [WS00]. The frame is divided into two sections: the first consist of N slots reserved for slotted-ALOHA-like short packet transmission, while the second is used for long message transmission, using a dynamic connection-oriented approach. An advantage of the framed approach is that ACKs are sent on a frame-by-frame basis (that is, one ACK list at the end of the frame) instead of on a packet-by-packet basis. Note that all the designs cited above are fully or partially based on random access multiple access. Further, slotted and non-slotted forms of the simple ALOHA protocol are employed. In the next chapter we will see that these protocols are rather suboptimal and that high-throughput alternatives exist.

3.5 On Data Collection Network Capacity

We define the *data collection network capacity*, denoted by V, as the total volume of data that can be collected by the system over a given time span ΔT. In the following, we consider a simplified network topology consisting of M source nodes, a single relay node s (the satellite) and one sink node (the gateway), as shown in Figure 3.3. The variables c_u and c_d denote the uplink and downlink capacities, respectively. By *capacity* we mean the rate that can be achieved by a

[6]This protocol is reviewed in the next chapter.

[7]This protocol is also reviewed in the next chapter.

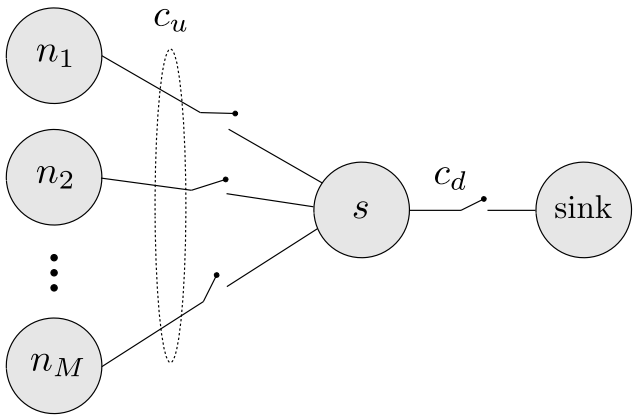


Figure 3.3 – Network model.

single user exploiting the whole available bandwidth, considering the link budget and practical modulation and coding schemes. We assume that the satellite can be either in data collection mode or in forwarding mode, so only one link is active at a time. The fraction of time over which each link is active, *i.e.*, the link duty cycle, is denoted by τ_u for the uplink and τ_d for the downlink, with $\tau_u + \tau_d \leq 1$. Clearly, the information volume that can be collected by the system during ΔT is upper bounded by

$$V = \Delta T \min\{c_u\tau_u, c_d\tau_d\}. \tag{3.8}$$

In other words, the network capacity is determined by the bottleneck capacity, which can either correspond to the uplink or to the downlink, according to the actual system design. If we denote by $T_{u/d}$ the effective uplink/downlink throughput at the link layer, we may write $T_{u/d} = \eta_{u/d}^{(\mathrm{MAC})} c_{u/d}$, where $\eta_{u/d}^{(\mathrm{MAC})}$ is the uplink/downlink link layer efficiency. Thus, Eq. (3.8) becomes $V = \Delta T \min\{T_u\tau_u, T_d\tau_d\}$.

Since the downlink is a point-to-point link, packet errors are mainly due to transmission errors (*e.g.*, after a signal fade). Also, as mentioned above, high-efficiency link layer protocols can be employed for a maximum use of the channel bandwidth. Therefore, we may assume $\eta_d^{(\mathrm{MAC})} \approx 1$. For the uplink, if we consider a contention-based multiple access scheme, then in addition to transmission errors there is also a non-negligible probability of packet losses due to collisions. In this case, a significant fraction of the capacity is typically wasted, and hence $\eta_d^{(\mathrm{MAC})} < 1$.

A numerical example is given in Table 3.1. We consider that, over a time span $\Delta T = 24$ [h], the data collection payload is active during 23.5 [h], while the downlink is active during 1/2 [h][8].

[8]For the downlink, this is roughly the total daily access time for a single ground station. We assume that the rest of the time the satellite is in data collection mode.

Uplink	**Downlink**
$\tau_u = 23.5/24$	$\tau_d = 1/48$
$c_u = 9600$ [bps]	$c_d = 100$ [kbps]
$\eta_u^{(\mathrm{MAC})} = 0.1$ [bps]	$\eta_d^{(\mathrm{MAC})} \approx 1$
$T_u = 960$ [bps]	$T_d = 100$ [kbps]
$\Delta T \tau_u T_u \simeq 81$ [Mb]	$\Delta T \tau_d T_d = 180$ [Mb]

Table 3.1 – A numerical example of the daily data collection capacity bounds for a practical system.

The link capacities have been selected taking into account the limitations of a typical nanosatellite platform. For the downlink capacity we adopted a conservative value of $c_d = 100$ [kbps], though current radios provide rates beyond 1 [Mbps]. For the uplink we assumed $c_u = 9600$ [bps]. This rate can be achieved with a SNR per bit $E_b/N_0 > 20$ [dB] by a terminal transmitting at 1 [W] to a satellite in a 600 [km] LEO orbit at the minimum elevation angle $\epsilon_{min} = 10°$. However, some designs adopt lower rates (*e.g.*, 1200 [bps] in [AKZ07; Bed13]) The uplink access scheme is assumed to be based on the slotted ALOHA random access protocol, which provides an efficiency $\eta_u^{(\mathrm{MAC})} \simeq 0.1$ for a packet error rate equal to 0.1 (see next chapter). The results of this example show that the data collection capacity is limited by the uplink, even if it is active during a much higher fraction of time. Further, the volume that can be downloaded is more than twice the volume that can be uploaded. It should be clear that the uplink-downlink capacity asymmetry can be even higher in practical systems. This simple example illustrates two important things: first, it shows the impact of the MAC efficiency on the overall system capacity. Indeed, if we improved the uplink efficiency by a factor 2.2, the downlink capacity would have been reached. Second, it highlights the need for an accurate evaluation of $\eta^{(\mathrm{MAC})}$.

3.6 Summary

In this chapter, some fundamental topics on the design of the overall communications system have been reviewed. The need for an optimised, cross-layer network architecture is discussed, and a generic architecture is introduced. The uplink and downlink budgets are examined. For the downlink, we showed that several alternatives for capacity improvement exist. By contrast, we have seen that the uplink is subject to more severe limitations, and that only low information rates can be adopted in general.

The adoption of an appropriate multiple access scheme for the uplink channel is discussed. Due to the traffic properties of typical IoT/M2M networks and to the nature of the LEO environment,

we conclude that a suitable PHY/MAC protocol design should meet the following requirements:

- random operation;
- support for short packet transmission;
- scalability;
- delay tolerance;
- energy efficiency;
- high MAC efficiency and low packet error rate;
- support for varying propagation delay, power and high Doppler shifts;
- long communication range;
- efficient use of the channel bandwidth W for $R << W$, where R is the single-user rate.

Thus, these stringent constraints suggest to follow a cross-layer PHY/MAC design. Our focus in the remaining chapters will be mainly on the design and evaluation of novel random access schemes aiming to respond to these requirements.

"Every time you power up your mobile phone or use that phone to establish a voice, SMS, or Internet connection, the very first packet transmitted is sent via an ALOHA random access channel."

Norman Abramson (1932 – present)

4

Random Access Communications

Contents

It becomes clear from the previous discussions that random access protocols are fundamental in both satellite and IoT/M2M networks. Consequently, this chapter introduces basic concepts and as well as modelling conventions employed in the design of random access systems. In addition, a short review of the state of the art on random access protocols is provided.

4.1 Introduction

Since the introduction of the ALOHA protocol by Abramson in the early 1970s [Abr70], Random Access (RA) schemes have played a fundamental role in communications networks, particularly in wireless radio. They are currently an essential part of cellular and satellite systems, where they are used for initial network connection, capacity allocation demand or even as the main multiple access scheme. Moreover, the adoption of highly efficient RA protocols will be paramount for the development of massive M2M communications and the IoT [Zan+13] in the near future.

With ALOHA, Abramson proposed a simple access solution in which the users transmit over a broadcast channel as soon as they have a packet in their buffers, without following any type of coordination. In case of *collision*, that is, when two or more packets interfere each other, Abramson assumed that all packets would be automatically lost. With this hypothesis, known as the collision channel model, he showed that a maximum channel utilisation (or *throughput*) equal to $1/(2e)$ was achievable. Soon after the introduction of the ALOHA system Roberts showed [Rob75] that if the packet transmissions are accommodated into slots of length equivalent to the packet duration, then a maximum throughput of $1/e \simeq 0.37$ [packets/slot] can be achieved. Thus, ALOHA and the slotted ALOHA, both of which will be reviewed later on in this chapter, have become the basis for the consolidation of RA as an active research field within the context of information and communications theory

As pointed out in 3.4.3, RA schemes are suitable for large networks with *bursty* traffic, specially for short message transmission and when the number of nodes in the network is high. In this

context, if conventional fixed-assignment TDMA is adopted then the access delay would increase linearly with the network size. On the other hand, adopting a channel reservation scheme might be too inefficient due to the connection set-up overhead. Thus, RA schemes are usually the most simple and efficient solution in such environments.

While it is true that RA allows to keep a low access delay, this usually comes at the cost of higher packet error rates and, consequently, more energy consumption. Further, without some kind of congestion control mechanism, the packet error rate can quickly deteriorate up to a point of no return in which the system capacity goes to zero [EV89]. This is the well-known stability problem of RA systems and will be briefly discussed in Section 4.2.

In this thesis the focus will be mainly on the throughput and packet error rate performance metrics, which will be studied later on within the context of time- and frequency-asynchronous random access. We start by providing some basic definitions in Section 4.2. Then, in Section 4.3 we describe some of the most commonly used reception or channel models. In the two following sections, an overview of the state-of-the-art on random access protocols is provided. Finally, Section 4.7 summarizes this chapter.

4.2 Basic Definitions and Figures of Merit

In this section we review some basic modelling notions and performance metrics that are usually employed to characterise RA systems, following classical conventions [Abr70; BG92]. Let us start by considering a generic system in which multiple users send packets to a single receiver node or sink. We assume that the aggregate traffic from all users has mean λ [packets/s]. If the packet length is fixed to T_p [s], then the average number of packet arrivals during this interval is $G = \lambda T_p$, which is expressed in [packets/packet duration] or erlangs [E] and is usually referred to as the *normalized channel load* or simply the *MAC load*. In slotted RA—where the transmissions are accommodated into time slots of length equal to the packet duration—G is usually expressed in [packets/slot]. The aggregate traffic on the network is usually modelled as a Poisson source. In this case, if K denotes the number of packet arrivals during during T_p [s], the probability of having $K = k$ is

$$P(K = k) = \frac{G^k}{k!} \exp[-G]. \tag{4.1}$$

We will resort to this formula several times throughout this thesis.

4.2.1 Throughput and Packet Error Rate

The system *throughput*, denoted by T, is defined as the average amount of packets that are successfully transmitted through the MAC channel during a given time unit. As G, it is usually

convenient to normalise T so that it is expressed in [packets/packet duration]. Also, when physical layer parameters are to be taken into account, it is usually convenient to express the throughput in [b/s/Hz] according to the underlying spectral efficiency.

Mathematically, we define the normalised throughput T as

$$T(G) \triangleq G \times [1 - P_e(G)] \tag{4.2}$$

where $P_e(G)$ is the packet error rate (PER). According to the underlying reception model (see next section), P_e may or may not include transmission errors. In most cases, it only considers packet losses due to collisions and not to other signal impairments caused by the propagation channel.

From Eq. (4.2), the PER is simply given by

$$P_e(G) = 1 - T(G)/G. \tag{4.3}$$

Relation with the MAC efficiency

In Section 3.4 we introduced the notion of MAC efficiency $\eta^{(\mathrm{MAC})}$ in order to describe the channel utilization of a generic MAC protocol with respect to a single-user (or point-to-point) system. This definition is equivalent to the throughput of a random access system, as long as the receiver follows a conventional single-user architecture. In this case, a maximum of 1 [packets/packet duration] can be processed by the receiver and $0 \leq T \leq 1$.

4.2.2 Stability and Delay

In practice, packets generated by higher layers arriving at the MAC unit are stored in buffers of finite length until they are transmitted on the channel. When the MAC layer must provide full reliability, packets are buffered until they are *successfully* transmitted, which is acknowledged by the central station through a feedback channel. In this case, some kind of retransmission policy has to be implemented at the MAC unit, and the traffic on the channel can be decomposed into newly generated packets arriving to the system and retransmitted packets.

The problem of ensuring system *stability* translates into selecting the appropriate transmission control mechanism that ensures that the probability of buffer overflow can be made arbitrary small, or equivalently, that the underlying Markov chain process is ergodic [RE88; EV89]. For this, the new-packet arrival rate of each user must be controlled in order to stay below a certain threshold that ensures stability.

Another measure that is directly related to the notion of stability is the packet *delay*. Indeed, in a stable system the delay experienced by any buffered packet must be bounded. Thus, if the

moments of the delay are finite for a given new-packet arrival rate then it is likely that the system will be stable.

The stability problem is in general complex and usually requires a different modelling approach. For slotted ALOHA, it has been partially tackled assuming a finite population network in which the arrival process is Bernoulli. Indeed, the *stability region*[1] of the slotted ALOHA has only been precisely determined for the two-user system [TM79] while for the general N-user system only upper and lower bounds have been found [RE88].

In general, practical implementations of RA systems are particularly susceptible to stability issues. In slotted ALOHA, *e.g.*, it is possible to show through an intuitive analysis that even if the new-packet arrival rate is below the maximum throughput, a short traffic fluctuation can lead the system to an unrecoverable operation state [Mas88]. For these reasons, practical systems are usually designed to support relatively low traffic loads (for the classic slotted ALOHA, it is roughly below 0.1 [packets/slot]). Alternatively, stable operation can be reinforced by means of different approaches, which include the use of collision resolution and congestion-avoidance policies or simply by allowing the system to be rebooted. These solutions, unfortunately, incur the cost of a more complex MAC design.

4.2.3 Normalized Efficiency

As an alternative measure of the throughput efficiency for average-power-limited systems, the *normalized efficiency* has been proposed [Abr77; Liv11]. It is defined as:

$$\eta^{(\mathrm{RA})} = C_i/C, \tag{4.4}$$

where C is the Gaussian multiple access channel capacity and C_i is the capacity of a given RA scheme i. If P denotes the aggregate power on the channel from all active users and N is the noise power, C is given by

$$C = \log(1 + P/N), \tag{4.5}$$

while C_i is defined as

$$C_i = T_i(G)\log(1 + P/(DN)). \tag{4.6}$$

In Equation (4.6), $T_i(G)$ is the RA system throughput and D is the ratio between the average power transmitted on the channel and the power used for transmitting a single packet.

[1] This corresponds to the set of new-packet arrival rates that ensure system stability.

4.2.4 Asynchronous vs Synchronous Random Access

In general, protocols based on pure ALOHA are referred to as *asynchronous* RA systems (AS). These correspond to the simplest form of channel access because the transmitters do not need to be synchronized with a network-wide timing reference. On the other hand, if packet transmissions are accommodated into slots, as proposed in slotted or framed ALOHA systems we speak of *synchronous* RA systems (SS).

Currently, most modern wireless systems architectures are based on some type of Multi-Frequency Time-Division Multiple Access (MF-TDMA) where the time-frequency space is divided in frames where different services and dedicated channels are multiplexed. This is the case for widespread protocols such as IEEE 802.11, 4th generation Long Term Evolution (LTE) and Digital Video Broadcasting (DVB). In these systems RA channels are typically slotted. Besides, before the apparition of high-throughput AS, SS have been believed to provide better performance under practical implementations. As a consequence, a significant amount of research has been devoted to enhance the performance of SS.

As already noted, the main drawback of SS is the need for slot timing synchronisation. This, in general, can be achieved by two different ways: the first is to rely on external reference signals such as, for instance, those provided by the Global Positioning System (GPS). The disadvantages with this approach is that additional circuitry is required and that external reference signals are not always available. The second alternative is to implement a dedicated broadcast signalling channel in the central node. In this case, the disadvantage is that some of the downlink network capacity is lost because of the signalling overhead. Either way, the requirement for slot synchronisation has an impact in the transmitter circuitry complexity (and consequently, in its cost) as well as in the network power consumption. Indeed, because of clock drift, tighter slot boundaries result in more expensive oscillators and/or more frequent signalling. These aspects, in particular, are discussed in detail in [Sad05]. Further, in wide-area wireless networks (as for instance, mobile satellite networks or LPWANs) some mechanism to compensate the highly variable propagation delays must be implemented. Otherwise, the slot guard times must be increased, resulting in throughput performance degradation (see [Ren01, Section 3.3.2]).

For these reasons, and since in both terrestrial and satellite IoT/M2M networks energy efficiency and terminal complexity are important aspects to take into account, AS are currently gaining increasing interest.

4.3 Reception Models for Random Access Systems

As discussed in Section 3.2.3, a natural consequence of the layered approach is that protocol design is done without much regard to adjacent layers, and MAC protocols are not an exception. In the context of RA systems, instead of dealing with complex PHY layer models, it was initially more

convenient to work upon simplistic, abstract models that allow for more tractable analysis. The collision channel proposed by Abramson for ALOHA [Abr70] is a good example of this approach. Along with the collision model, a variety of different reception models have been proposed in the literature. This need for new models was on one hand due to the inherent limitations of the collision model, which fails to give accurate performance bounds. On the other hand, it was due to the appearance of modern receiver architectures such as Multipacket Reception (MPR), Multiuser Detection (MUD), SIC-based systems and, in general, to the emergence of cross-layer designs [NV04; DS04].

Indeed, assuming that all packets involved in a collision are automatically discarded may represent a too pessimistic hypothesis, specially for packet radio systems where the received power levels are affected by time-varying fading and path loss. In reality, if one of the packets involved in a collision is received with sufficiently high power with respect to the other contending packets, then the receiver will likely be able to decode it, and the other packets would be perceived simply as background noise. When the latter occurs, we say that the receiver has *captured* the packet. The phenomenon is known as the *capture effect* and was first observed in FM receivers [Rob75]. Similarly, it has been found that through advanced signal processing techniques it is possible to achieve extremely low packet error rates, and thus a more detailed description of the PHY layer was required in order to study the system performance.

Thus, we may identify two main families of reception models that are relevant within the context of RA systems: 1) those that do not assume any particular PHY layer implementation, which we refer to as *PHY-independent models* and 2) those that do take into account PHY layer parameters such as the modulation and coding schemes, which we call *PHY-dependent models*. Some well-known models in these two categories are described in the following. Note that most of the references cited in this section are based on slotted RA systems, which have been more popular and also more easy to analyse.

4.3.1 PHY-independent Models

The most basic example within this category is naturally the collision channel. Additionally, there are two other models that have been widely employed in the literature.

- *The protocol model.* First proposed in [GK00] for studying the capacity of wireless networks, this model is based on a geometrical abstraction of the underlying PHY layer. Assuming a star topology, the protocol model can be defined as follows. Consider a node i located at a distance R_i from the central receiver. Then a packet from i is successfully decoded if

$$R_j \geq (1 + \Delta)R_i, \tag{4.7}$$

for any node $j \neq i$ that is transmitting simultaneously with node i. The term Δ is a parameter that can be adjusted according to some specific system characteristics (*e.g.* the propagation environment or the receiver sensibility). Due to its simplicity, this model has been mainly used to derive analytical bounds on the scalability of Wireless Sensor Networks.

- *The SINR-based capture model.* A variety of models in this category have been proposed in the literature (see for instance [Rob75; Abr70; ZP92; Nam84]). In general, a packet from user i is assumed to be captured by the receiver if

$$\gamma = \frac{P_i T}{\sum_{j \neq i} P_j T + N_0} > b \tag{4.8}$$

where γ is called the signal-to-interference-plus-noise ratio (SINR), P_j is the received power of node j, T is the symbol period and N_0 is thermal noise. The constant b is known as the *capture ratio* and depends on the receiver implementation. In general b can be any positive constant, though it can be negative in spread-spectrum systems. The term $\sum_{j \neq i} P_j T$ corresponds to Multiple Access Interference (MAI).

The capture effect can be evaluated by means of the capture probability $P_C(k)$, which is expressed as a function of the number of colliding packets k, but also depends on their power distribution at the receiver input. Thus, if K is a random variable representing the number of packet arrivals, the throughput can be expressed as

$$T(G) = \sum_{k=0}^{\infty} P(K = k+1; G) P_C(k) \tag{4.9}$$

It can be shown that, the more variable the received power is, the higher the maximum throughput. While power imbalance is already an inherent characteristic of packet radio networks, some authors have also proposed to enhance the capture effect by forcing the terminals to transmit at random power levels [Met76; BH]. However, this is achieved at high packet error rates and thus is not a recommended approach when energy efficiency is a concern. Furthermore, it has been shown that, if we want to optimise both the throughput and energy efficiency of an ALOHA system, then the single power level scheme is optimal [WE02]. When the signal power is mainly affected by the propagation environment, then the system performance depends directly on the channel statistics and the spatial distribution of the users. The interested reader will find abundant documentation on these topics (*e.g.* [Abr77; Nam84; AB87; SYW90; LL92; LHV92]).

4.3.2 PHY-dependent Models

While SINR-based capture models already provide a better description of signal propagation aspects, they are particularly less accurate for asynchronous RA schemes[2]. In this case, the MAI shows more rapid variations and the average packet SINR becomes a less representative figure of merit. A more accurate evaluation of the capture effect can be achieved if we take into account PHY layer parameters. In order to do so, we start by decomposing the PER as follows [Abr90]

$$\begin{aligned} P_e = & P(\text{decoding error}|\text{preamble detection \& synch}) \\ & \times P(\text{preamble detection \& synch}) \end{aligned} \tag{4.10}$$

where $P(\text{preamble detection \& synch})$ is the probability that the packet preamble is successfully detected and that the signal parameters are correctly estimated; and $P(\text{decoding error}|\text{preamble detection \& synch})$ is the probability that at least one bit is found in error after demodulation and error correction (if applied), given that the former event occurs. Notice that a high preamble detection probability also entails a high probability of successfully estimating the signal parameters since this process is typically preamble-aided. Hence, we may assume

$$P(\text{preamble detection \& synch}) \simeq 1 - P(\text{preamble collision}) \tag{4.11}$$

In general, Eq. (4.10) can be simplified if the system is designed in order to achieve a probability of preamble collision that is at least one order of magnitude below $P(\text{decoding error})$, in which case $P_e \simeq P(\text{decoding error})$. This can be achieved, for instance, by the use of an appropriate set of direct-sequence preambles with low cross-correlation properties.

In addition, it is possible to use direct-sequence preambles with which the preamble collision probability can be made significantly smaller. In any case, it is only important to achieve a low preamble collision probability under practical channel loads. In high load scenarios, even if the preamble is perfectly acquired, packets can still be irrecoverable due to the high MAI. Therefore, the MAI is in general the dominant effect in the packet error probability (see for instance the numerical results provided in [CDH07]).

A first attempt to study the joint effect of modulation, coding, packet length and the channel statistics in the throughput of slotted ALOHA was made by Zhang in [ZP92]. The proposed approach consisted in evaluating the throughput considering the statistics of the decision variable for each symbol of a given test packet. Thus, assuming a packet of L_p symbols, the decision variable for the l-th symbol given K colliding packets can be written

[2]Spread-spectrum systems, as for instance Spread ALOHA, are an exception since the interference can be better modelled as a Gaussian process.

$$z_l = A_0 b_0(l) + \sum_{k=1}^{K} A_k(l) \cos[\phi_k(l)] b_k(l) + n_l, \tag{4.12}$$

where $A_0(l)$ and $b_0(l)$ are the amplitude and the transmitted symbol for the test packet and $A_k(l)$, $b_k(l)$ and $\phi_k(l)$ are the amplitude, transmitted symbol and phase, respectively, for an interfering packet k, $1 \leq k \leq K$.

Similar models have been proposed for the evaluation of the slotted ALOHA throughput performance under different scenarios, including the LEO satellite channel [RWS99; Ren+00]. In particular, detailed PHY models are required for the study of cross-layer RA schemes introducing signal processing techniques, such as [CDH07; Liv11; TB00]. We will review some of these techniques later on in this chapter.

4.4 ALOHA and Some Derivations

In this section we review some of the more fundamental RA protocols proposed in the literature, starting with Abramson's ALOHA system. Examples of both synchronous and asynchronous approaches will be considered. For some simple cases, we also derive closed-form expressions for the throughput performance.

4.4.1 The ALOHA System

Next we review the original ALOHA system proposed in [Abr70]. Consider a multiuser network consisting of an infinite number of terminal nodes which transmit packets very infrequently. We assume that the aggregate traffic load from all users—including new-packet arrivals and lost packets that have been scheduled for retransmission—forms a Poisson source[3] of mean λ [packets/s], so that the normalized channel load is $G = \lambda T_p$, where T_p is the packet length. In ALOHA, a given packet is assumed to be successfully transmitted if and only if no other packet transmission starts within a $\pm T_p$ time-window, in which case a collision occurs and none of the packets involved can be recovered by the receiver.

Since the number of packet arrivals K during an interval of length $2T_p$ follows a Poisson distribution of intensity $2\lambda T_p = 2G$, we have

$$P(K = k) = \frac{(2G)^k}{k!} \exp[-2G]. \tag{4.13}$$

[3]In fact, it is generally assumed that newly generated packets follow a Poisson distribution and that lost packets are retransmitted after a random, sufficiently long time, so that the resulting channel load can still be considered Poisson distributed. We note that, while this approach is particularly convenient for throughput performance evaluation purposes, it does not guarantees system stability.

and the probability that a packet is successfully transmitted is simply

$$\begin{aligned} P_s &= P(K=0) \\ &= \exp(-2G). \end{aligned} \tag{4.14}$$

The normalized throughput of the ALOHA protocol is therefore given by

$$\begin{aligned} T(G) &= G \times P_s \\ &= G\exp(-2G). \end{aligned} \tag{4.15}$$

On the other hand, the MAC packet error rate can be obtained from Eq. (4.3)

$$\begin{aligned} P_e(G) &= 1 - T(G)/G \\ &= 1 - \exp(-2G). \end{aligned} \tag{4.16}$$

Note, from Eq. (4.15), that the maximum throughput of ALOHA is $T_{max} = 1/(2e)$ packets/packet duration and is achieved at a MAC load of $G = 1/2$ packets/packet duration. However, it is clear that operating an ALOHA system close to its maximum is highly inefficient since $P_e(G = 1/2) \simeq 0.63$. As consequence, it is often preferable to operate ALOHA over the nearly linear operation region. As an example, consider a system where the target maximum PER is 10^{-1}. In this case, the maximum MAC load is about $G = 0.053$ and the corresponding maximum throughput is 0.045.

4.4.2 Slotted ALOHA

As we have already noted, the first improvement over ALOHA, known as slotted ALOHA (SA), was proposed by Roberts [Rob75] and consisted in dividing the time into slots of length T_p s, so that each user can only transmit at the start of a time slot. With this constraint, a packet is successfully decoded if no other packet is transmitted over the same time slot. In this case, the number of colliding packets K is a Poisson random variable of mean $\lambda T_p = G$. Hence, $P(K = 0) = \exp[-G]$ and the throughput is

$$T(G) = G\exp(-G), \tag{4.17}$$

while the PER is given by

$$P_e(G) = 1 - \exp(-G). \tag{4.18}$$

Maximising Eq. (4.17) yields $T_{max} = 1/e$ [packets/slot], which is obtained for $G = 1$ and represents a 100% improvement with respect to ALOHA. Similarly, if we consider again a maximum PER equal to 10^{-1}, the maximum channel load is now about $G = 0.105$ [packets/slot], which is also twice what we obtained with ALOHA.

4.4.3 Diversity Slotted ALOHA

Proposed in the early 1980s as a generalization of SA, Diversity Slotted ALOHA (DSA) [CR83] was specially designed to operate on satellite communications systems. The principle is simple: instead of transmitting randomly a single version of a packet as in SA, a terminal transmits multiple *replicas* in order to exploit time and/or frequency diversity. This technique slightly reduces the maximum achievable throughput, but does improves the packet error rate by over one order of magnitude with respect to SA, thus allowing to operate at higher MAC loads.

The DSA throughput can be computed as follows. Let us assume that each user transmits k packet replicas over different random slots within a MAC frame, and that the packet generation process is Poisson. If G denotes the average packet arrival rate per slot, the effective load on each slot is kG. Under the collision channel model, the probability that a packet replica is successfully transmitted is therefore

$$P'_s = \exp(-kG). \tag{4.19}$$

Then, the PER equals the probability that all k replicas of a packet are lost:

$$P_e = (1 - P'_s)^k, \tag{4.20}$$

so the DSA throughput is simply

$$T(G;k) = G \times \left[1 - (1 - \exp(-kG))^k\right]. \tag{4.21}$$

4.4.4 Selective Reject ALOHA

[Ray84] introduces a variation of ALOHA called Selective Reject (SREJ) ALOHA that allows to achieve a maximum throughput close to the slotted ALOHA peak of $\sim 1/e$, but in a complete asynchronous manner. The technique consists in dividing the messages into several subpackets so that when a collision occurs only those subpackets that were lost are retransmitted, following a selective Automatic Repeat Request (ARQ) retransmission approach. Besides, it is shown in [Ray84] that a similar technique can be used in slotted ALOHA systems with variable packet lengths yielding the same throughput performance. Because of these advantages, SREJ-Aloha has been considered as a reference scheme for the design of some store-and-forward satellite systems (see for instance [WS00; Had98]).

4.4.5 Spread Spectrum ALOHA

In [Abr94], Abramson proposes Spread Spectrum ALOHA (SSA) as a different view of the ALOHA protocol, by examining a scenario in which a much more wider frequency band is available. This way, packets can be transmitted using a spreading code pattern, as in a single-code CDMA system. In fact, such a technique would not introduce any performance gain. However, SSA can be enhanced by the use of advanced forward error correction (*e.g.* Turbo codes) and relatively small packet sizes [ZC05]. The result is remarkable: if we evaluate the normalized throughput performance for a given BER target, SSA significantly outperforms SA. In addition, SSA provides a more extended linear throughput region, which allows to operate at higher MAC loads. Another important advantage of SSA is that it is a fully asynchronous approach. The main disadvantage of SSA is that, as conventional CDMA systems, its performance is severely harmed under power unbalance[4] conditions [DD12].

4.5 High-throughput Synchronous Random Access

4.5.1 Contention Resolution Diversity Slotted ALOHA

A major breakthrough in RA communications has been made recently with the introduction of Contention Resolution Diversity Slotted ALOHA (CRDSA) [CDH07]. Based on DSA, in CRDSA the packets transmissions are organized in MAC frames consisting of N_{slots} slots. When a user has a packet to transmit, it sends two replicas in random slots within the frame. In contrast to DSA, each replica also contains a pointer with the relative location of its twin. In this way, when one replica is successfully decoded at the receiver, the other one can be removed from the frame by means of interference cancellation (IC) techniques. Further, since cancelling the interference of a replica may allow to decode one or more additional replicas, a successive interference cancellation (SIC) algorithm is proposed. This process is illustrated in Figure 4.1 for a frame size $N_{slots} = 6$ and three active users A, B and C. The receiver first decodes replica A_1 from user A, which is free of interference. Then, its twin A_2 is removed from the frame. Since replica B_2 is now free of interference, it can be decoded, and its twin B_1 removed from the frame. The packet from user C can be recovered either from replica C_1, after removing B_1 from the frame, or, independently, from replica C_2, which was initially free of interference.

As DSA, CRDSA was designed to further improve the performance of random access satellite channels, and has been included in the Digital Video Broadcasting (DVB) - Return Channel via Satellite (RCS) standard [Dvb].

[4] This is the *near–far* problem mentioned in Section 3.4.1.3.

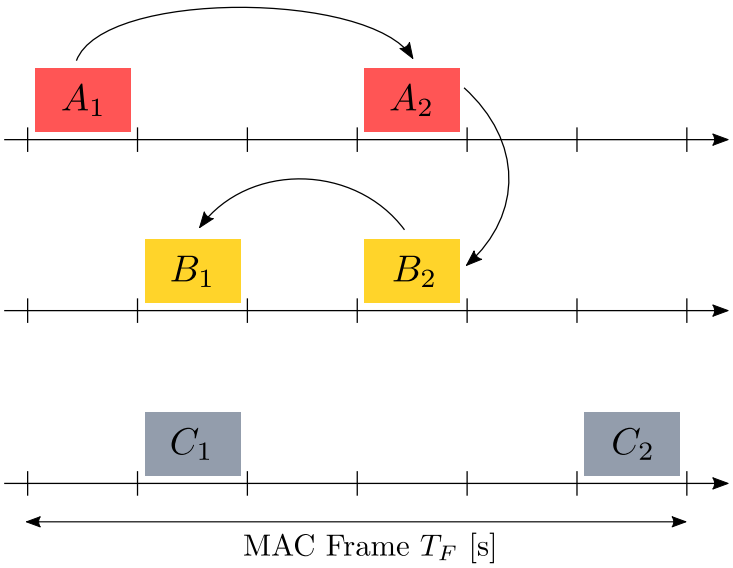


Figure 4.1 – A representation of the SIC process in CRDSA.

4.5.2 Irregular Repetition Slotted ALOHA

In [Liv11], Liva proposes Irregular Repetition Slotted ALOHA (IRSA) and generalizes the principles of CRDSA by allowing each terminal to transmit different number of packet replicas. The repetition rate is randomly selected following a distribution $\{\Lambda_l\}$. Noting that the iterative interference cancellation process can be represented as a message-passing decoding algorithm, Liva establishes a strong connection with low density parity check (LDPC) decoding, which provides a mathematical framework for throughput optimization. Thus, he shows that an asymptotic throughput (that is, what can be achieved with very large frames) close to 0.97 [packets/slot] can be achieved.

4.5.3 Coded Slotted ALOHA

Introduced in [PLC11], Coded Slotted ALOHA (CSA) is based on the application of erasure coding at the packet layer—an idea that has been originally proposed in [LS90]—together with interference cancellation at the receiver side. The principle is somewhat similar to that of IRSA, but instead of transmitting packet replicas, users transmit encoded blocks derived from the original packet. The process can be summarized as follows. First, a packet is divided in k subpackets of equal length. Then, the k subpackets are encoded through a (n_i, k) erasure code $\mathcal{C}_\rangle$, which is randomly selected from a set $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_{n_c}\}$ of n_c codes. Also, within each encoded block, a reference to the other $n_i - 1$ blocks is appended, which allows to perform SIC. The n_i encoded subpackets are then

transmitted over random slots in a MAC frame of size N_{CSA}. Notice that in contrast to IRSA, with CSA redundancy rates greater than 1/2 can be achieved. Results reported in [PLC11] show that, as $N_{CSA} \to \infty$, an asymptotic throughput performance close to 0.97 [packets/slot] can be achieved, which is the same result obtained for IRSA.

4.6 High-throughput Asynchronous Random Access

4.6.1 Contention Resolution ALOHA

As an alternative to simplify the transmitter design in CRDSA/IRSA, [Kis11] proposes to eliminate the slot boundaries within the MAC frame and to allow terminals to transmit their replicas at any location within the MAC frame. At the receiver side, a SIC algorithm is applied in order to resolve the replica collisions in a similar way as in CRDSA. The protocol, called Contention Resolution ALOHA (CRA), has paved the way for the development of high-throughput asynchronous RA protocols. In terms of throughput performance, simulation results in [Kis11] show that CRA actually outperforms CRDSA, achieving a near linear throughput for normalized channel loads close to 1[5].

4.6.2 Enhanced Spread Spectrum ALOHA

Recently introduced by [DD12] as a RA solution for satellite-based M2M systems, Enhanced Spread Spectrum ALOHA (E-SSA) significantly improves SSA through a sliding-window interference cancellation algorithm implemented at the gateway receiver. E-SSA not only provides a high-throughput/low-PER performance, but also it is more robust against power unbalance than its predecessor. Another interesting feature in E-SSA is that it operates in a fully asynchronous manner and without the need for packet replicas, which results in a higher efficiency. A maximum throughput of 1.9 [b/s/Hz] was reported in [DD12], obtained for a QPSK and turbo FEC 1/3 PHY layer and a log-normal channel environment. E-SSA has been selected as one of the return link access schemes in S-band Mobile Interactive Multimedia (S-MIM) [Sca+13], a novel standard to provide mobile messaging services over GEO satellite systems.

4.6.3 Asynchronous Contention Resolution Diversity ALOHA

While CRA relaxes the tight synchronisation requirements of CRDSA, it still requires the terminals to be synchronized with the frame timing. The first fully asynchronous, high-throughput RA scheme proposed in the literature was Asynchronous Contention Resolution Diversity ALOHA

[5] Assuming a PHY layer consisting of QPSK modulation, FEC 1/2 and an operating signal-to-noise ratio of 10 [dB].

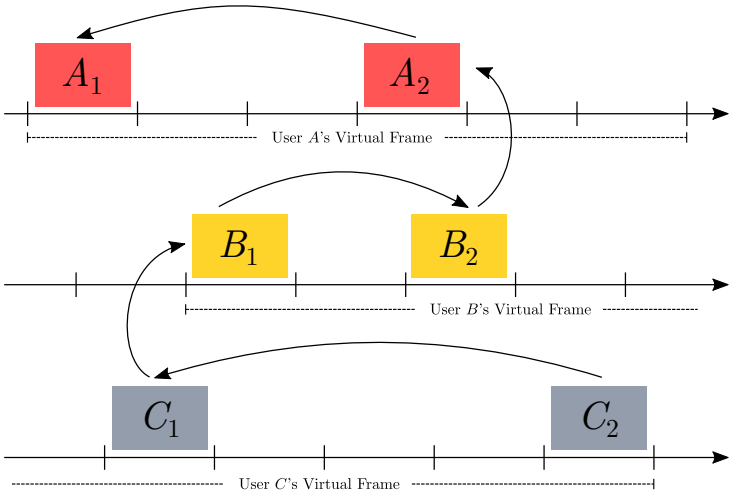


Figure 4.2 – A representation of the SIC process in ACRDA.

(ACRDA) [De +14]. Instead of following a global frame reference as in CRDSA, with ACRDA each terminal keeps a local frame reference called *virtual frame* (VF) and transmits packet replicas in random slots within its own VF. With respect to CRA, this approach allows to simplify the signalling overhead required to indicate the relative replica locations. At the receiver side, a sliding-window architecture similar to that of E-SSA is implemented. Figure 4.2 illustrates the iterative decoding process of ACRDA for a virtual frame of size $N_{slots} = 6$ and three active users. Considering a PHY layer based on QPSK and turbo FEC 1/3, ACRDA achieves a maximum throughput near 1 [b/s/Hz] for a signal-to-noise ratio equal to 10 [dB] and perfect power control; and near 1.6 [b/s/Hz] under log-normal power imbalance.

4.7 Summary

In this chapter, we have reviewed fundamental concepts of random access communications, including performance metrics such as throughput and packet error rate, which will be used later on in the analysis of time-asynchronous and frequency-asynchronous random access. We presented the main reception models adopted in the literature, and discussed some of the underlying assumptions upon which they are based. Finally, we have also provided a short review of the state-of-the-art in random access protocols, starting from pure ALOHA to high-throughput schemes implementing successive interference cancellation. Table 4.1 summarizes the throughput performances provided by the main schemes we have considered. For those schemes that are based on the collision channel model and do not assume any particular PHY layer configuration, we have considered (for practical reasons) that the underlying transmission parameters are such that the spectral efficiency is $\eta = 1$

[b/s/Hz][6].

RA scheme	**Peak T_{max} [b/s/Hz]**	**T at target $P_e \leq 10^{-3}$ [b/s/Hz]**	**Observations**
SA	$1/e \simeq 0.37$	10^{-3}	Under the collision model.
DSA	$\simeq 0.31$	$\simeq 1.7 \times 10^{-2}$	Under the collision model.
SSA	$\simeq 0.62$	$\simeq 0.5$	Turbo FEC, small packets and perfect power control.
CRDSA	$\simeq 0.52$	$\simeq 5 \times 10^{-2}$	Under the collision model.
IRSA	$\simeq 0.97$	$\simeq 0.97$	Under the collision model. Infinite frame size
ACRDA	$\simeq 1.6$	$\simeq 1.5$	Under the Gaussian, log-normal channel with $\sigma = 3$ [dB] and $E_b/N_0 = 10$ [dB]. QPSK, Turbo FEC.
CSA	$\simeq 0.97$	$\simeq 0.97$	Under the collision model. Infinite frame size.
E-SSA	$\simeq 1.9$	$\simeq 1.9$	For a Log-normal channel with $\sigma = 3$ [dB] and $E_b/N_0 = 13.7$ [dB]. QPSK, Turbo FEC.

Table 4.1 – Throughput (b/s/Hz) for different RA schemes.

[6]This is roughly what we can obtain with QPSK modulation and rate 1/2 FEC.

Part II

Contributions

"Someone told me that each equation I included in the book would halve the sales."

Stephen Hawking (1942 – present)

5

A Delay-Tolerant Multiple Access Approach

Contents

In this chapter we explore some benefits obtained by exploiting the delay-tolerant nature of the network. Our focus is on simple techniques that can be easily and rapidly implemented in practical systems.

5.1 Introduction

It it clear from the previous discussions that improving the uplink capacity represents a major challenge in data collection networks using small satellite platforms and small terminals. Before

studying a detailed architecture, in this chapter we explore some improvements that can be obtained simply by taking advantage of the delay-tolerant nature of the network and that can be implemented transparently. We explore two different approaches, though they can be used in combination. The first, which we present in Section 5.2, consists simply on adopting a transmission control mechanism through which nodes transmit only above a minimum elevation angle. By increasing this threshold the propagation path length is reduced—resulting also in reduced transmission power— but the number of access opportunities is reduced as well. Therefore, there is trade-off between energy efficiency and access delay that must be evaluated.

In the second approach, presented in Section 5.3, the key idea is to postpone the reconstruction of the transmitted message packets until they are downloaded by the gateways, which are, in principle, considerable more powerful than typical CubeSat platforms. To illustrate this concept, we will use as baseline model the well known slotted Aloha (SA) random access (RA) protocol combined with an optimal erasure code. Our objective is to study the throughput, packet error rate (PER) and energy efficiency (EE) properties of this solution in the context of a nanosatellite machine-type data collection network. In particular, the throughput–energy efficiency trade-off is discussed. The subjects discussed in this Chapter have been included in [AF17b].

5.2 Power Reduction Through Transmission Control

In the following we assume that each terminal node with a satellite in view is able to obtain an estimate $\hat{\epsilon}$ of the elevation angle ϵ at time $t = k\tau$, for $k = 0, 1, 2, ...$ and some $\tau > 0$ that depends on the estimation algorithm. The estimate can be obtained through several ways. One alternative is to use the Doppler information, as proposed in [Ali+98], which can be obtained using the satellite beacon. Another alternative consists on computing the elevation angle from the terminal and satellite positions. Naturally, this method requires a positioning system in both ends.

When a new packet is generated at some node, it is transmitted in the next transmission interval $t = k\tau$ with probability

$$p(t = k\tau) = \begin{cases} 1, & \text{if } \hat{\epsilon} \geq \epsilon_{min}, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

Our objective now is to evaluate the EE of this simple transmission control scheme.

The EE of a point-to-point link, noted η_{EE} [bits/joule], is defined as the information rate divided by the transmission power [Che+11]:

$$\eta_{EE} \triangleq \frac{R_b}{P_t}. \tag{5.2}$$

Following the analysis in Section 3.3, we assume that the uplink is designed in order to achieve a minimum SNR γ_0 at the minimum elevation angle ϵ_{min}:

$$\gamma_0 = \frac{P_t G_t G_r}{L_{FS}(\epsilon_{min}) N_0 R_b}, \tag{5.3}$$

where P_t and G_t are the terminal power and antenna gain, respectively; G_r is the satellite antenna gain, N_0 the noise density at the satellite receiver input, $L_{FS}(\epsilon)$ the free space loss attenuation and R_b is the uplink data rate[1]. It follows that

$$P_t = \overbrace{\left(\frac{\gamma_0 N_0}{G_t G_r}\right)}^{1/\alpha} L_{FS}(\epsilon) R_b, \tag{5.4}$$

where α is a constant. Combining Eq. (5.2) and Eq. (5.4), we get

$$\eta_{EE} \propto \frac{1}{L_{FS}(\epsilon_{min})}. \tag{5.5}$$

Thus, Eq. (5.5) can be used to describe the link EE as a function of the minimum elevation angle at which terminals may transmit.

5.2.1 The Energy Efficiency–Access Delay Trade-off

Given that $L_{FS}(\epsilon_{min})$ decreases monotonically with ϵ_{min}, a system designer interested in maximizing η_{EE} may feel tempted to select a high value for ϵ_{min}. However, as the threshold ϵ_{min} is increased, the access delay (*i.e.*, the waiting time experienced by a newly generated packet until it is successfully transmitted to the satellite) increases as well because the satellite footprint is reduced. Naturally, higher access delays translate into reduced communication capacity. We are thus faced with a trade-off between EE and access delay.

Computing the average access delay as a function of ϵ_{min} is usually done through a computer simulation tool such as the System Tool Kit (STK)®. Alternatively, we may compute the average satellite passes per day (PPD)—a metric that is directly related to access delay. A closed-form expression to compute the approximate average PPD is given in [Bur13]. Figure 5.1 shows the EE–PPD trade-off for a target latitude $L = 45°$ and a single-satellite network consisting of one near-polar orbiting satellite at an altitude of 600 [km]. For convenience, we plotted the normalized EE defined as $\hat{\eta}_{EE}(\epsilon_{min}) \triangleq \eta_{EE}(\epsilon_{min})/\eta_{EE}(\pi/2)$.

[1]Again, others sources of attenuation have been ignored in order to simplify the analysis, though these can be easily incorporated for more accurate results.

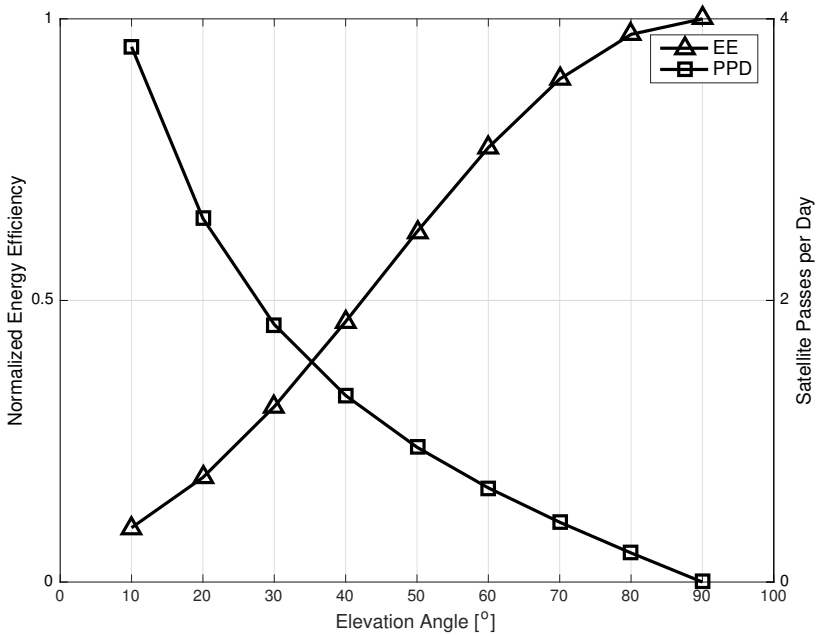


Figure 5.1 – EE-PPD relation for a single satellite network in near-polar orbit.

In particular, the results show that increasing ϵ_{min} from 10° up to the 20° – 30° region involves a significant reduction of the PPD. This can be explained by the fact that a great portion of the coverage area corresponds to low elevation angles. For instance, the footprint area for all $\epsilon > \epsilon_{min}$ is $A \approx 9.69 \times 10^6$ [km^2] and $A \approx 4.53 \times 10^6$ [km^2] for $\epsilon_{min} = 10°$ and $\epsilon_{min} = 20°$, respectively. In spite of this, reducing the PPD by about 50% (from $\epsilon_{min} = 10°$ to $\epsilon_{min} = 30°$) allows to improve the EE by more than 300%.

Practical Issues

It is important to note that this transmission control scheme would only be useful in the ideal scenario wherein both the satellite and terminal antennas have an ideal, omnidirectional radiation pattern. In practice, this is not the case and the resulting attenuation is not only a function of the propagation path length but also on the combined gain in the terminal–satellite direction. An alternative transmission control scheme can be based on estimates of the SNR [JB93], which a terminal can also obtain by monitoring the satellite beacon. In this case, the transmission probability is given by

$$p(t = k\tau) = \begin{cases} 1, & \text{if } \hat{\gamma} \geq \gamma_0, \\ 0, & \text{otherwise;} \end{cases} \tag{5.6}$$

where $\hat{\gamma}$ is the estimated SNR and γ_0 the target SNR. Note that adopting this transmission control scheme also involves an EE-delay trade-off since the access opportunities are reduced as γ_0 is increased. An analysis of this case has not been included since it has been considered too system-dependent (*i.e.*, it depends on the actual radiation patterns).

5.3 Throughput Improvement Through Packet-layer Coding

In Chapter 4 we reviewed some high-throughput RA schemes recently proposed, in particular those targeting GEO satellite systems. Most of these techniques, however, have been conceived to operate over powerful gateway receivers. As GEO systems are in general non-delay tolerant, packet retransmissions come at a high cost due to the access delay component associated to the long propagation distance. With high-throughput RA schemes retransmissions are avoided by implementing complex signal processing techniques allowing to resolve some packet collisions. Thus, high throughputs can be achieved while keeping a negligible packet error rate, resulting also in increased EE. Because nanosatellite platforms have very limited signal processing capabilities, implementing high-throughput RA schemes might not always be possible. Therefore, novel solutions presenting a good compromise between complexity and performance must be conceived. In this section we evaluate the use of packet-layer coding as an alternative to improve the throughput performance without requiring additional protocol overhead at the satellite side.

Let us start by recalling the network model described in Chapter 2. As mentioned, we consider a data collection application in which the information flow is unidirectional. Messages are generated and transmitted by terminal nodes and then, after being collected by the satellite(s), are gathered at the ground station gateways, at which point we assume they are available to the end-users. In this context, an end-to-end transmission occurs between terminal nodes and gateways.

Following the standard layered architecture, end-to-end reliability can be provided through some retransmission scheme implemented at the transport layer. In addition, point-to-point reliability can be provided by the link layer through some error control strategy. The approach that we consider here to improve the overall reliability is the use of cross-layer forward error correction (FEC). More specifically, we consider cross-layer FEC at the link layer as a means to improve the throughput performance in a RA-based uplink channel. Globally, the idea is as follows. At the transmitter side, divide each application message into k blocks, then encode the blocks through a (n, k) erasure code and finally transmit the n encoded blocks according to the channel access scheme adopted. This process is illustrated in Figure 5.2. At the receiver side, the original message can be fully recovered if at least k out of n blocks are correctly transmitted.

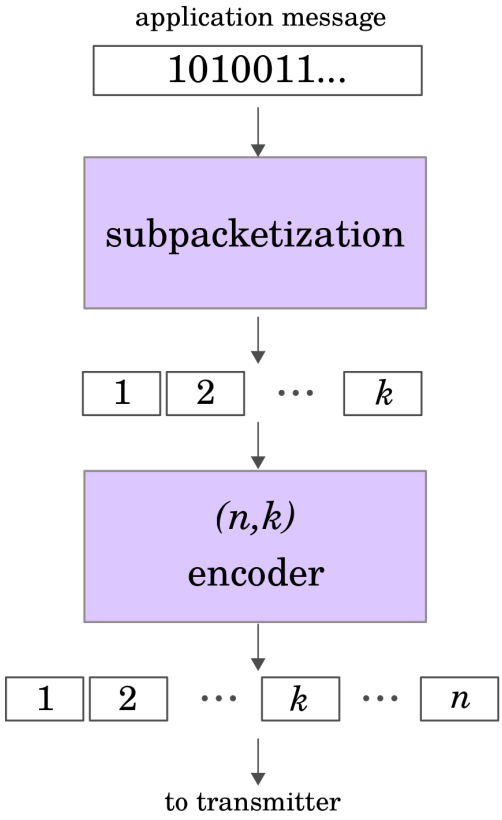


Figure 5.2 – Encoder diagram.

We note that, since the messages are available to the end-users after being received by the ground station gateways, the decoding process can be delayed until the blocks are received by the gateways, without compromising the application performance. Hence, from the point of view of the satellite, this process can be completely transparent as no additional logic/signal processing is required. Essentially, the satellite simply keeps seeing a flow of packets and the additional decoding complexity is assumed by the gateway receivers, which are considerable more powerful.

In terms of protocol architecture, this approach can be implemented at the transport layer because it operates as an end-to-end reliability mechanism. In this case, a whole application message is encoded forming n packets, and we talk about transport-layer coding. However, as discussed in Section 3.2.4, a transport layer is rather dispensable in our context, so a link-layer implementation could be considered in order to simplify the design. In this case, k link-layer packets are aggregated and then encoded to form n packets—this mechanism is referred to as packet-layer coding. Alternatively, we may consider that a single packet message can be divided into k *subpackets*. End-to-end packet-layer coding can be viewed as a delay-tolerant approach to cope with the packet collisions in RA communications. For a general discussion on higher-layer

FEC, the reader is referred to [Cor07, Section 4.5]).

5.3.1 Performance Evaluation

The use of erasure codes as an alternative to improve the throughput performance of RA systems has been originally proposed by Lam with the introduction of Time-Hopping and Frequency-Hopping Multiple-Access (TH/FH-MA) [LS90]. With Lam's approach, a message packet is encoded forming n subpackets. Each subpacket is then transmitted following a time-hopping and/or a frequency-hopping pattern. Hence, a time division/frequency division of the channel is required. While this approach is interesting—specially for performance evaluation purposes—it is not optimal in our context because the design of hopping patterns proposed by Lam does not take into account the impact of the time-varying LEO channel; rather, it is based on the collision channel model. Certainly, by forcing the transmissions to follow predefined patterns in a time-varying channel, several subpackets can be transmitted during deep fades, resulting in performance degradation. As discussed above, a transmission control strategy based on the estimated SNR is preferred in this scenario and will be considered in our analysis.

Our objective now is to evaluate the throughput performance that can be obtained with this scheme in a typical SA RA system. The main assumptions considered in our analysis are the following. The aggregate packet arrival process follows a Poisson distribution of rate λ [packets/s]. The packet duration is fixed to T_p [s]. The normalized channel load G is defined as the number of packets arriving to the system during the packet duration, *i.e.*, $G = \lambda T_p$ [packets/slot]. Each terminal node uses the same (n, k) erasure code to encode and transmit its packets. We assume that this code is optimal (*i.e.* it is a maximum distance separable (MDS) code), so that each message can be decoded if any k of the n subpackets are received correctly. The network is composed by a large number of terminal nodes, so we may assume that each transmission is generated by a different node. The uplink channel is divided in time slots of duration equal to the subpacket duration of T_p/k [s]. We will base our analysis of the uplink channel on the SINR-based capture reception model with a capture threshold $b = 3$ [dB], as well as on the collision channel model. These models are described in Section 4.3.1. As mentioned, we consider that each subpacket is transmitted following the transmission control scheme described by Eq. (5.6). To model the imperfections in the SNR estimation algorithm (for instance due to uplink-downlink channel decorrelation or due to important channel fluctuations within subsequent estimates), we assume that the received power for each subpacket at the satellite receiver input follows a log-normal distribution [Cor07]:

$$f_Z(z) = \frac{10}{\sqrt{2\pi}\ln(10)z} \exp\left[-\frac{(10\log_{10} z - \mu)^2}{2\sigma^2}\right], \tag{5.7}$$

where the random variable $Z = A_k^2$ is the square of the packet amplitudes A_k's, and μ and σ are

the mean and standard deviation, expressed in dB, respectively. Note that we are also assuming that the channel is flat at least during T_p/k (*i.e.*, the channel coherence time is greater than the subpacket duration). Since in general $T_w >> T_p$, T_w being the visibility window, we suppose that the n subpackets generated per message can always be transmitted during the satellite passage. Finally, we assume that only the uplink impacts the overall packet error rate. On the downlink, subpackets from different users are multiplexed and transmitted reliably (using for instance, an ARQ error control scheme and a robust PHY layer).

We note that, if the subpackets transmissions are sufficiently randomized over time (which is likely due to the underlying transmission control scheme), then we may assume that the load on the channel is also Poisson. In a slot of duration T_p/k the average number of packet arrivals is $\lambda T_p/k = G/k$. Since n subpackets are generated for every packet, the effective normalized load on the channel (in [subpackets/slot]) is

$$G' = (n/k)G. \tag{5.8}$$

Now, since k successful subpacket transmissions are required to recover a packet, the MAC packet error rate P_e is simply given by

$$P_e = \sum_{i=n-k+1}^{n} \binom{n}{i} \delta^i (1-\delta)^{n-i}, \tag{5.9}$$

where δ is the subpacket error rate. For the collision channel model, we have

$$\delta = 1 - e^{(-nG/k)}, \tag{5.10}$$

which is simply one minus the probability of having zero subpackets transmitted in a slot. For the capture channel with transmission control, the subpacket error probability δ is obtained through computer simulation, as closed-form expressions cannot be derived. Then, we can compute the MAC packet error rate following the same steps presented above. The throughput is finally obtained using the definition $T = G(1 - P_e)$.

Figure 5.3 depicts the throughput–PER performance for both the collision and the capture channel of the SA using a $(n, k) = (60, 20)$ erasure code. The throughput–PER performance of the SA has also been provided as reference. For the capture channel, we set SNR = 20 [dB], $\mu = 0$ [dB] and $\sigma = 3$ [dB]. We observe that some extent of power imbalance has a positive impact on the performance, allowing to reach throughputs close to 0.38 for a PER $P_e \leq 10^{-3}$. Also, the maximum throughput obtained for the capture channel is almost twice the maximum obtained for the collision channel. In either case, the performance gain for low packet error rates is significantly higher with respect to plain SA.

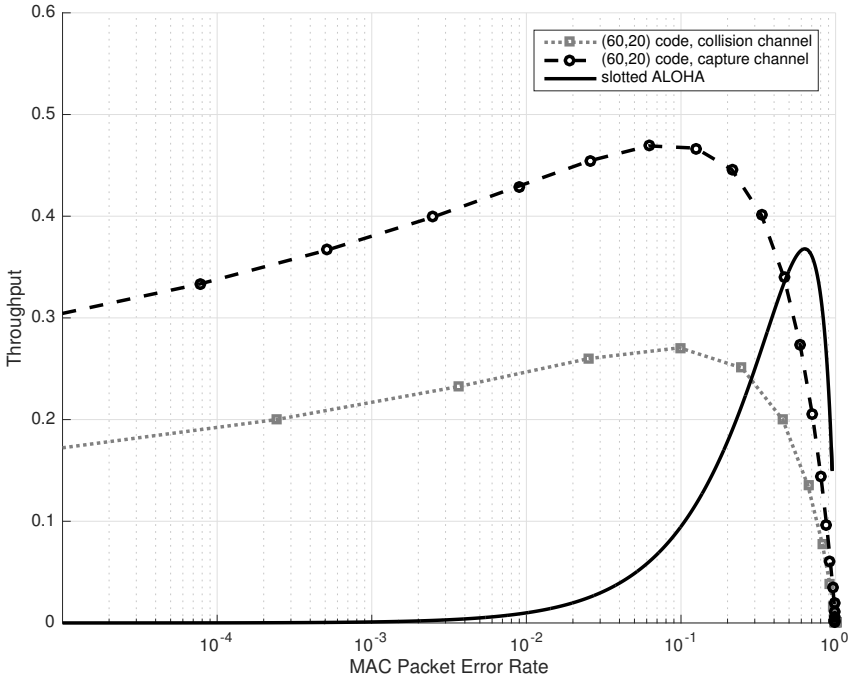


Figure 5.3 – Throughput vs PER for the SA system using a (60,20) erasure error correcting code.

5.3.2 The Throughput–Energy Efficiency Trade-off

While it is clear that the use of erasure coding at the packet level provides a notable improvement in the throughput–PER performance, we have not yet accounted for the additional energy cost that this method entails. In order to do so, we start by defining the EE of a MAC system with a single-user receiver architecture as

$$\eta_{EE}^{MAC} \triangleq \frac{SC}{GP_t} = \frac{C(1-P_e)}{P_t}, \tag{5.11}$$

where C is the point-to-point (or the single-user) link capacity and GP_t is the average power on the channel. Now, we may use Eq. (5.11) noting that the effective power on the channel is scaled by n/k. For convenience, we normalize Eq. (5.11) by the factor C/P_t, since both parameters are constant. In Figure 5.4 we have plotted the throughput–EE trade-off for SA using tree coding rates and the same capture channel considered above. Again, the performance of plain SA is also provided as reference. Note that for each coding rate there is an important region where the throughput–EE performance is significantly improved with respect to the SA system. This is in contrast to what is observed in plain SA, where power imbalance has a negative impact in terms of EE. Note also that the maximum normalized EE achieved by each coding rate is limited by

the factor $1/(n/k)$. Naturally, plain SA offers higher EE, but at lower throughputs. Finally, these results suggest that there is an optimal coding rate below which the throughput–EE performance decreases, but above which a higher EE can be achieved at the cost of decreased throughput.

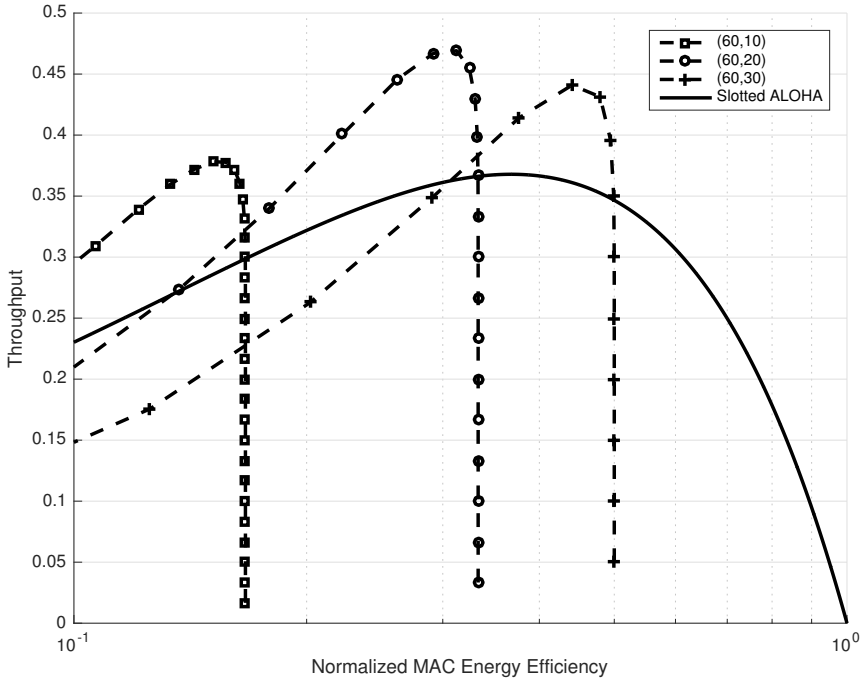


Figure 5.4 – Throughput–Energy Efficiency trade-off for different erasure codes and plain SA.

Practical Issues

An obvious problem with this approach is that additional protocol overhead is required to transmit a message in several blocks. This is particularly true in the short-packet regime. Packet-layer coding using packet aggregation helps to avoid this issue, but it also increases the end-to-end delay significantly (specially in applications where each node generates low traffic). Another alternative is to use packet repetition (as in diversity ALOHA), however this provides only coding rates $k/n < 1/2$ and hence less flexibility to find a good compromise in terms of throughput–EE.

5.4 Summary

In this Chapter, we have explored simple mechanisms that can be implemented in a delay-tolerant small satellite data collection system. We have seen that a transmission control scheme can help

reduce the transmission power significantly, as long as there is flexibility in terms of delay requirements. We have also evaluated the use of packet-layer erasure coding to improve the performance of the uplink RA channel while providing improved reliability. The throughput-EE merits of this approach have also been discussed.

"What is not surrounded by uncertainty cannot be the truth."
Richard Feynman (1918 – 1988)

6

Time- and Frequency-Asynchronous ALOHA

Contents

Time- and Frequency-Asynchronous Aloha (TFAA) is a novel random access scheme in which packet transmissions are neither coordinated in the time domain nor in the frequency domain. Its original interest follows from the adoption of Ultra Narrow Band (UNB) modulation techniques

in terrestrial LPWANs [GG15] based on Sigfox's protocol. By allowing to trade transmission data rate with communication range (or transmission power), TFAA is particularly attractive in power constrained applications such as LPWAN or M2M over satellite, which is the main reason why it has been considered in this work.

The main objective of this chapter is to study the throughput performance of TFAA over three different reception models: the collision channel, a SINR-based capture channel, and also over a more detailed PHY model that takes into account the transmission parameters. For the collision channel, an exact, closed-form expression is derived, while for the other models we have developed a semi-analytical approach. In addition, a computer simulation tool has been developed in order to validate the results obtained for the PHY model. In this case, our results show that the performance gap between ALOHA and TFAA is rather small, and suggest that important improvements might be achieved by means of interference mitigation techniques (such as forward error correction). The main results presented throughout this chapter have been published in [AF17c].

6.1 Introduction

The idea of adopting a frequency-asynchronous, ALOHA-like random access approach has come recently with the introduction of UNB techniques as the radio interface in Sigfox's[1] LPWANs [DGG14a; GG15], though other systems have used a similar approach several years before (*e.g.* the Argos system). There is, in fact, no clear consensus regarding the name of this access technique: in [DGG14a] it is introduced as Random Frequency Division Multiple Access (R-FDMA), while in [Ant+15] it is referred to as time-frequency ALOHA (TFA). In this work, the name time- and frequency-asynchronous ALOHA (TFAA) has been adopted in order to indicate explicitly that both the time and frequency domains are non-slotted.

In UNB systems, the signal bandwidth—typically in the order of 100 Hz—is comparable to the oscillator frequency uncertainty. Hence, the use of a FDMA approach is inefficient, as the required guard bands become significantly large with respect to the signal bandwidth. Similarly, the use of a slotted time reference, as in TDMA, is not practical because of the heterogeneous propagation delays experimented by different terminal nodes within a LPWAN cell, and the additional signalling overhead required to keep network-wide synchronization. Thus, TFAA appears as a natural solution, specially for *bursty* traffic conditions and short packet lengths.

While unslotted ALOHA-based systems—which are time-asynchronous only—have received considerable attention over the past decades, very few studies on TFAA can be found currently. The first attempt to study the network capacity of TFAA in a UNB LPWAN context is presented in [DGG14a; DGG14b]. The analysis is based on the bit error probabilities and does not take

[1] See https://www.sigfox.com/

into account the effects of the system dynamics on the interference power, which is assumed to follow a Gaussian distribution. We will show later on that this approach can be misleading in packet radio networks since the MAI may vary significantly in a symbol-by-symbol basis and, at the same time, the symbol error probabilities are not independent. A different approach based on the same UNB scenario is adopted in [Ant+15], where the MAC protocol performance is studied following conventional assumptions employed for ALOHA random access systems. In particular, the authors consider a collision channel model and Poisson-distributed arrivals, and they also study the impact of the high Doppler effects that are present in LEO satellite systems. Finally, a MAC-level performance evaluation framework is proposed in [Li+16]. In contrast to [Ant+15], this work is based on a capture, PHY-independent channel and a constant traffic model (which is more suited for some sensing applications).

In this thesis, we study the TFAA technique from a more general perspective, *i.e.*, without restricting it to the UNB scenario. Broadly speaking, we consider a random access system characterized by 1) narrow-band (or UNB) packet transmissions that do not follow any slot timing reference and use random carrier frequencies within a given *system bandwidth* W, 2) a *signal bandwidth* B smaller than W and 3) Poisson-distributed packet arrivals. (Further modelling details will be given throughout this chapter). In this context, our objective is to study the system throughput performance.

In the next section, we highlight some of the main features of TFAA. Then, in Section 6.3, we provide a high-level description of the protocol and present a general system model upon which we will base the performance evaluation. A discussion of the main underlying assumptions is also included. This general system model is complemented by three different reception models: the collision channel, the capture, SINR-based model and the PHY model, which are studied separately in Sections 6.4, 6.5 and 6.6, respectively. In contrast to previous studies, we derive a general expression for the throughput performance of TFAA under the collision channel model which is applicable to both the UNB and narrow-band regime. While our main focus is in the study of the PHY model (since no results on a similar scenario have been published so far), a brief analysis of the classic capture model is also included for completeness.

6.2 Advantages and Issues of TFAA

Narrow-band or ultra narrow-band TFAA offers some characteristics that make of it a suitable radio interface for both terrestrial and satellite LPWAN. Among its main advantages, we may mention the following:

1. ***Relaxed synchronization requirements***. TFAA operates in a fully time-asynchronous manner, which simplifies the terminal design and reduces network signaling overhead. On

the other hand, frequency synchronization requirements are significantly relaxed at the transmitter side, since packets can be transmitted over any random frequency within the system bandwidth. Thus, by avoiding the requirement for accurate oscillators, cheaper radio transmitters may be considered.

2. ***Energy efficiency***. TFAA provides some flexibility in terms of data rate because it does not impose tight channel boundaries. A system designer has the possibility to optimize the terminal energy efficiency by adopting an optimal transmission data rate [2] according to the system specifics, as suggested in [Mia+09]. Note also that cheaper oscillators—that can be used in TFAA as discussed above—are in general less power-consuming [Sad05].

3. ***Robustness to Doppler shifts***. Since the system is already designed to support a high amount of frequency uncertainty, relatively high Doppler shifts should be managed transparently. As a consequence, the need for Doppler shift pre-compensation at the transmitter side may be avoided in most terrestrial applications. However, high Doppler effects—as those experimented in LEO satellite systems in high frequency bands—may entail additional issues in UNB operation, as we discuss below.

4. ***Increased FEC collision resolution capabilities***. Compared to pure ALOHA, in TFAA the MAI can be more easily coped with by the help of low-rate forward error correction (FEC), as we will see on the next chapter. While this aspect was already identified in time-asynchronous schemes [Kis11]—for which the FEC collision resolution capabilities are improved with respect to slotted ALOHA systems—in Section 7.4 we present results that show that this effect is enhanced when frequency-asynchronism is introduced.

5. ***Resilience to power imbalance***. Unlike SS systems—whose throughput drops quickly in the presence of near-far (or power imbalance) conditions [DD12], the throughput performance of TFAA increases in such conditions. This is an important feature in heterogeneous environments with near-far conditions, specially in the absence of tight power control mechanisms.

6. ***Flat fading operation***. In TFAA packets are transmitted in either narrow-band or UNB. It follows that the signal bandwidth is likely to be smaller than the channel coherence bandwidth in most cases.

7. ***Co-channel interference mitigation***. If some sub-bands in W are affected by narrow-band interference, the system capacity can be easily improved by exploiting time/frequency diversity techniques.

[2] Note that in some scenarios the data rate cannot be adjusted, *e.g.* when it is determined by application requirements.

There are, however, several practical issues that must be taken into account, specially if UNB signals are considered. First, the demodulator complexity is increased since there is a higher degree of uncertainty with respect to the center frequency of the incoming packets. Hence, the receiver must be able to scan the whole system bandwidth in order to detect the presence of packets. Second, in the UNB regime, the effects of Doppler rate become more significant, specially when the transmitter-receiver relative velocity is high. As we have seen in Section 2.3.2, the Doppler rate in typical LEO systems can reach about 100 Hz/s in UHF band and over 700 Hz/s in L band. These scenarios are analysed in [Ant+15] for the collision channel model, where it is shown that the Doppler rate has a slight impact in the TFAA MAC throughput. However, this issue may depend heavily on the underlying PHY implementation and hence more in-depth research is still needed. Finally, as we will show in this chapter, TFAA in its simplest form provides a relatively poor throughput performance. In the next chapters, we will show that TFAA can be significantly enhanced by the use of coding and SIC techniques.

6.3 General System Model

In this section we present the basic definitions and assumptions that will be used for the study of TFAA, whichever the reception model adopted. The full available system bandwidth is W, while each message's bandwidth is limited to B[3], as sketched in Figure 6.1. In general, $W = kB$, where $k = 1$ for pure ALOHA and $k > 1$ for TFAA. The center frequency and the arrival time of some packet i, as observed by the receiver, are noted f_i and t_i, respectively. Packets are transmitted at any frequency within the continuous interval $I_f = [f_m, f_M]$, where $f_m = -B([W/B] - 1)/2$ and $f_M = B([W/B] - 1)/2$. Note that if the W/B ratio is set to 1, then $f_m = f_M = 0$ and all packets are transmitted over the center frequency $f = 0$. In this case the protocol operates exactly as ALOHA. The main assumptions considered throughout this chapter are summarized as follows:

1. Poisson arrivals: as in the ALOHA model previously described, we consider an infinite population RA system with Poisson-distributed packet arrivals where each new arriving packet is assumed to be generated at a new terminal. The average number of packets generated by all the terminals in the system during the packet duration T_p is denoted by G_a. Since in TFAA several packets can be decoded simultaneously over the entire band W, G_a can be much greater than 1. Consequently, we define the *normalized channel load* G as the average number of packets that are transmitted over the time-frequency space required by a single message.

2. Ideal time-limited and bandwidth-limited transmission: each message's energy is fully contained within a time-frequency surface of bandwidth B and time length T_p, *i.e.* out-of-band

[3]Here, the bandwidth is defined as the full frequency support of a signal.

emissions and inter-symbol signal components are assumed to be negligible.

3. Random transmission frequency: in practice, and according to the system specifics, f_i can be affected by Doppler effects or oscillator instabilities. We account for this uncertainty by considering f_i as a random variable (rv). Further, we will consider that f_i follows a continuous uniform distribution with support on the interval I_f, *i.e.* $f_i \sim \mathcal{U}[f_m, f_M]$.

4. Packets are transmitted at constant power. Regardless of the underlying propagation environment, the received symbol energy is constant for all the symbols within a given packet. By contrast, different packets may be received at different power levels—this case will be considered in Section 6.6.

5. Three different reception models, based on those described in Section 4.3, are taken into account:

   (a) *The collision channel model.* In the original pure ALOHA system [Abr70] a test packet i is successfully transmitted if and only if no other packet arrival occurs within the interval $[t_i - T_p, t_i + T_p]$. Here, we extend this model to account for the frequency separation of concurrent packet transmissions. Following Assumption 2), we will consider that only the packets transmitted within a frequency distance B can interfere with the test packet i. Thus, packet i will be successfully decoded by the receiver if and only if no other packet arrival occurs over the two-dimensional plane $[f_i - B, f_i + B] \times [t_i - T_p, t_i + T_p]$.

   (b) *The capture, SINR-based model.* Here we consider also a PHY-independent model in which the collisions are not necessary destructive. We assume that the overall MAI is proportional to the sum of the overlap areas between the test packet and the interfering packets. Thus, a packet transmission is considered if the *average* signal-to-interference plus noise ratio is above a given threshold b.

   (c) *The PHY model.* In this case the performance is obtained by taking into account the underlying transmission parameters. We will see that the so-called Gaussian assumption—which is usually employed in the analysis of RA protocols with capture—cannot be considered here since the interference amplitude may vary significantly between consecutive symbols and also depends on the frequency distance of the interfering packets. Assuming that all packets can be perfectly acquired by the receiver (that is, under perfect preamble detection and signal parameter estimation), a given test packet will be correctly transmitted depending on the actual MAI plus noise level over each transmitted symbol.

Assumption 1) has been widely employed in the related literature, becoming a standard model for the analysis of random access models. Thus, we adopt it in order to facilitate comparison

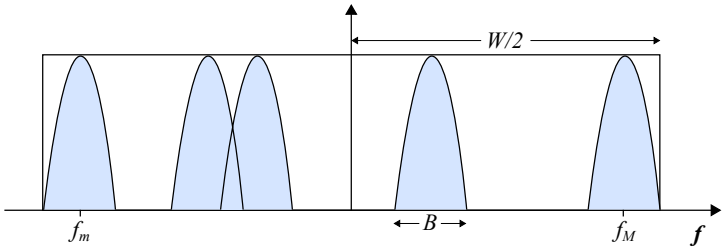


Figure 6.1 – Baseband spectrum representation of TFAA.

between our results and those reported for other relevant RA protocols. However, the suitability of the Poisson arrival model for LPWAN and M2M applications is arguable, since in this context it is sometimes common to observe correlated traffic patterns (for instance, several terminals in a network may be programmed to send report messages on a periodic basis). Also, it does not describes well high traffic burstiness.

Assumption 2) is essential for throughput evaluation. By considering that the packet energy is band-limited to B, only those packets transmitted at frequency distances smaller or equal to B may affect a given test packet. While in practice the energy of a time-limited signal has infinite frequency contributions, the use of pulse-shaping filters allows to concentrate most of the signal energy within a low-pass bandwidth equal to $(1 + \beta)/2T$, where β is the roll-off factor and T the symbol period.

Having uniformly distributed center frequencies as considered in Assumption 3) represents a best-case scenario. In practice this can be approached by assigning different frequencies to the terminals in the network. Alternatively, in scenarios with high and variable Doppler shifts as LEO systems, each terminal may employ its own fixed transmission frequency, which has a similar effect at the receiver side.

Assumption 4 is reasonable since TFAA is in principle a narrow-band protocol. Therefore, in most applications the fading envelope will vary slowly.

The collision channel model proposed in Assumption 5a) is indeed unrealistic but it facilitates significantly the analysis and therefore provides valuable insight. Besides, it allow us to obtain a straightforward lower-bound of the system performance. In general, it is reasonable to expect that partly overlapped packets can also be recovered. This case, that we consider in Assumptions 5b) and 5c), is however much more complex to describe analytically. As the collision channel, the former has the advantage of being independent of the PHY layer implementation, while at the same time provides a more closer description of the actual reception process. Since this model is based on the total interference power during the packet transmission, it can be particularly useful in systems employing long packets and interleaving, in which case the MAI power can be more easily be counteracted with FEC.

In Assumption 5c) we have considered an ideal receiver with perfect preamble detection and

signal parameter estimation. This is justified by the fact that in TFAA the event of having packets fully overlapped in both time and frequency domains occurs with low probability, specially under practical traffic loads. The preamble collision probability is therefore reduced with respect to ALOHA and SA. More details regarding packet reception modelling have been provided in Section 4.3.2.

6.4 Throughput Analysis Under the Collision Channel Model

In this section, we derive the throughput performance of TFAA under the collision channel model described above. We define the *aggregate throughput* S_a as the average number of successful packet transmissions over the whole system bandwidth W during the packet duration T_p. The throughput components due to successful transmissions over the frequency intervals $I_+ = [0, f_M]$ and $I_- = [f_m, 0)$ are noted S_+ and S_-, respectively. Thus, we may write $S_a = S_+ + S_-$. By symmetry, and by observing that the throughput contribution from packets with transmission frequency $f = 0$ is negligible, we have

$$S_a = 2S_+. \tag{6.1}$$

Now we focus on determining S_+. We start by considering the traffic load over a differential frequency band of width df, that is, the fraction of packet arrivals whose transmission frequency is in df. Since $f_i \sim \mathcal{U}[f_m, f_M]$ for any packet i, packet transmissions in df occur with probability $(W - B)^{-1}df$. It follows that the traffic over the differential df is also Poisson with rate

$$\frac{\lambda}{W - B}df, \tag{6.2}$$

where the term $\frac{\lambda}{W-B}$ is here defined as the *traffic density* and is expressed in packets/s/Hz. Denoting by P_s the probability of successful packet transmission, the throughput contribution dS from the traffic carried over a differential df during T_p may be expressed as [Abr77]

$$dS = \frac{\lambda T_p}{W - B}df \cdot P_s. \tag{6.3}$$

Now we need to derive P_s. Since packets transmitted close to the boundaries of the frequency band are less exposed to packet collisions, P_s is in general a function of the transmission frequency. Hence, we have $P_s = P_s(f)$. It is easy to note from Figure 6.2 that the surface of the vulnerability region for a packet transmitted at (t, f), with $f \in I_+$, is given by

$$\Lambda(f) = \begin{cases} 4T_pB, & \text{for } 0 \leq f \leq f_M - B \\ 2T_p(f_M - f + B), & \text{for } f_M - B < f \leq f_M. \end{cases} \tag{6.4}$$

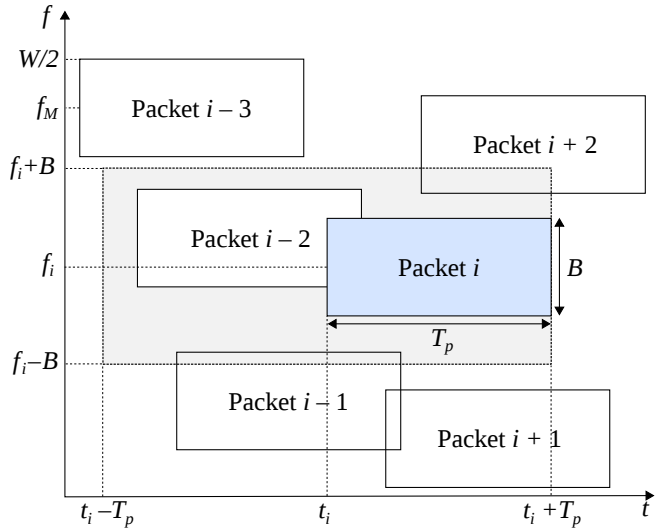


Figure 6.2 – A two-dimensional representation of TFAA. The gray-shaded area corresponds to the set of points (t, f) over which any transmitted packet will interfere with packet i. For the collision channel model, the overlap of packets i and $i-2$ results in both packets being lost.

It follows that K, the number of packets arriving over $\Lambda(f)$, obeys a Poisson distribution whose intensity is obtained by multiplying $\Lambda(f)$ by the traffic density:

$$P(K=k) = \frac{\left[-\lambda\Lambda(f)/(W-B)\right]^k}{k!} \exp\left[\frac{-\lambda\Lambda(f)}{W-B}\right]. \tag{6.5}$$

Hence, according to our definition of the collision channel, $P_s(f)$ is simply the probability that no packet arrives over $\Lambda(f)$:

$$\begin{aligned} P_s(f) &= P(K=0) \\ &= \exp\left[\frac{-\lambda\Lambda(f)}{W-B}\right]. \end{aligned} \tag{6.6}$$

Now we are ready to compute S_+ by simply summing all the throughput contributions over the positive frequency interval I_+, that is

$$S_+ = \int_{I_+} dS \tag{6.7}$$

By replacing the result of the integral in Eq. (6.1), we obtain the aggregate throughput of TFAA

$$
\begin{aligned}
S_a &= 2S_+ \\
&= \frac{\lambda T_p(W-3B)}{W-B}\exp\left[\frac{-4B\lambda T_p}{W-B}\right] \\
&\quad + \exp\left[\frac{-2B\lambda T_p}{W-B}\right]\left(1-\exp\left[\frac{-2B\lambda T_p}{W-B}\right]\right).
\end{aligned}
\tag{6.8}
$$

It is now convenient to derive the *normalized throughput* $S(G) \triangleq S_a \times \frac{B}{W}$ as a function of the *normalized channel load* $G \triangleq G_a \times \frac{B}{W}$. Multiplying Eq. (6.8) by $\frac{B}{W}$ yields

$$
\begin{aligned}
S(G) = \frac{G(W-3B)}{W-B}\exp\left[\frac{-4GW}{W-B}\right] \\
+ \frac{B}{W}\exp\left[\frac{-2GW}{W-B}\right]\left(1-\exp\left[\frac{-2GW}{W-B}\right]\right).
\end{aligned}
\tag{6.9}
$$

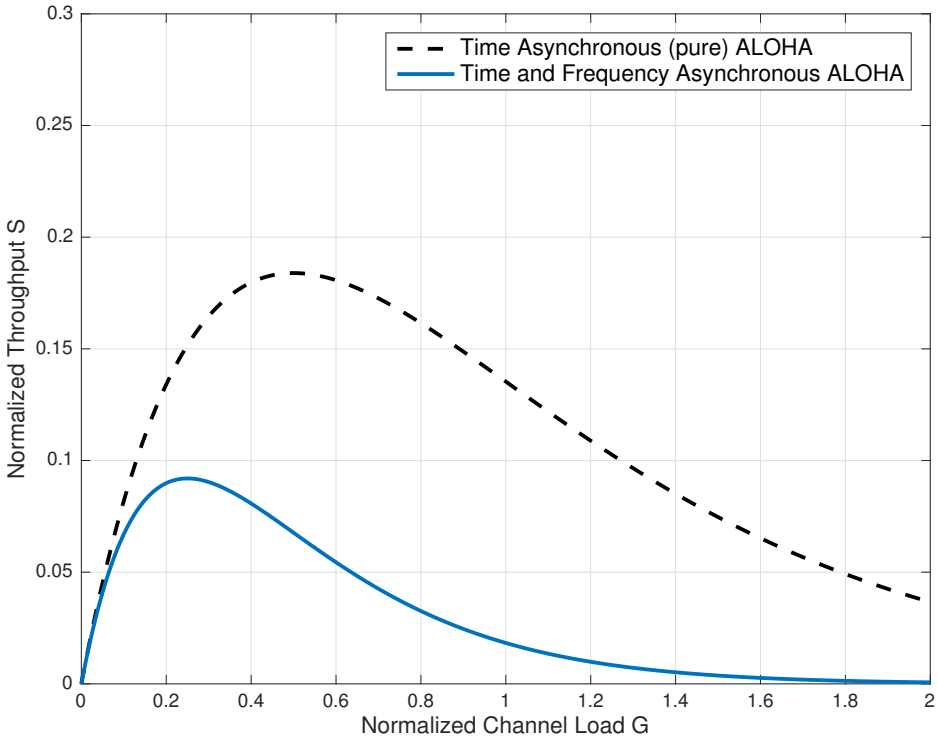
Finally, we observe that, as $B/W \to 0$ (that is, if B is very small compared to W), Eq. (6.9) converges quickly to

$$
S = G \cdot \exp\left[-4G\right]. \tag{6.10}
$$

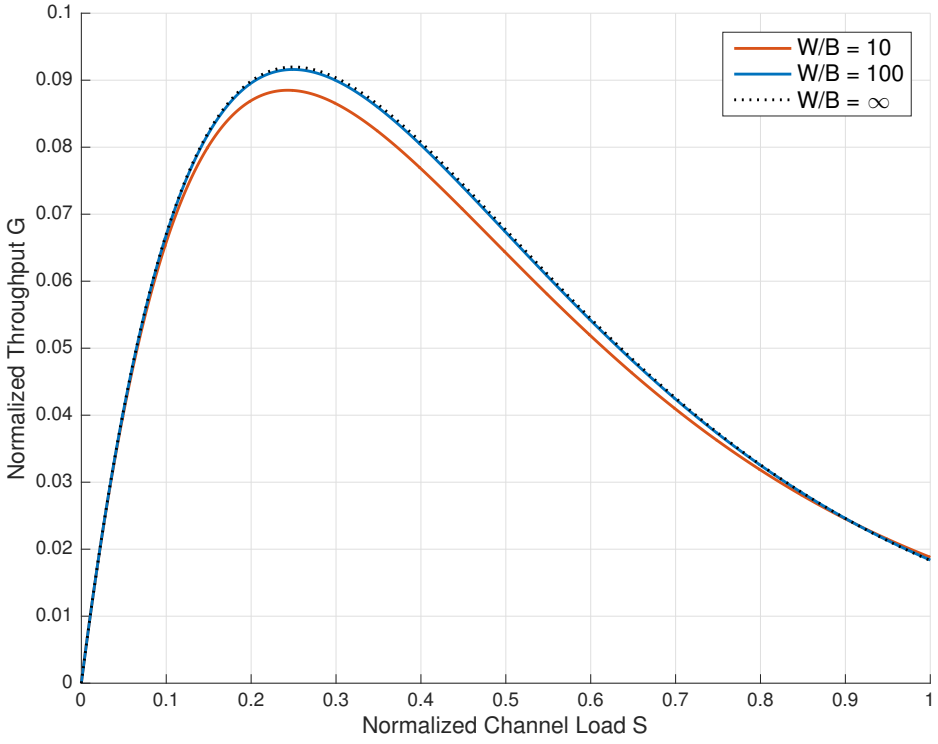
Even though the collision channel model may bring too pessimistic results, it gives interesting insight. In particular, it can be noted that the normalized throughput does not depend on any design parameter (as the packet size or the bandwidth) as long as $B << W$. Also, we observe that the maximum throughput is achieved at $G^* = 1/4$ and equals $S(G^*) = 1/(4e)$. Figure 6.3a compares the normalized throughput of ALOHA with that of TFAA given by Eq. (6.10). In Figure 6.3b we have plotted the TFAA throughput as given by Eq. (6.9) for $W/B = 10, 100$ and ∞. As one would expect intuitively, it is clear that the curves converge quickly with the W/B ratio. Therefore, the simple formula given by Eq. (6.10) approximates well the TFAA throughput for virtually any value of the W/B ratio.

6.5 Throughput Analysis Under the Capture, SINR-based Reception Model

In the following section we will derive the TFAA throughput based on the capture, SINR-based reception model introduced in Section 6.3. While it is reasonable to expect that this model describes better the underlying reception mechanism, and that may therefore provide more accurate



(a) Throughput comparison between TFAA and ALOHA



(b) TFAA throughput for different W/B ratios.

Figure 6.3 – Normalized throughput versus MAC load of TFAA under the collision channel model.

performance results, its analysis is somewhat more complex. A particular scenario based on this model has been studied in [Li+16], where an analytical framework based on stochastic geometry is proposed. The authors provide a closed-form solution for the MAI distribution for the 1-interferer case ($K = 1$), while for the general case ($K > 1$) only a numerical approach is proposed. Also, in [Li+16] the focus is given on some outage probabilities which are related to the geometrical configuration of the interference. Here, we will briefly develop a simple semi-analytic method to obtain the throughput performance considering a Poisson traffic model, and will use these results as an additional benchmarking reference.

The variables and definitions given in the previous section are also applicable here and will be used throughout this section. In order to simplify the analysis, we focus on the reception of a test packet, which we denote by the index 0, that is affected by the interference of K packets. We assume that the test packet is located at some frequency $f_0 \in [f_m, f_M]$ and that the packet bandwidth is very small compared to the system bandwidth W (*i.e.* $B/W \to 0$). This assumption is supported by the results obtained for the collision channel, where we showed that there is little difference between the asymptotic throughput, as given by Eq. (6.10), and the exact throughput, as given by Eq. (6.9), even for W/B ratios as low as 10. With this assumption, the vulnerability region for the test packet is independent of the frequency location and equals $\Lambda = 4T_pB$ (as shown in Figure 6.2 for packet i), while the the traffic density is approximately λ/W packets/s/Hz. Hence, the number of interfering packets K follows a Poisson distribution of intensity

$$4T_pB\frac{\lambda}{W} \simeq 4G, \tag{6.11}$$

where G is the normalized channel load defined in the previous section. The power of the test packet and the K interfering packets at the receiver input is denoted P_k ($0 \leq k \leq K$).

We define the signal-to-interference-plus-noise (SINR) ratio for the test packet as

$$\gamma = \frac{P_0 T}{I + N_0} \tag{6.12}$$

where T is the symbol period, N_0 the thermal noise density and I the total interference energy from the K overlapping packets. In the reception model adopted in this section, a packet is assumed to be successfully transmitted if its SINR is higher than some capture threshold, *i.e.* if $\gamma > b$, for some $b > 0$. Note that we only admit positive values of b since TFAA is by nature a non-spread spectrum system. In order to determine the interference power I_k due to a colliding packet k, we assume that the overlap level is independent between both time and frequency domains. We quantify the overlap level by means of two random variables: α_k for the time-domain and β_k for the frequency domain. These variables are assumed to be distributed over the interval $[0, 1]$, so that for a fully synchronized packet we assign an overlapping factor of 1, and a value 0 for a non-colliding packet. In the example given in Figure 6.4, we have $\alpha_k = 0.5$ and $\beta_k = 0.6$. Further,

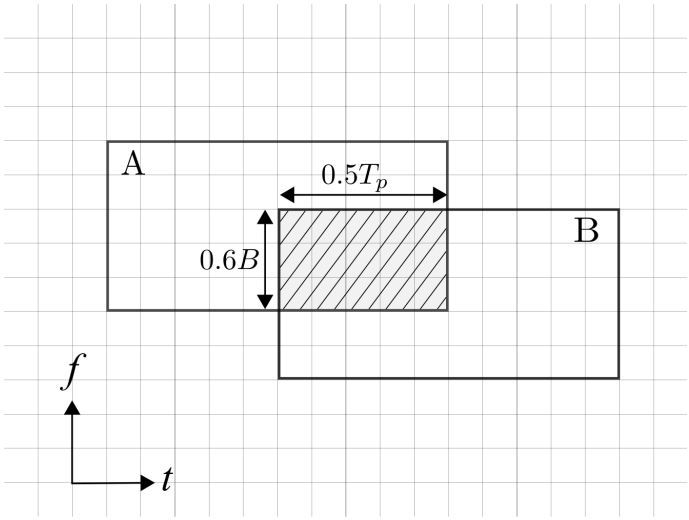


Figure 6.4 – A collision between packets A and B.

we assume that α_k and β_k are independent, identically distributed as $\mathcal{U}[0,1]$. The total amount of interference contributed by packet k is therefore $I_k = \alpha_k \beta_k P_k T$, and the total interference energy is given by

$$I = \sum_{k=1}^{K} \alpha_k \beta_k P_k T. \tag{6.13}$$

When the received powers P_k's are equal, that is, for $P_0 = P_1 = ... = P_K = P$, Eq. (6.13) becomes $I = PT \sum_{k=1}^{K} \alpha_k \beta_k$. For this case, we define the random variable $\Theta_K = \sum_{k=1}^{K} \alpha_k \beta_k$ so that we may write the conditional packet capture probability as

$$\begin{aligned} P_{c|K} &= P\left(\frac{PT}{\Theta_K PT + N_0} > b\right) \\ &= P\left(\Theta_K \leq \frac{1}{b} - \frac{N_0}{PT}\right) \\ &= F_{\Theta_K}(1/b - N_0/PT; K), \quad \text{for } K > 0. \end{aligned} \tag{6.14}$$

In Eq. (6.14), $F_{\Theta_K}(\theta; K)$ is the cumulative distribution function (cdf) of Θ_K with parameter K. When $K = 0$, the capture probability is 1, so in general we have

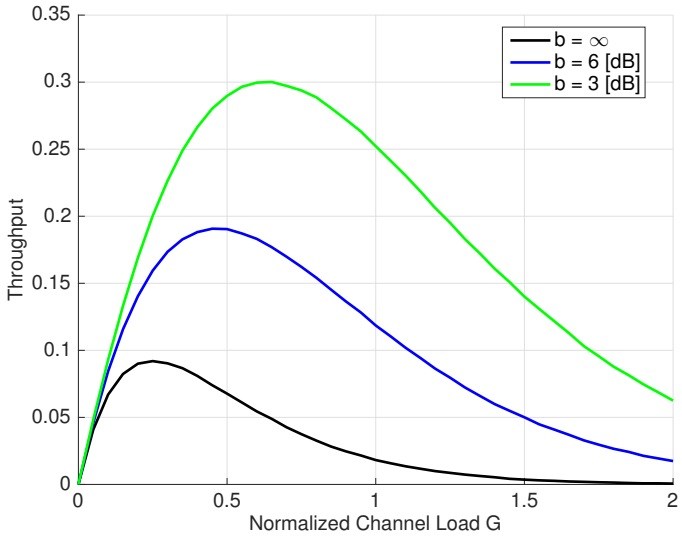


Figure 6.5 – Normalized throughput versus MAC load for TFAA with different capture thresholds.

$$P_{c|K} = \begin{cases} 1, & \text{if } K = 0, \\ F_{\Theta_K}(1/b - N_0/PT; K), & \text{if } K > 0. \end{cases} \tag{6.15}$$

The normalized throughput for a capture reception model can be defined as [ZP92]

$$T(G) = G \sum_{k=0}^{\infty} P_{c|K} P(K = k; G), \tag{6.16}$$

where $P(K = k; G)$ is given by

$$P(K = k; G) = \frac{(4G)^k \exp -4G}{k!}. \tag{6.17}$$

As in [Li+16], we may find $F_{\Theta_K}(\theta; K)$ in closed-form for the case $K = 1$, while obtaining a general expression for any K does not seem feasible. Therefore, we obtain the throughput as given by Eq. (6.16) by first computing Eq. (6.16) using the Monte Carlo method. Figure 6.5 shows the throughput for capture thresholds $b = 2, 4$ and ∞. Note that in the latter case, when $b \to \infty$, $P_{c|K} = F_{\Theta_K}(-N_0/PT; K) = 0$ for all $K > 0$, and Eq. (6.16) converges to $T(G) = G \exp -4G$, which is equivalent to the result obtained for the collision channel.

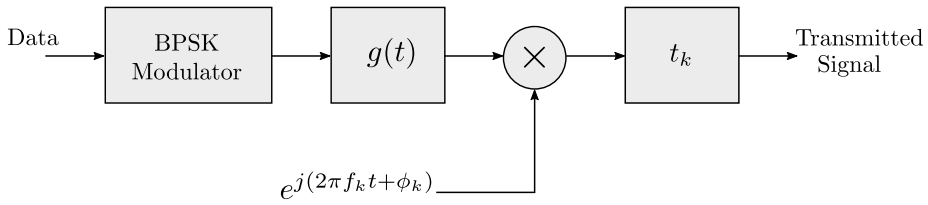


Figure 6.6 – Simplified TFAA transmitter block diagram.

6.6 Throughput Analysis Under the PHY Channel Model

In this section, we study the throughput performance of TFAA by also taking into account the transmission waveform parameters and the capture model described in Section 6.3. We assume that all packets are transmitted using uncoded binary phase shift keying (BPSK). This modulation scheme has been chosen because of its simplicity but also because it is of practical interest, since it has been adopted by Sigfox systems. After modulation, a pulse shaping filter $g(t)$ of unit energy is applied. The receiver follows a conventional, single-user matched filter architecture. We focus on the reception of a test packet, denoted as packet 0, which is subject to time-varying interference from a set $\mathcal{C} = \{1, 2, ..., K\}$ of K colliding packets having different frequency offsets and relative delays with respect to the test packet. We assume that all packets are of equal length and contain a fixed number of L_p information symbols.

A simplified transmitter diagram is presented in Figure 6.6. Along with the arrival time t_k and the transmission frequency f_k, each packet k $(0 \leq k \leq K)$ is characterized by an initial random phase ϕ_k (at $t = 0$) and an amplitude A_k. The amplitudes $\{A_k\}$ are constant during the packet transmission time T_p and are independent, identically distributed according to some given distribution. For the test packet we assume perfect parameter estimation, and, without loss of generality, we set $(t_0, f_0, \phi_0) = (0, 0, 0)$. As in the previous section, we consider the case where $B << W$, so that the number of interfering packets K follows a Poisson distribution with intensity $\simeq 4G$, where G is the normalized load. Hence, for $k \in \mathcal{C}$, we have $t_k \sim \mathcal{U}[-T_p, T_p]$, $f_k \sim \mathcal{U}[-B, B]$ and $\phi_k \sim \mathcal{U}[0, 2\pi]$. Assuming that the system is also affected by AWGN, the baseband equivalent of the received signal can be expressed as

$$r(t) = s(t) + I(t) + n(t), \quad 0 \leq t \leq T_p; \tag{6.18}$$

where $n(t)$ is a complex noise signal and $s(t)$, $I(t)$ are the signals corresponding to the test packet and the interfering packets, respectively, and may be expressed as

$$
\begin{aligned}
s(t) &= A_0 \sum_{l=0}^{L_p-1} b_0(l) g(t - lT), \\
I(t) &= \sum_{k=1}^{K} \sum_{l=0}^{L_p-1} A_k b_k(l) g(t - lT - t_k) e^{j(2\pi f_k t + \phi_k)}.
\end{aligned}
\tag{6.19}
$$

Here, the sequences $\{b_k(l),\ 0 \leq l < L_p\}$ take random, equiprobable values in $\{\pm 1\}$ and thus represent the BPSK symbols corresponding to packet k; and T is the symbol duration.

6.6.1 Multiple Access Interference Analysis

As shown in the receiver model depicted in Figure 6.7, the decision variable z_l for the l-th symbol of the test packet is taken from the output $z(t)$ of the matched filter $f(t) = g(-t)$ at the sampling instant $t = lT$. Hence, we have

$$
\begin{aligned}
z_l &\triangleq z(t = lT) \\
&= \int_{-\infty}^{\infty} \Re\{r(\alpha) f(lT - \alpha)\} d\alpha, \\
&= A_0 b_0(l) + \sum_{k=1}^{K} I_k(l) + n_l,
\end{aligned}
\tag{6.20}
$$

where $\Re\{x\}$ is the real part of x, n_l is a noise term of variance $N_0/2$ and $I_k(l)$ is the individual interference contribution from each of the K packets in $\mathcal{C}$. Thus, the term

$$
\omega_K(l) = \sum_{k=1}^{K} I_k(l) \tag{6.21}
$$

is a random variable representing the MAI from the set of K colliding packets over the l-th symbol of the test packet.

Let us now focus on the term $I_k(l)$, which comes from the convolution

$$
I_k(l) = A_k \int_{-\infty}^{\infty} \Re\{ \sum_{i=0}^{L_p-1} b_k(i) g(\alpha - iT - t_k) g(-[lT - \alpha]) e^{j(2\pi f_k \alpha + \phi_k)} \} d\alpha \tag{6.22}
$$

Obtaining an analytical solution of Eq. (6.22) is in general a cumbersome task. However, when rectangular pulse shaping is considered, Eq. (6.22) is greatly simplified (see [Axf92], where

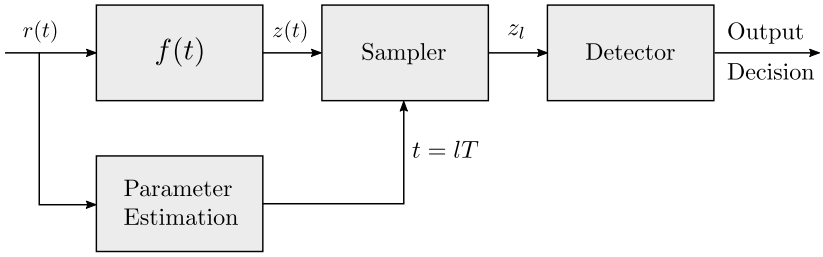


Figure 6.7 – Simplified matched-filter receiver model.

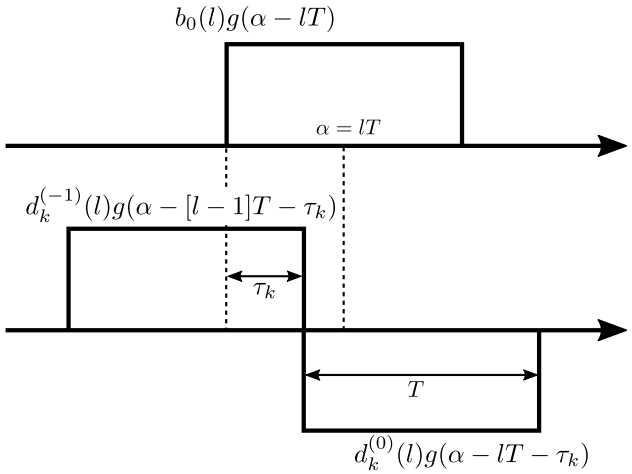


Figure 6.8 – A representation of the convolution given by Eq. (6.24)

a similar system is studied under the continuous transmission regime). We will shortly review this configuration next. We assume that the filter $g(t)$ has a rectangular shape and a duration T, as shown in Figure 6.8. We observe that, since the packets are transmitted asynchronously, a given symbol of the test packet is, in most cases, affected by two consecutive symbols from each interfering signal. In other words, a symbol $b_0(l)$ can be interfered by some pair of symbols $(b_1(m), b_1(m-1))$ from packet 1, another pair $(b_2(n), b_2(n-1))$ from packet 2, and so on. We denote these two symbols by $d_k^{(-1)}(l)$ and $d_k^{(0)}(l)$, and note that their values will depend on the arrival time t_k. For instance, if t_k is an integer multiple of T, then the signal k is perfectly synchronised (at the symbol level) with the test packet, in which case we assume $d_k^{(-1)}(l) = 0$. On the other hand, if the interfering signal is just starting (or ending) at $t = lT$, then either $d_{k,l}^{(-1)} = 0$ or $d_{k,l}^{(0)} = 0$. Similarly, if transmission of packet k started after $[lT, (l+1)T]$ or ended before this interval, then $d_{k,l}^{(-1)} = d_{k,l}^{(0)} = 0$.

In addition, we define the timing offset between the test packet and packet k as

$$\tau_k = \begin{cases} \mathrm{mod}(t_k, T), & \text{if } t_k \geq 0 \\ T - \mathrm{mod}(|t_k|, T), & \text{otherwise.} \end{cases} \tag{6.23}$$

We may now express $I_k(l)$ as

$$\begin{aligned} I_k(l) &= A_k \int_{-\infty}^{\infty} \Re\{d_k^{(-1)}(l) g(\alpha - [l-1]T - \tau_k) g(-[lT - \alpha]) \\ &\qquad + d_k^{(0)}(l) g(\alpha - lT - \tau_k) g(-[lT - \alpha])\} e^{[2\pi f_k \alpha + \phi_k]} d\alpha \\ &= A_k \int_{T(l-1/2)}^{T(l-1/2)+\tau_k} d_k^{(-1)}(l) \cos[2\pi f_k \alpha + \phi_k] d\alpha \\ &\qquad + A_k \int_{T(l-1/2)+\tau_k}^{T(l+1/2)} d_k^{(0)}(l) \cos[2\pi f_k \alpha + \phi_k] d\alpha \\ &= A_k \left[d_k^{(-1)}(l) I_k^{(-1)}(l) + d_k^{(0)}(l) I_k^{(0)}(l) \right], \end{aligned} \tag{6.24}$$

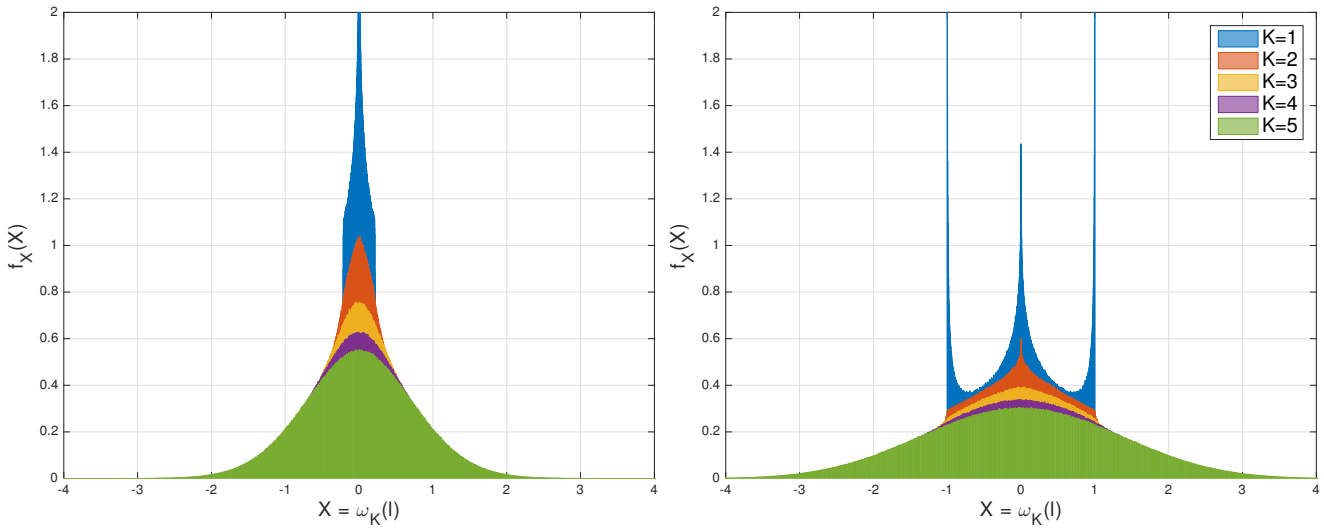
where the terms $I_k^{(-1)}(l)$ and $I_k^{(0)}(l)$ are given by

$$\begin{aligned} I_k^{(-1)}(l) &= \left[\frac{\sin[2\pi f_k (T(l-1/2) + \tau_k) + \phi_k] - \sin[2\pi f_k (T(l-1/2)) + \phi_k]}{2\pi f_k} \right] \\ I_k^{(0)}(l) &= \left[\frac{\sin[2\pi f_k (T(l+1/2)) + \phi_k] - \sin[2\pi f_k (T(l-1/2) + \tau_k) + \phi_k]}{2\pi f_k} \right] \end{aligned} \tag{6.25}$$

It is clear from the above equations that the interference level is not independent from symbol to symbol. Indeed, the discretized MAI signal $\omega_K(l)$ is a stochastic process that depends on the set of parameters $\{t_k, f_k, \phi_k, A_k, 1 \leq k \leq K\}$ and the data sequences $b_k(l)$'s. When the interference is time-asynchronous only, that is when $f_k = 0$ for all $k \in \mathcal{C}$, the term $I_k(l)$ reduces to

$$I_k(l) = \cos\phi_k \left[d_k^{(-1)}(l) m(\tau_k - T) + d_k^{(0)}(l) m(\tau_k) \right], \tag{6.26}$$

where $m(t) = g(t) * g(-t)$, and $*$ denotes convolution. It is worth noting that this scenario has been object of several investigations aiming to describe the effects of co-channel interference. In particular, it has been possible to determine the distribution of $I_k(l)$ analytically for rectangular pulse shaping and typical linear modulation schemes [TV99; Chi97; Chi; GC05]. However, the general case in which the frequency offsets f_k's are not null, does not seem to admit an analytical solution. Figure 6.9 shows the probability density function (pdf) of the random variable $\omega_K(l)$ for both the time- and frequency-asynchronous (TFA) system (Figure 6.9a) and the time-asynchronous



(a) Time- and frequency asynchronous system.

(b) Time-asynchronous system.

Figure 6.9 – pdf of the MAI signal $\omega_k(l)$ for $K = 1, 2, ..., 5$.

(TA) system (Figure 6.9b), and for $K = 1, 2, 3, 4$ and 5. In the former, the pdf is obtained by means of Monte Carlo simulation, while for the latter we used the expressions given in [Chi97]. As one might have expected, the variances of $\omega_k(l)$ for the TFA system are always smaller than those of the TA system. Also, we observe that $\omega_K(l)$ converges faster to a Gaussian distribution in the TFA system. However, this fact has little relevance in terms of throughput evaluation since the bit errors are not independent and hence the so-called Gaussian assumption has limited validity.

So far, our results already reveal us some information about the throughput performance of TFAA. On the one hand, it is already obvious that, for a given channel load, TFAA packets are more exposed to collisions than ALOHA packets, since the vulnerability area is increased in the former. On the other hand, we conclude from Figure 6.9 that in TFAA each collision produces, in average, less interference. This behaviour suggests us that the performance of TFAA may be significantly improved if some amount of interference resilience is added to the system, which can be accomplished by trading bandwidth for error correction capabilities.

6.6.2 Bit Error Probability

Before studying the packet error rate and, consequently, the throughput performance of TFAA, we analyse in this section the error probability at the *bit* level, which we refer to as the bit error rate (BER). To do so, we focus on a random symbol l of the test packet. If we assume that the interference magnitude $\omega_K(l)$ is known, and that a symbol $b_0(l) = -1$ has been transmitted, then

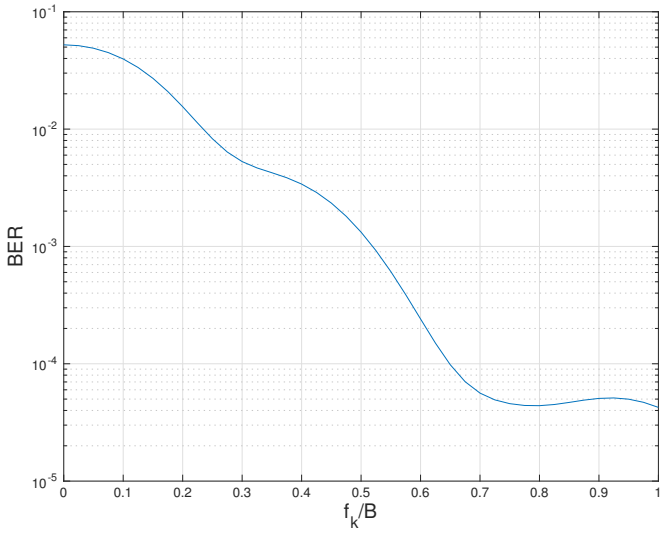


Figure 6.10 – BER as a function of the normalized frequency distance f_k/B. $K = 1$, $E_b/N_0 = 10$ [dB].

the decision variable z_l will follow a Gaussian distribution with mean $-A_0 + \omega_K(l)$ and variance N_0. The conditional bit error rate is therefore defined as the probability of observing $z_l > 0$ given $b_0(l) = -1$ and $\omega_K(l)$:

$$\begin{aligned} \mathrm{BER}[l; K, \omega_K(l)] &\triangleq P\left(z_l > 0 | K, \omega_K(l)\right) \\ &= Q\left(\frac{A_0 - \omega_K(l)}{\sqrt{N_0}}\right), \end{aligned} \tag{6.27}$$

where $Q(\cdot)$ is the standard Q-function defined as

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left[\frac{-u^2}{2}\right] du \tag{6.28}$$

Note that this approach differs from the standard Gaussian assumption, in which case the interference is treated as Gaussian noise and is added to the denominator of the argument of the Q-function.

As an example, we consider a continuous transmission regime[4] in which $K = 1$ (*i.e.* there is only one interfering signal) and compute the BER over an arbitrary bit l as a function of the frequency distance f_k, assuming that both the desired and the interfering signals have equal power

[4]Note that in a continuous transmission regime the terms $d_k^{(-1)}(l)$ and $d_k^{(0)}(l)$ are always non-zero.

($A_0 = A_k = 1$) and that the parameters (τ_k, ϕ_k) are random. Dropping the dependence on l, Eq. (6.27) becomes thus [Axf92]

$$\begin{aligned} \mathrm{BER}[f_k; K=1, \omega_1] &= \mathbb{E}\left[Q\left(\frac{A_0-\omega_1}{\sqrt{N_0}}\right)\right] \\ &= \sum_{\left(d_k^{(-1)}, d_k^{(0)}\right)\in\{-1,1\}^2} P(d_k^{(-1)}, d_k^{(0)}) \times \\ &\qquad \mathbb{E}_{\tau_k}\left[\mathbb{E}_{\phi_k|\tau_k}\left[Q\left(\frac{A_0-\omega_1}{\sqrt{N_0}}\right)\right]\right] \end{aligned} \tag{6.29}$$

where $\omega_1 = A_k\left[d_k^{(-1)} I_k^{(-1)} + d_k^{(0)} I_k^{(0)}\right]$, the terms $I_k^{(-1)}$ and $I_k^{(0)}$ are given by Eq. (6.25) and the sum is over the set of combinations

$$\{d_k^{(-1)}, d_k^{(0)}\} = \{(-1,-1), (-1,1), (1,-1), (1,1)\}. \tag{6.30}$$

Also, since the binary data sequences are equiprobable, $P(d_k^{(-1)}, d_k^{(0)}) = 1/4$. By computing Eq. (6.29) numerically for $E_b/N_0 = 10$ [dB], we obtain the curve shown in Figure 6.10, where the x-axis is normalized with respect to the signal bandwidth $B = 2/T$,[5]. In particular, we observe that the BER already drops below 10^{-4} for $f_k > 0.65B$. These results confirm the hypothesis that colliding packets at $f_k > B$ can be neglected in the throughput calculation.

6.6.3 Computation of the Exact Packet Error Probability

Our interest now is to compute the packet error probability by taking into account the complete interference process $\boldsymbol{\omega_K} = [\omega_K(0), \omega_K(1), ..., \omega_K(L_p - 1)]^t$, and the error probability at each transmitted symbol of the test packet. Assuming, without loss of generality, $\boldsymbol{b}_0 = [-1, -1, ..., -1]^t$, *i.e.*, the test packet corresponds to the all-zero data packet, the conditional symbol error probability for the l-th symbol can be written

$$\begin{aligned} \mathrm{BER}[l; K, \boldsymbol{\omega_K}] &\triangleq P(z_l > 0 | K, \boldsymbol{\omega_K}) \\ &= Q\left(\frac{A_0 - \omega_K(l)}{\sqrt{N_0}}\right), \end{aligned} \tag{6.31}$$

Now, the conditional packet error probability $P_{e|K,\boldsymbol{\omega_K}}$ can be evaluated as follows:

[5]We assume that the packet bandwidth is contained inside the first null of the side lobes. Since we are considering a rectangular pulse shaping filter, this occurs at $\pm 1/T$.

$$P_{e|K,\boldsymbol{\omega_K}} = 1 - \prod_{l=0}^{L_p-1} (1 - \text{BER}[l; K, \boldsymbol{\omega_K}]). \tag{6.32}$$

Note that the random vector $\boldsymbol{\omega_K}$ depends on multiple random variables, namely the transmitted data sequences $\boldsymbol{b}_k = [b_k(0), b_k(1), ..., b_k(L_p - 1)]^t$ and the set of signal parameters $\{t_k, f_k, \phi_k, A_k\}$. Therefore, in order to determine the average packet error rate $P_{e|K}$ given K colliding packets, we must average (6.32) over all possible interference realizations, that is

$$P_{e|K} = \mathbb{E}_{\boldsymbol{\omega_K}} \left[P_{e|K,\boldsymbol{\omega_K}} \right], \tag{6.33}$$

where $\mathbb{E}_{\chi}[\cdot]$ denotes expectation with respect to the random variable χ. Since the joint distribution $P(\omega_K[0], \omega_K[1], ..., \omega_K[L_p - 1])$ is not known, we compute Eq. (6.33) numerically using the Monte Carlo integration method.

Now, since packet arrivals are Poisson distributed, the average packet error rate (PER) is given by

$$P_e = \sum_{K=0}^{\infty} P_{e|K} \frac{(4G)^K}{K!} \exp(-4G). \tag{6.34}$$

Finally, the throughput can be readily obtained according to the definition

$$T(G) \triangleq G \times (1 - P_e). \tag{6.35}$$

6.6.4 Numerical Results

Next we present numerical results for the PHY model studied in this section. The exact performances were obtained by first evaluating $P_{e|K}$ using the Monte Carlo method, and then computing (6.34) and (6.35). We refer to this method as the semi-analytic method. Additionally, we have tested the protocol performance by means of a physical layer simulator implementing the full transmission chain of our model. For both the simulation and semi-analytic methods, the packet length is $L_p = 120$ bits and we have considered a rectangular shaping filter. Hence, the packet bandwidth is assumed to be equal to $B = 2R_s$, where R_s is the symbol rate.

The simulation approach follows closely the model described at the beginning of this section: for each realization, a test packet is first generated and centred at $(t, f) = (0, 0)$. Then, since we are assuming $B << W$, a random number of interfering packets is generated following a Poisson process of intensity $4G/\eta$, where η is the spectral efficiency and is defined as $\eta = R_b/R_s$[6],

[6]Here, as in the related literature, the pulse shaping filter excess bandwidth is not included in the calculation of the normalized throughput.

R_b and R_s being the information bit rate and the symbol rate, respectively. Thus, for uncoded BPSK we have $\eta = 1$ [b/s/Hz]. These packets are located in random positions within the surface $[-T_p, T_p] \times [-B, B]$, and their initial phase $\phi_k(t = 0)$ is randomly selected from $[0, 2\pi]$. A Gaussian noise process is then added to the resulting signal. Finally, the received signal is matched filtered and demodulated assuming perfect parameter estimation. Note that, as the interfering signals can be centred at any frequency in $[-B, B]$, the required system bandwidth spans $[-1.5B, 1.5B]$. Therefore, the number of samples per symbol at the output of the shaping filter must be set to $T/T_s = 3L$, where T and T_s are the symbol and the sample duration, respectively; and L is typically 8 or higher. In the scenarios studied next, each simulation point is obtained by averaging 10^6 interference realizations over a test packet. Results are expressed in [b/s/Hz] according to the underlying spectral efficiency η.

Perfect Power Control

We start by considering a scenario where each packet arrives at the receiver with exactly the same power, so $A_0 = A_1 = \ldots = A_K = A$. The signal-to-noise (SNR) ratio per bit is set to $E_b/N_0 = A^2/N_0 = \infty$, so that the packet errors are dominated by packet collisions. The performance of pure ALOHA under the same PHY model has also been evaluated as reference. Results for the throughput and the PER are presented in Figure 6.11, where it can be seen that the performance gap between TFAA and ALOHA is rather small compared to that obtained under the collision channel model. This difference is even smaller for practical channel loads, that is, for $G < 0.4$, where the response is roughly linear. Under these conditions, we may conclude that there is no significant performance loss when moving from a TA to a TFA system.

Log-normal Fading

We have also evaluated the performances over a more realistic scenario in which the power levels $Z = A_k^2$ for both the desired and the interfering packets are subject to a log-normal fading process of mean μ [dB] and standard deviation σ [dB]. For this scenario, we have considered $\mu = 0$ [dB] and $\sigma = 0, 3$ and 5 [dB]. The average SNR per bit is set to $E_b/N_0 = 15$ [dB]. The results in Figure 6.12 show that, as opposed to the behaviour observed in spread-spectrum RA schemes, the throughput performance of TFAA is actually expanded in the presence of power imbalance[7]. However, the increase in throughput comes at the cost of a very high packet error rate, which is not tolerable in practical systems. For $G < 0.6$ we observe a slight performance degradation with the level of power imbalance σ, explained by transmission errors rather than by collision events.

[7]See for instance [DD12, Section IV-D] for Spread ALOHA. It should be noted that a completely different behaviour is observed when SIC is employed: in this case, the performance of spread-spectrum techniques is improved under power imbalance.

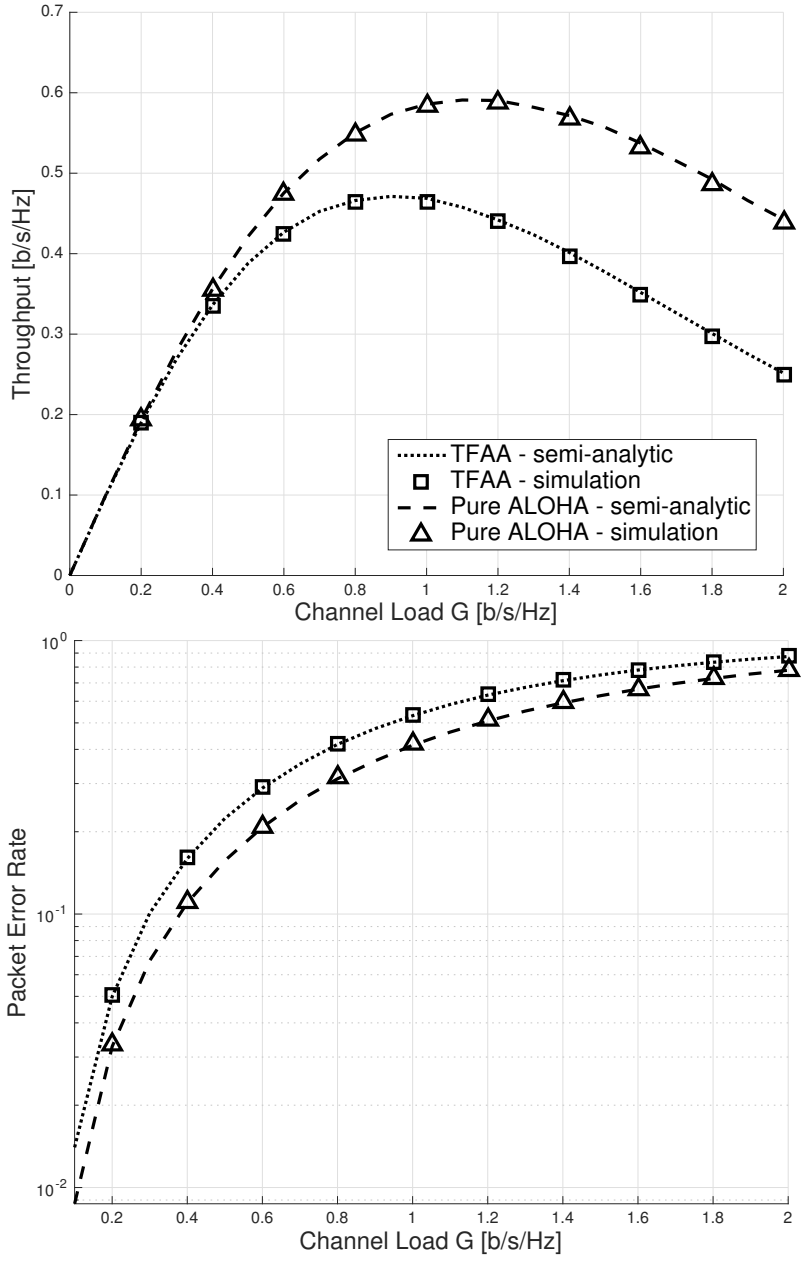


Figure 6.11 – Throughput and PER performance of TFAA and ALOHA over the PHY model with perfect power control. $E_b/N_0 = \infty$, $\eta = 1$ [b/s/Hz].

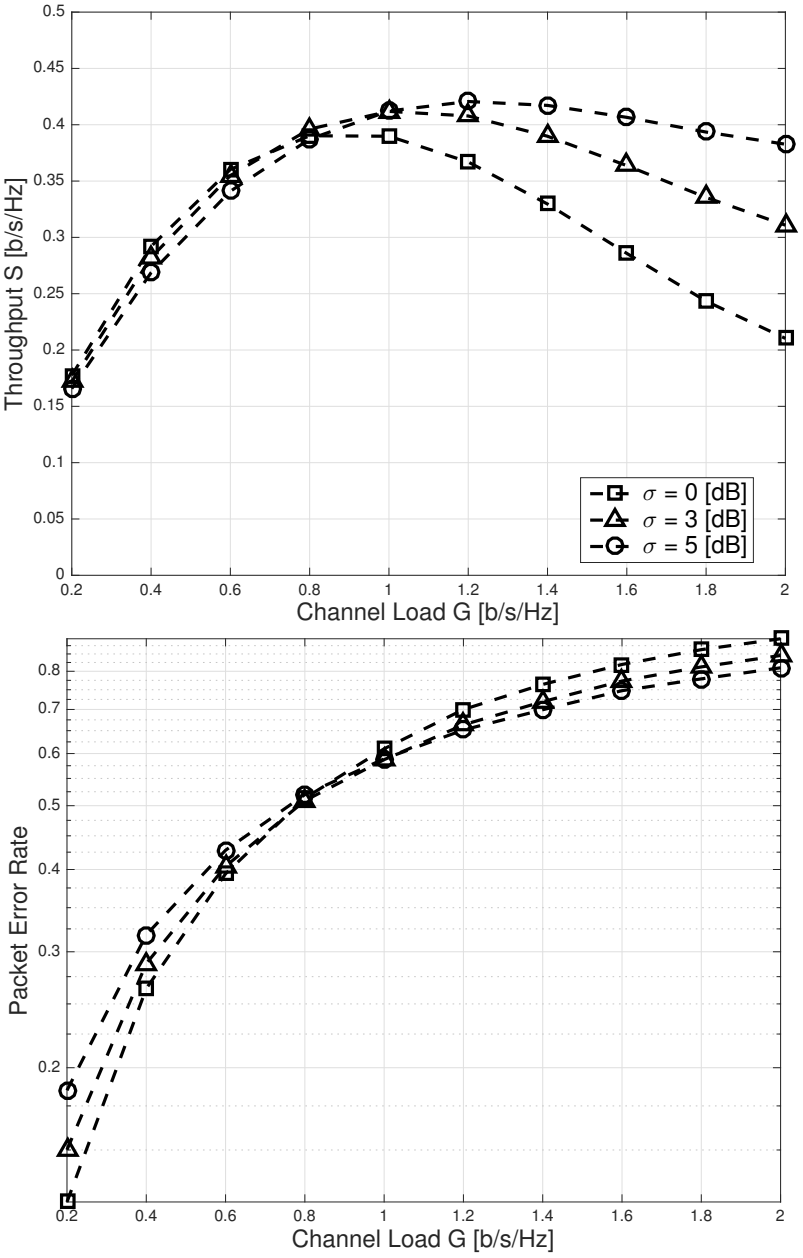


Figure 6.12 – Throughput and PER performance of TFAA under log-normal fading and AWGN. The average SNR is set to $E_b/N_0 = 15$ [dB] with standard deviation σ [dB], $\eta = 1$ [b/s/Hz].

6.7 Summary and Conclusions

In this chapter, we have first introduced a model framework for throughput performance evaluation of TFAA systems. Upon the baseline model, three different reception models have been studied. For the collision channel, an exact, closed-form expression has been derived through a simple analysis, which provides an absolute lower bound of the system performance. In particular, we have shown that a maximum normalized throughput $S = 1/(4e)$ can be achieved. Since the results obtained for the collision channel are rather pessimistic, we have also studied a capture channel based on well-known models described in the literature. The suitability of this model in the time-asynchronous regime is however questionable, since the average SINR may not be a representative figure of merit. Then, we have considered a more accurate model which takes into account the underlying physical layer implementation. In this case, the time-varying MAI level that affects the transmitted packets is considered instead of the average SINR. We have also investigated the characteristics of the MAI for the case of rectangular pulse shaping. Finally, exact throughput and PER performances were obtained using the Monte-Carlo method and validated by computer simulation. Globally, we conclude that adding an additional degree of asynchronism does not involve a significant performance loss.

Future work should address the issues related to UNB operation, in particular the impact of Doppler effects in systems where the transmitter-receiver relative velocity is high. While this aspect has been studied for the collision channel, a more detailed analysis of the PHY channel is still required. Finally, more system-specific issues such as the performance degradation due to imperfect parameter estimation should also be considered in future work. Other fundamental aspects, such as the impact of modulation order, coding and the pulse shaping filter, will be addressed in the next chapter.

" Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin."

John von Neumann (1903 – 1957)

7

Transmission Parameter Optimisation for TFAA Systems

Contents

Following the baseline time- and frequency-asynchronous ALOHA system described in the previous chapter, in particular the proposed PHY model, our interest here is to determine in which ways the underlying transmission parameters, including the use of Forward Error Correction, influence the overall MAC performance. In order to do so, a series of simulation experiments have been performed, the results of which are reported in this chapter.

7.1 Introduction

In the previous chapter, we have studied the performance of TFAA considering a very simple PHY layer design based on plain BPSK. This is, in fact, the most common configuration used in practical UNB networks and is particularly appealing because of its simplicity and low complexity requirements at the transmitter side, which results in a low-cost, energy-efficient terminal. Indeed, because of the practical constraints imposed by LPWANS, further improvements in system performance should be pursued without increasing significantly the terminal complexity. On the other hand, since the packets transmitted in LPWANs are typically short, the advantages provided by complex modulation and coding schemes are limited [Yoo+10].

Our focus in this chapter will be in three main PHY layer aspects: modulation order (or constellation size), forward error correction (FEC) and pulse shaping filtering. These blocks are fundamental since they characterise the signal waveform, and, more importantly, they also have a direct impact on the MAC performance.

The transmitter model that we consider here is very similar to the one studied in the previous chapter, the main difference being that an optional FEC block is added before the modulator, as shown in Figure 7.1; and that a non-rectangular pulse shaping filter is now considered. While there is a large number of FEC schemes proposed in the literature, here we have only considered convolution and turbo coding mainly because they are relatively simple to implement, specially at the transmitter side, and provide relatively good performance. Furthermore, these codes are appealing in the short packet regime [Yoo+10]. Similarly, we consider square root raised cosine (SRRC) pulse shaping filtering since it is largely employed in practical systems and has attractive spectral properties—though other pulses might provide better performance.

Our interest here is not to find an optimal PHY layer configuration nor to perform an exhaustive study on the performance of coded M-PSK TFAA systems. Rather, we limit ourselves to obtain some general insight about the trade-off between spectral efficiency and system throughput at the MAC level, considering the PHY channel model described in the previous section.

The rest of this chapter is organized as follows. In Section 7.2 we briefly describe the simulation model used to carry out our experiments. The results obtained for each of the three blocks considered are discussed separately in Sections 7.3, 7.4 and 7.5. Finally, Section 7.6 summarizes this chapter.

7.2 Simulation Approach

As noted in the previous chapter, the MAI in TFAA forms a stochastic process, which implies that, according to the set of random signal parameters that define each interfering signal, the MAI levels at each symbol of the test packet are not independent. As a consequence, any closed-form expressions available in the literature for assessing the performance of coded modulation schemes

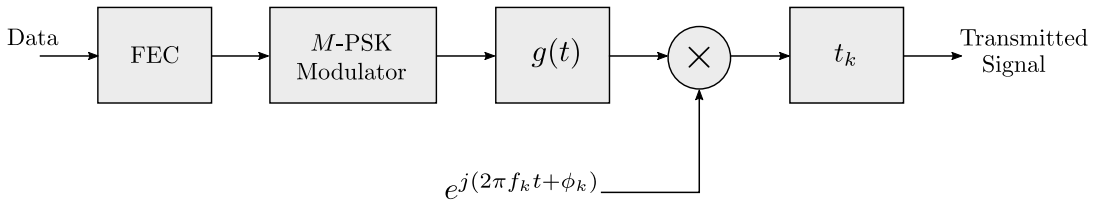


Figure 7.1 – TFAA transmitter block diagram for packet k.

cannot be directly applied—specially if we are interested in obtaining accurate insight—since they are typically based on Gaussian fading channel models with known statistics. For this reason, we resort to computer simulation.

Instead of transmitting packets randomly over the whole system bandwidth W, a simplified simulation model in which only one test packet is transmitted at each realization has been considered, following the approach described in Section 6.6.4. We assume that this packet is transmitted at $(t, f) = (0, 0)$ and that the interfering packets are randomly placed within the region $[-T_p, T_p] \times [-B, B]$. By contrast, since a FEC block has now been included, the number of information bits carried by each packet is now $L_b = \rho \log_2(M) L_p = \eta L_p$, where L_p is the packet length in symbols, M the modulation set size, $0 < \rho \leq 1$ the coding rate and η the spectral efficiency, which is in general defined as

$$\eta = \rho \log_2(M). \tag{7.1}$$

Similarly, the number of interfering packets K follows now a Poisson distribution of intensity $4G/\eta$, where G is the useful channel load in b/s/Hz (for more details regarding the simulation model, the reader is referred to Section 6.6.4).

7.3 Choice of the Pulse Shaping Filter Roll-off Factor

Due to the nature of the interference in TFAA, a question that arises is whether the use of spectrum-efficient pulse shaping filters may have a beneficial impact on the system performance. Indeed, one might feel tempted to think that if the frequency support of a packet signal can be narrowed down then one can reduce the average number of packet collisions in the system. However, we know from time–frequency analysis that limiting a signal's bandwidth necessarily involves an increase in its time span. This is known as the time–frequency uncertainty principle, or the Heisenberg–Gabor limit, and has been formalized by several authors (see for instance [Har33]).

While decreasing the signal bandwidth does reduce the packet vulnerability region—at least from a high-level point of view—the time-domain signal expansion has also a negative impact in the MAI characteristics. In other words, since a symbol waveform in fact spans several times the

symbol duration, reducing its bandwidth would further spread the MAI over time. We refer to this effect as *cross-symbol interference*.

It is therefore convenient to find a trade-off regarding the time–frequency characteristics of the transmitted signal to minimise the MAI and thus improve the system performance. In order to gain some insight into this question, we consider in the following a SRRC pulse shaping filter having the form [Pro07]:

$$g(t) = \mathrm{sinc}(\pi t/T)\frac{\cos(\pi\beta t/T)}{1-4\beta^2t^2/T^2}, \tag{7.2}$$

where T is the symbol duration, $\beta \in [0,1]$ is called the *roll-off factor* and $\mathrm{sinc}(x) = \sin(x)/x$. The roll-off factor is a direct measure of the spectral efficiency, as the pulse frequency support is mainly contained in the interval $[-(1+\beta)/2T, (1+\beta)/2T]$. For instance, if $\beta = 0$ the frequency response is perfectly rectangular with support on $[-1/2T, 1/2T]$, which corresponds to $\pm$ the Nyquist bandwidth. In this case, we say that the excess bandwidth is 0% and the spectral efficiency is to the maximum. Note also that in this case the pulse is $g(t) = \mathrm{sinc}(\pi t/T)$ and hence its time response is unbounded. On the contrary, a roll-off factor equal to 1 yields a pulse bandwidth which is twice the Nyquist frequency, so the excess bandwidth is 100% in this case.

In the following, we investigate the impact of the pulse shaping filter roll-off factor on the MAI. We start by reviewing the time-asynchronous case, for which some analytical results have been already derived. In particular, it has been found that the MAI variance from a set of K uncorrelated, equal power BPSK/QPSK interfering signals has the form

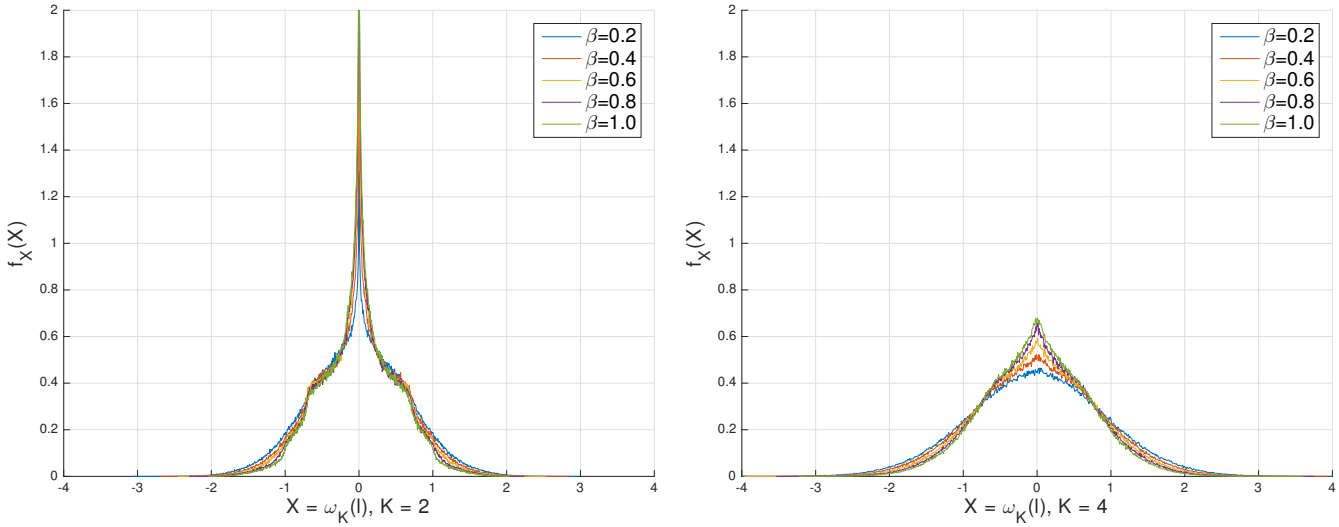
$$\mathbb{E}\left[\omega_K^2\right] = KF\mathbb{E}[A_k^2], \tag{7.3}$$

where A_k is the amplitude of each interfering signal and F, which depends on the receiver's matched filter output $m(t) = g(t) * g(-t)$, is given by [TV99]

$$F = 1/T\int_{-\infty}^{\infty}|m(t)|^2. \tag{7.4}$$

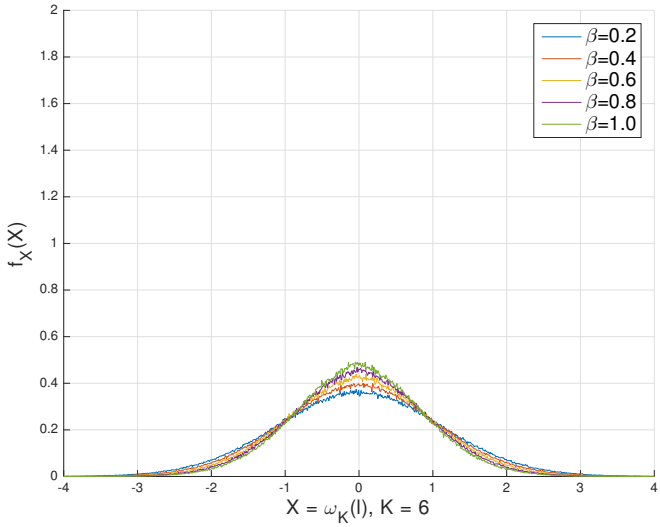
For our election of SRRC pulse shaping, we have $F = 1-\beta/4$. Therefore, as we have pointed out in the discussion above, the MAI variance decreases with β since a wider pulse shape produces less cross-symbol interference.

Let us now return to the time- and frequency-asynchronous case that concerns us here. Since no closed-form expressions of the MAI statistics are available, we compute numerically the probability density function (pdf) of the MAI amplitude from Equations (6.21) and (6.22), following the approach in [TV99]—though this time we use the SRRC pulse defined above instead of the rectangular pulse considered in the previous chapter. The obtained MAI pdf's are shown in Figure 7.2 for roll-off factors $\beta = 0.2, 0.4, 0.6$ and 0.8, where each subfigure corresponds to a given number of interfering packets $K = 2, 4, 6$ and 8. As expected, the pdf's converge to a bell-shaped distribution

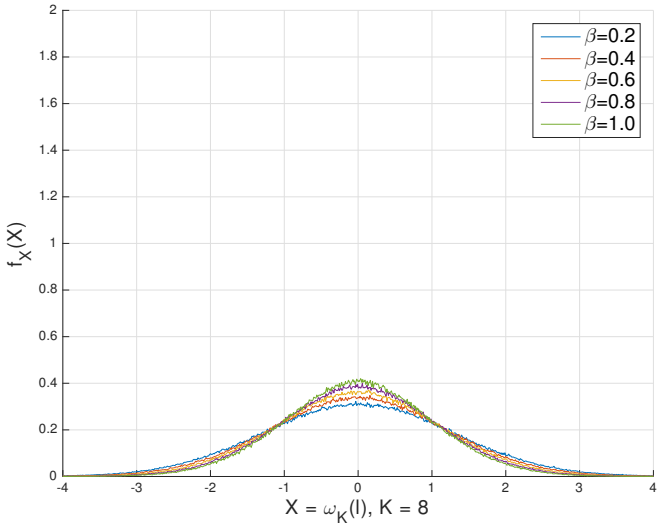


(a) $K = 2$.

(b) $K = 4$.



(c) $K = 6$.



(d) $K = 8$.

Figure 7.2 – Distribution of $\omega_K(l)$ for different number of interferers K and roll-off factors.

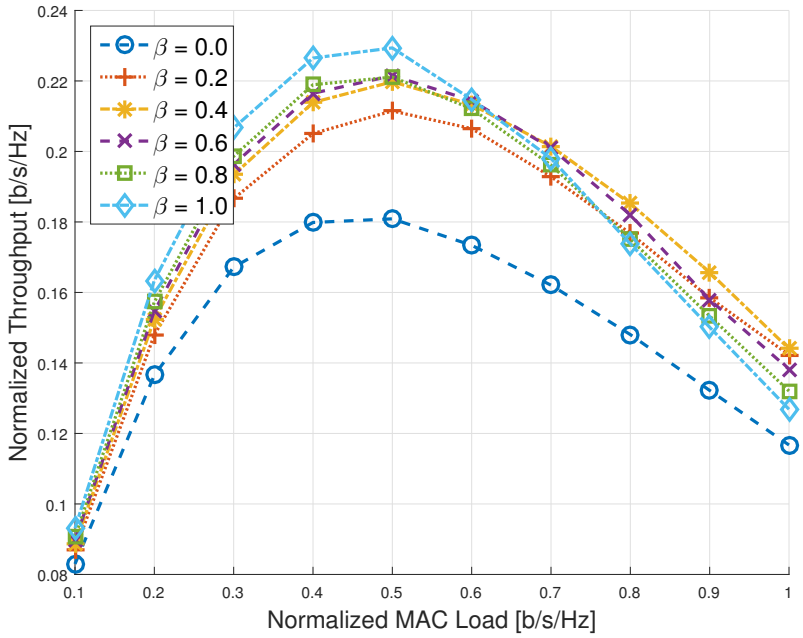


Figure 7.3 – Simulation results for the throughput of TFAA for different roll-off factors, $E_b/N_0 = \infty$, $M = 2$, $\rho = 1$ (no FEC), $\eta = 1/(1+\beta)$ b/s/Hz.

when the number of colliding packets increases, which is in line with the central limit theorem. Note also that, the probability of having high interference levels, *e.g.* $P(|\omega_K| > 1)$, increases as β decreases. In other words, the MAI variance decreases with β, as in the time-asynchronous case.

These results suggest that adopting a high value of β would allow us to minimise the effects of the MAI. However, to fully understand the impact of the pulse shaping filter roll-off factor on the system throughput, we must also consider the fact that increasing the excess bandwidth also increases the vulnerability region, which is given by $\Lambda = 4T_pB = 4T_p(1+\beta)R_s$[1]. Thus, in our simulation model, the intensity of the packet arrival process should be now computed considering that $\eta = \rho \log_2(M)/(1+\beta)$. Note that since the spectral efficiency is reduced by a factor $1/(1+\beta)$ the obtained throughput values are scaled down with respect to the other results presented in this thesis.

Figure 7.3 shows throughput results obtained by simulation for different values of β. The simulation scenario considers plain BPSK transmission, no FEC and $L_p = L_b = 120$. The channel does not induce any type of distortion, so we assume $E_b/N_0 = \infty$ and no attenuation. In this way,

[1]It is worth noting that, in all other throughput calculations included in this thesis, the pulse shaping filter excess bandwidth is not considered in order to provide results that can be fairly compared with similar random access schemes proposed in the literature.

we isolate the effects of the MAI in the throughput performance. We observe that, for relatively low channel loads, that is for approximately $G < 0.7$, expanding the pulse bandwidth has a positive impact and $\beta = 1$ achieves optimal performance. This is because at low channel loads minimising the cross-symbol interference allows us to cope with the additional number of colliding packets that results from the expansion of the vulnerability region. By contrast, for $G > 0.7$, the number of colliding packets is too important and minimising the cross-symbol interference does not bring significant benefits. Finally, it is interesting to note that setting $\beta = 0$—in which case the pulse corresponds to a sinc function— is always the worst option. This is not surprising because a sinc pulse has a longer time response (infinite, in theory) and hence entails a larger cross-symbol interference. Since it is typically preferable to operate a random access system at relatively low packet error rates (*e.g.* $P_e < 0.1$), we conclude that $\beta = 1$ is optimal for practical applications.

7.4 Impact of FEC

As we pointed out in Section 6.2, one of the most interesting characteristics of TFAA is that—according to our simulation experiments—it can be significantly improved through FEC techniques.

In TFAA a typical collision only involves a fraction of a packet. In fact, the average overlapping factor is $\bar{\Theta} = \mathbb{E}[\alpha_i\beta_i]$=1/4, α_i and β_i being random variables representing the overlapping factor in the time and frequency domains, respectively, for some colliding packet i. Therefore, it is likely that the corrupted bits might be recovered through FEC, as long as the affected area is not too large.

Note also that the error patterns due to collisions are similar to the so-called *burst* errors that are observed in some wireless communication systems. As a consequence, low-rate error correcting codes including an interleaving stage, which are usually employed in these environments, may also bring important performance gains in TFAA systems.

In our first simulation experiment, our objective is to study the trade-off between the bandwidth expansion due to the FEC overhead and the coding gain. To that end, we have considered convolutional FEC in our simulation model. The simulation scenario is as follows. As usual, the modulation format is BPSK with roll-off $\beta = 1$. The packet length, including FEC redundancy, is kept to 120 symbols. An ideal transmission channel is assumed. Also, we ignore the impact of the pulse shaping filter excess bandwidth, so that the intensity of the arrival process is $4G/\rho$. The code generators are standard maximum free distance codes with constraint length $K_L = 7$ obtained from [Pro07]. Hard-decision Viterbi decoding is considered. Figure 7.4 illustrates the performances obtained for coding rates $\rho = 1, 1/2, 1/3, 1/4, 1/5$ and $1/6$ over the infinite SNR regime. It is shown that TFAA benefits significantly from the use of low-rate FEC, even in the absence of thermal noise. In particular, for low traffic loads, we observe that the PER can be improved by about one order of magnitude. Is is also interesting to note that as we increase the FEC redundancy the throughput performance is always improved (at least for the set of coding rates considered here),

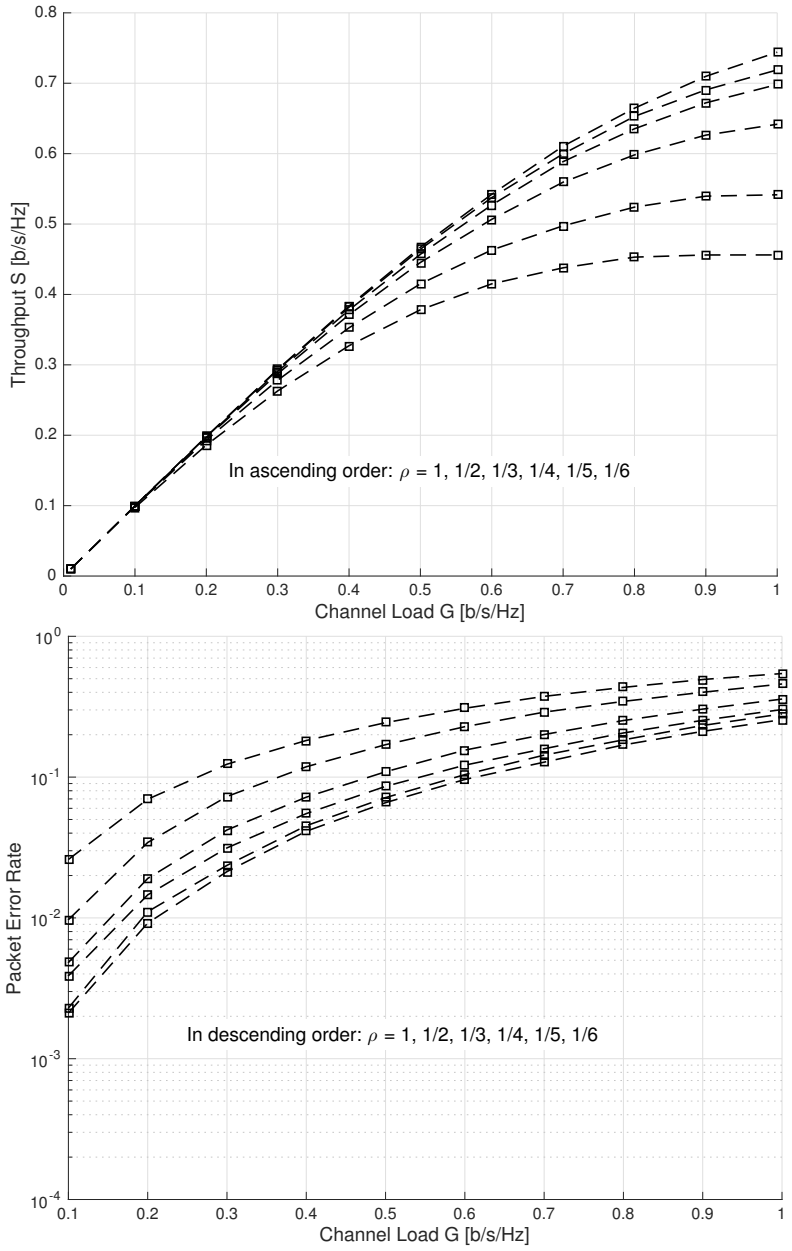


Figure 7.4 – Simulation results for the throughput and PER performance of TFAA over the capture channel model using low-rate convolutional FEC with hard-decision Viterbi decoding, $E_b/N_0 = \infty$, $\eta = \rho$ b/s/Hz.

though with diminishing marginal gains. For this particular configuration, coding rates between 1/4–1/3 provide a reasonable balance between performance gain and complexity.

In our second simulation experiment we have considered turbo FEC. The objective is two-fold. On the one hand, we are interested in determining whether TFAA can achieve better performance than ALOHA using efficient FEC in a more realistic transmission environment. To that end, we have considered $E_b/N_0 = 15$ dB and a flat fading channel where the packet amplitudes are lognormally distributed with mean $\mu = 0$ dB and standard deviation $\sigma = 3$ dB. On the other hand, since the channel estimates obtained by a typical soft demodulator are imperfect in a TFAA multi-user environment[2], we are also interested in determining whether the use of iterative soft decoding might provide significant performance gains. The results of this experiment are reported in Figure 7.5. We have chosen a rate 1/3 turbo code with a $(133, 171)$ generator (in octal form) and constraint length $K_L = 7$, which corresponds to a standard 3GPP encoder used in terrestrial mobile networks [3gp]. For the turbo decoder configuration we have set the number of decoding iterations to 8 and the algorithm is based on true *a posteriori* probabilities. The simulation scenario has minor differences with respect to the previous one. In particular, the channel that we consider now is non-ideal and the packet length, including FEC redundancy, is set to $L_p = 300$ symbols. The performances obtained by a rate 1/3 convolutional code using hard-decision Viterbi decoding are also included for comparison. Clearly, the results confirm that TFAA benefits more from FEC than ALOHA, reaching even higher throughputs. Furthermore, the performance gap is significant, specially when turbo FEC is considered. The results also show that the use of iterative turbo decoding brings additional benefits, even though the soft metrics are imperfect. There are, however, practical issues that should be taken into account in this scenario. In particular, estimating the MAI variance for computing the soft metrics is difficult in the short packet regime. Indeed, we expect that the performance gap between the turbo and Viterbi codes reduces as we decrease the packet length.

7.5 Impact of the Modulation Order

As we increase the modulation order M the bandwidth efficiency is improved according to Eq. (7.1). As a consequence, the number of packets required to convey a given amount of information is reduced, which means that the load on the channel is eased. The disadvantage of this is that, as the constellation size increases, the symbol error probability increases as well. For a typical point-to-point link, the use of higher modulation orders implies that more transmission power is required in order to achieve a given BER target. By contrast, in a multi-user environment, increasing the

[2]This is due to the fact that the log-likelihood ratios are computed assuming a Gaussian distribution of the MAI interference power and using the estimated variance over the packet duration. However, as pointed out earlier, the MAI power in TFAA may show significant fluctuations.

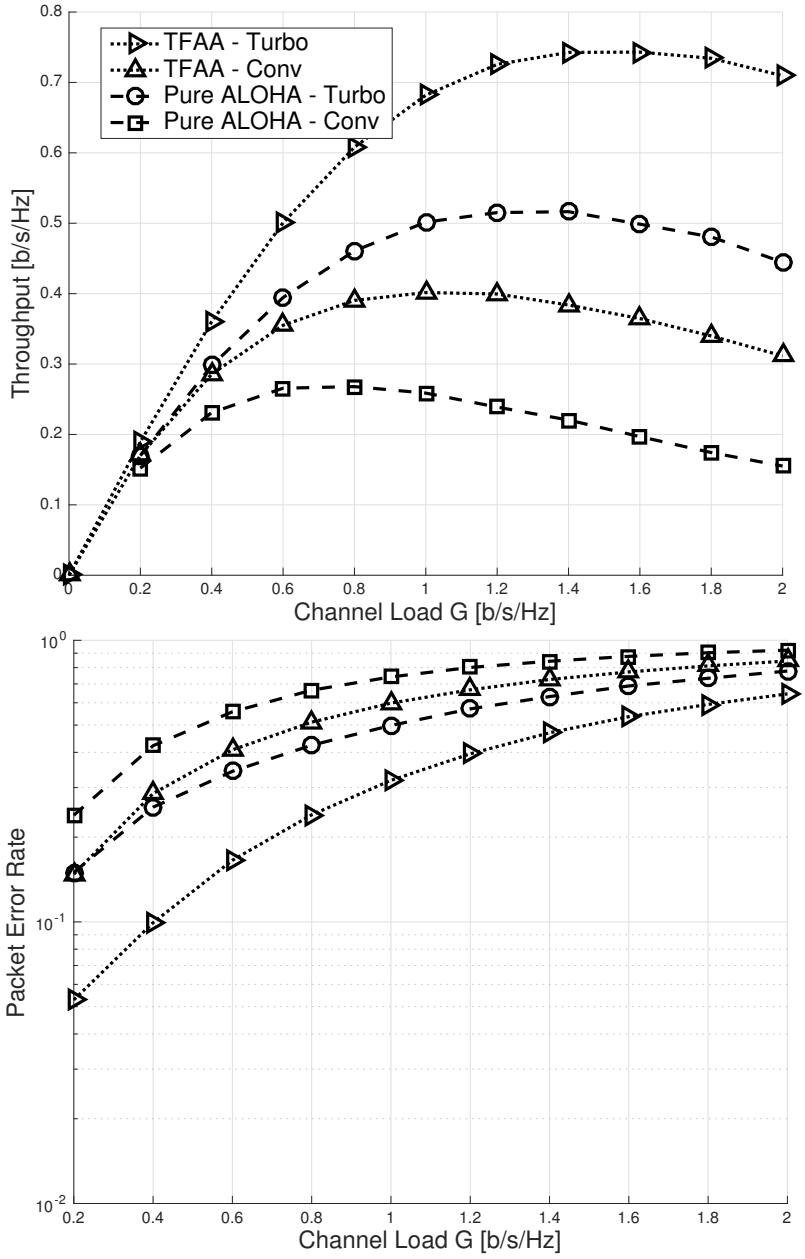


Figure 7.5 – Throughput and PER performance of turbo-coded and convolutional-coded TFAA and ALOHA over the PHY channel model with lognormal fading parameters $(\mu, \sigma) = (0, 3)$ dB and $E_b/N_0 = 15$ [dB]. $M = 2$, $\beta = 1$, $\eta = 1/3$ b/s/Hz.

transmission power also results in increased MAI power, which has a negative impact on the overall performance.

Again, we have studied this trade-off by taking into account only the effects of MAI and neglecting other transmission impairments, hence we assume $E_b/N_0 = \infty$ and no fading. The packets are transmitted using plain M-PSK with rolloff $\beta = 1$, so that the spectral efficiency is assumed to be equal to $\eta = \log_2(M)$ b/s/Hz. We have considered relatively low modulation orders ($M = 2, 4$ and 8) since it is difficult to implement high-order schemes in the short packet regime due to the short communication duty-cycle. In addition, even in the absence of MAI, transmitting at higher modulation orders typically requires a larger link budget, which is difficult to achieve in a LPWAN context. The throughput and PER performances obtained for this scenario are presented in Figure 7.6. We observe that for channel loads below 0.9 b/s/Hz, approximately, BPSK provides the highest throughput, though the difference is not significant. The difference is more apparent in terms of the PER, specially for the lowest channel loads considered. We may conclude that, since in this region the average number of colliding packets is already low, there is now much interest in improving the spectral efficiency and it results slightly more efficient to use a more robust modulation scheme. For $G > 0.9$ b/s/Hz, 4-PSK and 8-PSK provide the best performance, with very similar results. In this region, improving the spectral efficiency does help to reduce the MAI. However, as we pointed out before, optimising the performance at high channel loads is not compelling due to the high PER. Thus, for the considered scenario, BPSK is the optimal modulation scheme.

7.6 Summary

In this chapter, we have presented and discussed the results of several simulation experiments. In particular, we evaluated the impact on the MAC performance produced by varying the pulse shaping filter roll-off factor β, varying the modulation order M, and applying low-rate FEC. We have seen that improving the spectral efficiency whether by using higher modulation orders or by reducing the roll-off factor has a similar effect. In both cases, it does not seem to be an attractive strategy since it only allows to improve the performance at relatively high channel loads, *i.e.*, when the PER is too high for efficient operation. By contrast, we observed that applying low-rate FEC helps to improve the performance significantly. Further, the results show that TFAA can effectively outperform ALOHA, as we suggested in the previous chapter.

The results obtained here should be only considered as general guidelines for the design of practical systems. Indeed, an optimal PHY layer configuration must be selected taking into account system-dependent issues (as for instance, complexity and energy constraints) and a more detailed modelling framework.

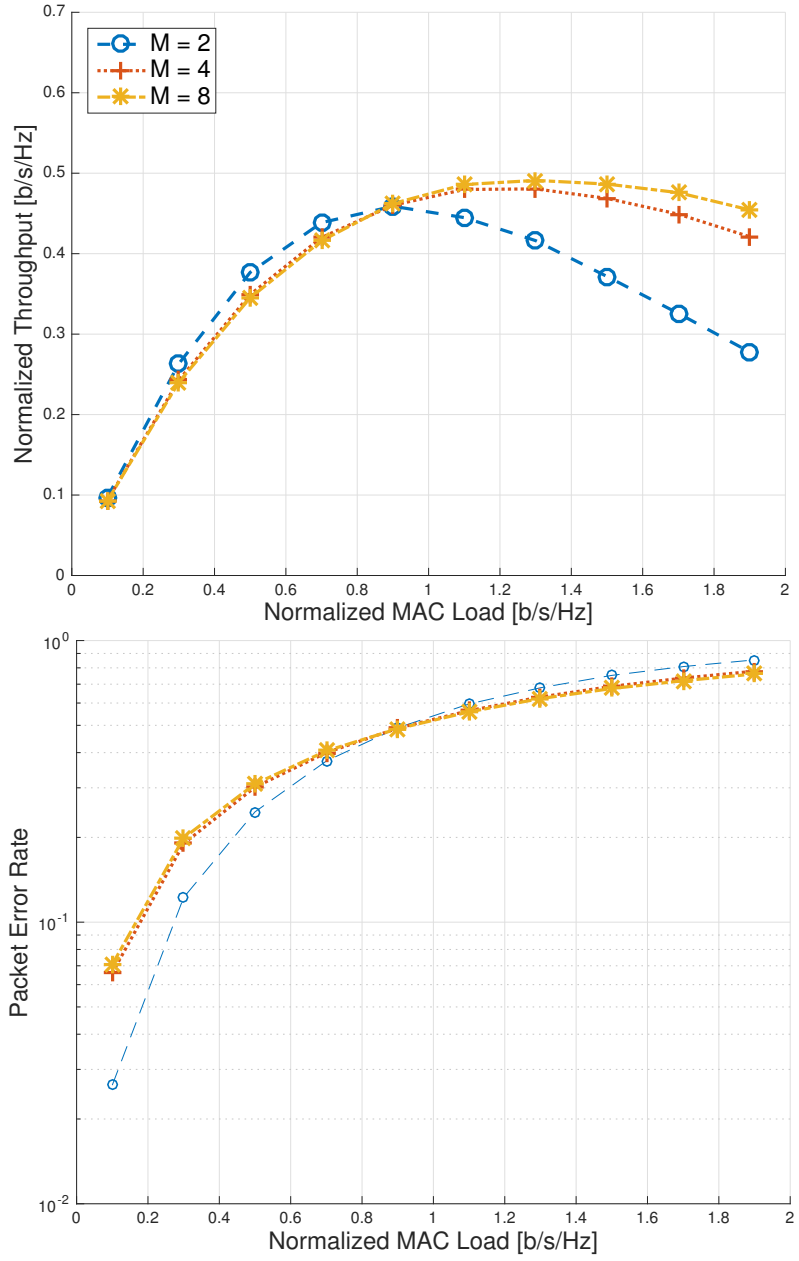


Figure 7.6 – Throughput and PER performance of TFAA for modulation orders $M = 2, 4$ and 8. $E_b/N_0 = \infty$, $\beta = 1$, no FEC, $\eta = \log_2(M)$ b/s/Hz.

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

John Von Neumann (1903 – 1957)

8

Contention Resolution Time- and Frequency-Asynchronous ALOHA

Contents

As we have seen, time- and frequency-asynchronous ALOHA (TFAA) provides interesting advantages for terrestrial low power wide area networks as well as for satellite machine-type data collection networks; namely, low power consumption, low terminal cost/complexity and long range communication capabilities. However, its throughput and packet error rate (PER) performance is rather poor with respect to state-of-the-art techniques. In this chapter, we present Contention

Resolution Time- and Frequency-Asynchronous ALOHA (CR-TFAA), an enhancement of TFAA which is in line with recent advances in asynchronous random access schemes, such as Asynchronous Contention Resolution Diversity ALOHA (ACRDA). In particular, CR-TFAA exploits time–frequency diversity and successive interference cancellation. The throughput and PER performances are evaluated through a semi-analytical method and also by computer simulation. The results show that, compared to ACRDA under similar conditions, CR-TFAA provides better PER performance over the linear throughput region. The work presented in this chapter has been published in [AF17a].

8.1 Introduction

In Chapter 4 we reviewed some random access protocols capable of providing high throughput at extremely low packet error rates. These developments have been made possible thanks to recent advances in signal processing techniques as well as to the adoption of cross-layer design strategies. Thus, packet collisions can be resolved at the PHY layer—an approach that is more efficient than the use of MAC-layer retransmission schemes. In general, high-throughput random access protocols are characterized by a combination of time diversity—which is achieved by transmitting two or more replicas of each packet over a frame structure spanning several time slots—, low-rate FEC codes and successive interference cancellation (SIC) at the receiver side. This approach is appealing because most of the additional complexity is introduced at the receiver side. Indeed, the main processing stage required at the transmitter is the FEC block, which for a typical convolutional encoder only consists of a few shift registers and, consequently, has a marginal complexity compared to the FEC decoder.

The objective of this chapter is to study the benefits brought by these techniques in the context of TFAA. While more advanced techniques have also been recently proposed, exploiting for instance soft packet combining [CKM16], here we focus on a more practical design approach following the main concepts introduced in Contention Resolution Diversity Slotted ALOHA [CDH07]—which has already been implemented in the DVB-RCS2 standard [Dvb]—and its non-slotted counterparts Contention Resolution ALOHA (CRA) [Kis11] and Asynchronous Contention Resolution Diversity ALOHA (ACRDA) [De +14]. The resulting design has been called Contention Resolution Time- and Frequency-Asynchronous ALOHA (CR-TFAA) and can be considered as a generalization of ACRDA. We will see that, as TFAA has better FEC collision resolution capabilities with respect to ALOHA and slotted ALOHA, CR-TFAA can also bring better performance than ACRDA and CRDSA.

In Section 8.2 we describe the CR-TFAA protocol and provide a model for its analysis. Then, in Section 8.3 we derive the throughput-PER performance following a semi-analytical approach, and we also describe the simulation model that was developed to validate our results. Numerical results obtained by the two methods are presented and discussed in Section 8.4. Section 8.5 concludes

this chapter.

8.2 CR-TFAA System Model

In order to facilitate comparison with other relevant high-throughput RA schemes such as ACRDA, the design that we adopt here follows some of the guidelines given in [De +14; CDH07]. In particular, the underlying PHY layer is based on turbo FEC and M-PSK modulation and, at the receiver side, a sliding-window decoding algorithm is considered. The main assumptions and notations used for the analysis of TFAA in Chapter 6 are also adopted here, namely:

1. The system bandwidth is denoted by W. Packets are of equal size and are assumed to be ideally contained within a time–frequency rectangle of size $B \times T_p$.

2. The overall packet arrival process is Poisson with intensity λ [packets/s]. We consider an infinite population system wherein each user generates a packet with very low probability, so that each new packet arrival is generated by a different user. The *normalised* throughput is defined according to Eq. (4.2):

$$T(G) \triangleq G \times (1 - P_e), \tag{8.1}$$

   where $G \triangleq \lambda T_p \times B/W$ is the *normalised* average channel load and P_e the PER. Hence, since the packets are transmitted randomly at any location within the frame, G is the intensity of packet arrivals over a section $B \times T_p$.

3. We assume flat fading operation. The E_b/N_0 of a packet is constant during T_p, but can vary from subsequent transmissions.

4. The channel model corresponds to the PHY model described in Section 4.3.2, in which the actual MAI plus noise is considered.

5. Channel and signal parameters are perfectly estimated by the receiver.

For a discussion on the validity of these assumptions, the reader is referred to Section 6.3.

Packet transmissions are organized in virtual frames (VF), as proposed in [De +14] for ACRDA. A VF is simply a local frame structure spanning several time slots and, in contrast to synchronous systems as CRDSA, its starting time does not follow a global network reference. The VF duration is $T_F = N_{\text{slots}} \times T_p$, where N_{slots} is the number of time slots per frame. When a new packet is generated, a VF is build and N_{rep} replicas of the packet are created and placed in random locations within the VF, *i.e.* at random time slots and at random carrier frequencies, but taking care of selecting non-overlapping positions among the packet replicas. As in TFAA, the carrier

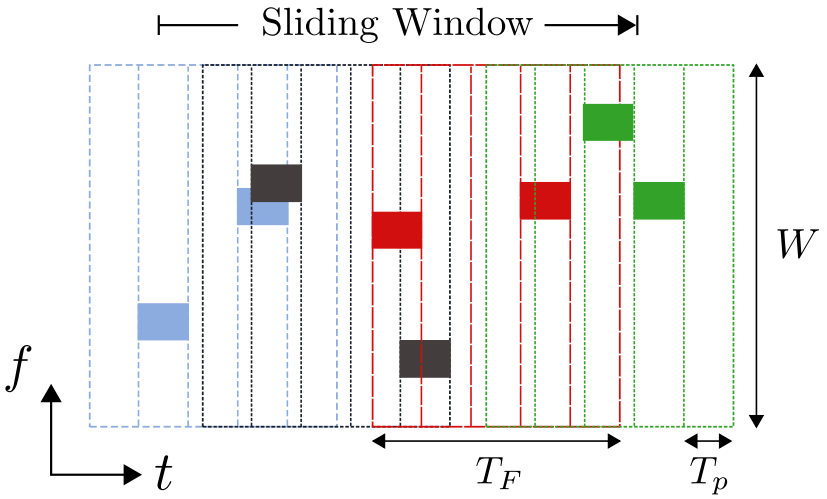


Figure 8.1 – Frame structure in CR-TFAA. Virtual frames from four different users are represented with different colours.

frequencies are selected randomly within the interval $[f_m, f_M]$, where $f_m = -B([W/B] - 1)/2$ and $f_M = B([W/B] - 1)/2$.

In CR-TFAA we define the frame size $N_{t\times f}$ as the number of replicas that can be placed without overlapping over the rectangle $W \times T_F$. It follows that $N_{t\times f} = N_{\text{slots}} \times \lfloor W/B \rfloor$. VFs are generated and transmitted asynchronously by each user and thus arrive at the receiver with random time offsets. The main difference between the VF structure in ACRDA and that of CR-TFAA is that in the latter the VF spans a bandwidth (W) greater than the packet bandwidth, so that replicas can be freely transmitted in the frequency domain. Figure 8.1 shows an example where the VFs from four different users are transmitted using $N_{\text{rep}} = 2$, $W/B = 10$ and $N_{\text{slots}} = 5$.

Along with the L_b payload bits, each replica should also contain a few extra bits where the relative position of the other $N_{\text{rep}} - 1$ replicas is signalled. Note that in the diversity scheme adopted here the VF time axis is divided in a discrete number of slots, while the frequency axis is non-slotted. The reason for this is that defining frequency slots demands a certain amount of frequency synchronization and oscillator stability, which cannot be achieved with UNB signals as discussed earlier[1]. The replicas' information bits are FEC encoded using a rate $\rho \simeq 1/3$ turbo code, modulated using M-PSK and filtered through a square root rise cosine (SRRC) filter $g(t)$

[1]In practice, the use of frequency slots—even if oversized—is convenient as it allows to minimise the signalling overhead. In either case, the relative frequency position of a replica can only be given approximately in UNB.

with roll-off factor β. Thus, the signal associated with some replica k is modelled as

$$r(t) = a_k \sum_{l=0}^{L_p-1} d_k(l) g(t - lT) e^{j(2\pi f_k t + \phi_k)}, \tag{8.2}$$

where L_p is the packet length in symbols, a_k^2 the channel gain, the sequence $\{d_k(l),\, 0 \leq l < L_p\}$ represents the replica's symbols, T is the symbol duration, $f_k \sim \mathcal{U}[f_m, f_M]$ is the carrier frequency and $\phi_k \sim \mathcal{U}[0, 2\pi]$ a random phase offset. It is assumed that the replicas are affected by a log-normal flat fading process of mean μ [dB] and standard deviation σ [dB], and by additive white Gaussian noise (AWGN). Note that a Doppler shift $f_D << W$ can also be considered, which further randomizes the observed frequency f_k. However, we do not account for the signal distortion that can occur under high Doppler effects, specially when UNB modulation is employed.

At the receiver side, the incoming signal is sampled using a sliding window as in [De +14]. The sliding window spans W_{VF}[2] VFs with a step size ΔW_{VF}, expressed as a fraction of one VF. The SIC process starts by scanning the sliding window samples over the whole system bandwidth in order to detect the presence of replicas (as mentioned above, we assume that the replicas can be perfectly located and their parameters perfectly inferred by the receiver). When a replica is correctly located, it is demodulated through a matched filter detector whose soft output is then fed to a turbo decoder. If the replica is successfully decoded, its interference contribution is removed from the sliding window buffer. The interference from other replicas of the same packet present in the sliding window is removed next. Note that here we are assuming that the interference cancellation process is ideal, that is, no residual signal components remain after removing a replica. If one or more packet replicas have not been fully received yet, the packet information is stored until the remaining incoming replicas are received and removed. The SIC process is repeated a number of times up to a maximum of N_{it} iterations. Then, the sliding window is shifted in time by $\Delta W_{\text{VF}} T_F$ [s] and the process is repeated. Finally, since our goal in this paper is to study the maximum performances that can be achieved by CR-TFAA, we consider an ideal SIC receiver that is able to locate the replicas and estimate the signal parameters perfectly [3].

8.3 Performance Evaluation

8.3.1 Semi-analytical Approach

The method that we present here draws heavily on the framework developed by del Río *et al.* in [dD14] for the analysis of synchronous random access protocols exploiting SIC, as well as on its

[2] In [De +14] the sliding window length is denoted W. Here we add the subindex VF to avoid ambiguity with the system bandwidth W

[3] Note that the ideal channel estimation hypothesis is commonly adopted in the related literature (see for instance [CDH07; Kis11; De +14]).

adaptation for the time-asynchronous case (*i.e.*, ACRDA) provided in [De +14] by De Gaudenzi *et al.* We will further adapt del Río's framework in order to account for the time- and frequency-asynchronous case. As in [De +14], we simplify the analysis by considering only the perfect power control scenario in which the packets from different users are received with exactly the same power (*i.e.*, the channel gain is $a_k^2 = 1$). We provide a synthetic overview of the method: some details and the derivation of some expressions are omitted for the sake of brevity. The reader is referred to [dD14] and [De +14] for a more detailed analysis. We do emphasize the main distinctions with respect to the aforementioned frameworks.

In Del Río's framework, the packet error probability is conditioned by the occurrence of a *loop*. A loop event occurs when the all the replicas of a test packet collide with those of one or more users. A loop involving the packets from l users will be referred to as an l-loop. When a loop occurs, it is assumed that the test packet cannot be decoded through the SIC process but can eventually be recovered if the FEC decoder can cope with the MAI for at least one of the N_{rep} replicas. On the other hand, when there is no loop, the SIC process is triggered and the packet can eventually be recovered during one of the N_{it} iterations. Thus, the PER can be expressed as

$$\begin{aligned} P_e(G; N_{\text{rep}}, N_{t\times f}) \simeq [P_e^{(N_{it})}(G; N_{\text{rep}})]^{N_{\text{rep}}} P_{loop}(0; G, N_{\text{rep}}, N_{\text{slots}}) \\ + \sum_{l=1}^{\infty} [P_e(l|\text{loop})]^{N_{\text{rep}}} P_{loop}(l; G, N_{\text{rep}}, N_{\text{slots}}), \end{aligned} \tag{8.3}$$

where $P_e^{(N_{it})}(G; N_{\text{rep}})$ is the probability of finding one packet replica in error after the SIC process, $P_e(l|\text{loop})$ is the probability of finding one packet replica in error in the presence of an l-loop and $P_{loop}(l; G, N_{\text{rep}}, N_{\text{slots}})$ is the probability of an l-loop event. To facilitate the analysis, De Gaudenzi *et al.* proposed to assume that the first replica of each packet is transmitted on the first time slot of the VF. This way, the probability of having an l-loop given that the first replica of t newly-generated packets collide with the first replica of the test packet is

$$P_{loop}^T(l; t, p) = \binom{n}{i} p^l (1-p)^{t-l}, \tag{8.4}$$

where p is the probability that the remaining $N_{\text{rep}} - 1$ replicas of the colliding packets are transmitted using the same replica positions. For ACRDA, $p = 1/\binom{N_{\text{slots}}-1}{N_{\text{rep}}-1}$ while for CR-TFAA it is approximately $p \simeq 1/\binom{N_{t\times f}-1}{N_{\text{rep}}-1}$. Then, the overall l-loop probability is obtained by averaging Eq. (8.4) over all t:

$$P_{loop}(l; G, N_{\text{rep}}, N_{\text{slots}}) = \sum_{t=0}^{\infty} P_{loop}^T(l; t, p) f_K(t; \lambda_l). \tag{8.5}$$

In Eq. (8.5), $f_K(\cdot)$ is the Poisson probability density function and λ_l is the intensity of new packet arrivals whose first replica is transmitted in the vulnerability region of the test packet. Hence, for CR-TFAA we have $\lambda_l = 4G/\eta$.

Computing $P_e(l|\text{loop})$ in CR-TFAA is somewhat difficult. For ACRDA, it is assumed that the total interference power χ from the l overlapping packets forms an Irwin-Hall distribution and that, after the FEC deinterleaving stage, the resulting interference plus noise can be approximated as a Gaussian process. Then, $P_e(l|\text{loop})$ is obtained using a polynomial approximation of the turbo code packet error rate obtained by curve-fitting simulation data. As we have already shown in the previous Chapter, approximating the MAI interference as a Gaussian process in CR-TFAA provides inaccurate results. In addition, the model considered for ACRDA does not takes into account the impact of the pulse shaping filter. As a consequence, we compute the average probability of packet error $\hat{P}_e[k]$ for a varying number of colliding packets $k = 0, 1, ..., 20$ within the vulnerability area $[-T_p, T_p] \times [-B, B]$, using the simulation tool described in the previous chapter. Thus, $P_e(l|\text{loop}) = \hat{P}_e[k]$. Figure 8.2 shows the results obtained for $\hat{P}_e[k]$ for a particular PHY-layer configuration.

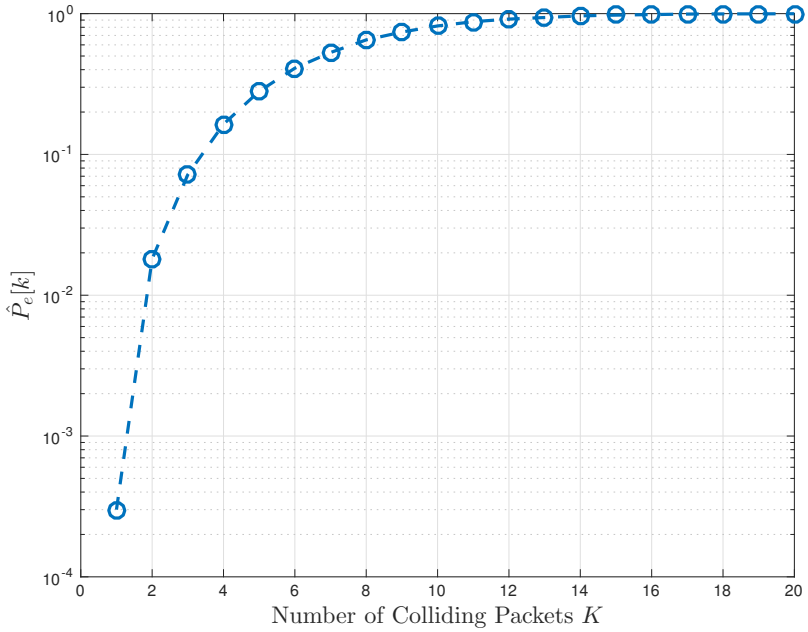


Figure 8.2 – Packet error probability $\hat{P}_e[k]$ vs number of colliding packets K obtained through simulation. QPSK modulation, $\beta = 0.2$, $E_s/N_0 = 10$ [dB], turbo FEC $\rho = 1/3$ and no power variations.

The term $P_e^{(N_{it})}(G; N_{\text{rep}})$ is given by

$$P_e^{(N_{it})}(G; N_{\text{rep}}) = \sum_{k=0}^{\infty} P_{loss}^{K,N_{it}}(k) f_K(k; \lambda_t) \tag{8.6}$$

where $\lambda_t = 4GN_{\text{rep}}/\eta$ is the intensity of the packet replica's arrival process over the vulnerability region of a test packet replica and is twice that of ACRDA; $P_{loss}^{K,N_{it}}(k)$ is the error probability after N_{it} SIC iterations given k colliding packets and equals

$$P_{loss}^{K,N_{it}}(k) = \sum_{r=0}^{k} P_{loss}^{R}(r) \binom{k}{r} q^r (1-q)^{k-r}, \tag{8.7}$$

with $q = [P_e^{(N_{it}-1)}(G; N_{\text{rep}})]^{N_{\text{rep}}-1}$. Note that since Eq. (8.6) is recursive, $P_e^{N_{it}}$ should be set to 1 for $N_{it} = 1$. Finally, the term $P_{loss}^{R}(r)$ is just the probability of a packet error in the presence of r interfering packets and, as $P_e(l|\text{loop})$, is obtained from simulation results, *i.e.*, $P_{loss}^{R}(r) = \hat{P}_e[r]$.

8.3.2 Simulation Approach

A computer simulator implementing in detail the model described above was written in Matlab®. The simulation approach is somewhat different to the one adopted for TFAA. In particular, transmitting a single test packet per realization is not sufficient since the likelihood of successfully decoding the packet also depends on the likelihood of decoding the other $N_{\text{rep}} - 1$ replicas as well as on the history of the SIC decoding process. As a consequence, a simulator realization is obtained by transmitting many consecutive VFs. The simulation step size is equivalent to 1 VF. At each simulation step, a random number of packets is generated according to a Poisson distribution of intensity

$$G_a = \left\lceil \frac{W}{B} \right\rceil N_{\text{slots}} G/\eta. \tag{8.8}$$

It is assumed that each of these packets is generated by a different user. Compared to the simulator implementation of [De +14] for ACRDA, the sampling granularity required for CR-TFAA is significantly higher. In fact, in [De +14] the time offset of a given packet is assumed to be in multiples of the symbol time T, therefore a single sample per symbol is required and a signal from a given packet replica k can be simply modelled as: $\boldsymbol{r_k} = \begin{bmatrix} d_k(0) & d_k(1) & \dots & d_k(L_p-1) \end{bmatrix}^T \times \exp[j\phi_k]$. Note that with this simplification the need for pulse shaping filtering is avoided. In our case, the pulse shaping filter has an impact on the performance, as shown in Chapter 7, so it has been included in the simulation model. As a consequence, the number of samples per symbol has been set according to the relation

$$\frac{T}{T_s} = \left\lceil \frac{W}{B} \right\rceil (1+\beta)L, \tag{8.9}$$

where T_s is the sample time and L is the upsampling rate. An advantage of this approach is that a higher granularity in the time offsets can be achieved; a replica can be transmitted at multiples of the sample time T_s. Also, it brings more accurate results. The main disadvantage is that the simulation time is considerable higher.

Thus, for each incoming packet, a VF is build and N_{rep} replicas are generated, each of which is FEC encoded, M-PSK modulated and filtered. The resulting complex baseband signal is then shifted in frequency through a complex carrier $\exp[j(2\pi f_k t + \phi_k)]$, with $f_k \sim \mathcal{U}[-f_m, f_M]$ and $\phi_k \sim \mathcal{U}[0, 2\pi]$. With respect to the VF starting time t_0, the replica is placed at a random multiple of the slot duration T_p. On the other hand, $t_0 \sim \mathcal{U}[t_{\text{sim}} - T_F, t_{\text{sim}} + T_F]$, t_{sim} being the current simulation time. Since we assume perfect channel estimation and replica detection, these random parameters are stored in a list that is accessible to the receiver module. The asynchronous VFs are combined in a channel buffer, where the log-normally distributed channel gains a_k^2's are applied and Gaussian noise is added to the resulting signal samples. The receiver now reads the last $[t_{\text{sim}} - T_F, t_{\text{sim}}]/T_s$ samples and proceeds with the iterative demodulation process until the sliding window spans all the new incoming samples. After that, the channel buffer is shifted by T_F/T_s samples and the overall simulator state is updated, which concludes the current simulation step. The process is repeated until a maximum of N_{frames} steps is reached. At this point, the PER is computed as the ratio between all successfully transmitted packets and the total number of transmitted packets.

8.4 Numerical Results

Next we present numerical results using the two performance evaluation methods described above. The main parameters used both for simulating the full system as well as for obtaining the probabilities $\hat{P}_e[k]$ for the semi-analytical method are given in Table 8.1.

8.4.1 CR-TFAA without diversity

When the repetition degree is set to $N_{\text{rep}} = 1$, the VF notion is no longer required (*i.e.* $N_{\text{slots}} = 1$) and, from the transmitter point of view, the protocol operates as pure ALOHA. This configuration is interesting because 1) it can be readily implemented in existing UNB systems since most of the additional signal processing is performed at the receiver side and 2) it provides advantages in terms of demodulator complexity as the need for replica channel estimation and interference cancellation is avoided. In addition, since in this case $T_F = T_p$, the sliding window only spans W_{VF} times the packet length, which is also beneficial in terms of demodulator complexity. In the simulations we considered BPSK modulation and $W/B = 10$. The rest of the parameters are those detailed in Table 8.1.

Parameter	**Value**
E_s/N_0	10 [dB]
Modulation	BPSK/QPSK
FEC rate ρ	$\simeq 1/3$
Encoder generator	(133, 171)
Turbo iterations	8
Constraint length K_L	7
Roll-off factor β	0.2
Filter span	4 [symbols]
Upsampling rate L	8
Payload length L_b	100 [b]
N_{it}	15
Sliding window size W_{VF}	3
Step size ΔW_{VF}	0.15

Table 8.1 – Simulation parameters.

The performance results obtained under this scenario are shown in Figure 8.3. We observe that a PER below 10^{-2} can be guaranteed for channel loads up to $G \simeq 0.8$ [b/s/Hz] under power imbalance conditions (*i.e.*, for $\sigma = 3$ [dB]) and $G \simeq 0.6$ [b/s/Hz] for perfect power control (*i.e.*, for $\sigma = 0$ [dB]). For a maximum PER of 10^{-3}, the achievable load is around $G \simeq 0.4$ [b/s/Hz] in both cases. These results suggest that the throughput of TFAA can be significantly improved by the use of simple SIC techniques and without requiring the transmission of additional packet replicas, which is explained by the fact that frequency randomization provides already a certain amount of diversity.

8.4.2 CR-TFAA with time-frequency diversity

Our interest in this section is to evaluate the performance of CR-TFAA exploiting time–frequency diversity and to contrast these results with those obtained with ACRDA. We observe that if the W/B ratio is set to 1, then $f_m = f_M = 0$. In this case, all replicas are transmitted over the same frequency $f = 0$ and the protocol operates exactly as ACRDA (thus, ACRDA can be considered just as a special case of CR-TFAA). Simulation results for CR-TFAA were obtained using $W/B = 5$ and $N_{\text{slots}} = 20$, while for ACRDA $W/B = 1$ and $N_{\text{slots}} = 100$. In both cases, $N_{\text{rep}} = 2$ and QPSK

modulation is adopted. Note that, in order to provide a fair comparison, approximately the same frame size has been considered for the two schemes ($N_{t\times f} = N_{\text{slots}} \times W/B = 100$). The results are presented in Figure 8.4, where it can be observed that both protocols provide very similar performances. In particular, the critical points (*i.e.* the values of G over which the system performance drops quickly) are very close. In terms of maximum achievable throughput, CR-TFAA is slightly below ACRDA under power imbalance conditions and is practically identical under perfect power control. It is worth noting, however, that CR-TFAA provides a PER performance that is at least about one order of magnitude below that of ACRDA for channel loads below $G \simeq 0.8$ [b/s/Hz] and $G \simeq 1.4$ [b/s/Hz] for $\sigma = 0$ and $\sigma = 3$ [dB], respectively. In other words, the PER performance of CR-TFAA is superior over practical channel loads (*i.e.* below the critical points).

To further validate these results, in Figure 8.5 the performances obtained through simulations are compared with those given by the semi-analytical method for the perfect power control regime. As for ACRDA, we observe that the semi-analytical approach provides a lower bound for the PER—which is probably due to the simplifications in the loop probability analysis—but provides a good description of the protocol behaviour. Also, it confirms that the CR-TFAA PER is improved with respect to ACRDA.

8.5 Summary

In this chapter, we introduced and evaluated the throughput performance of CR-TFAA, a random access scheme that combines SIC techniques with TFAA (which we described in previous chapters). This protocol can be regarded as a generalization of ACRDA and aims at improving the performance of LPWANs based on UNB. Numerical results for two main scenarios have been presented, namely with and without time–frequency diversity. In both cases, it is shown that CR-TFAA significantly improves the performance of TFAA. For the scenario including time–frequency diversity, the CR-TFAA performances were also compared with those of ACRDA. In particular, it was shown that CR-TFAA provides better PER performance for practical channel loads.

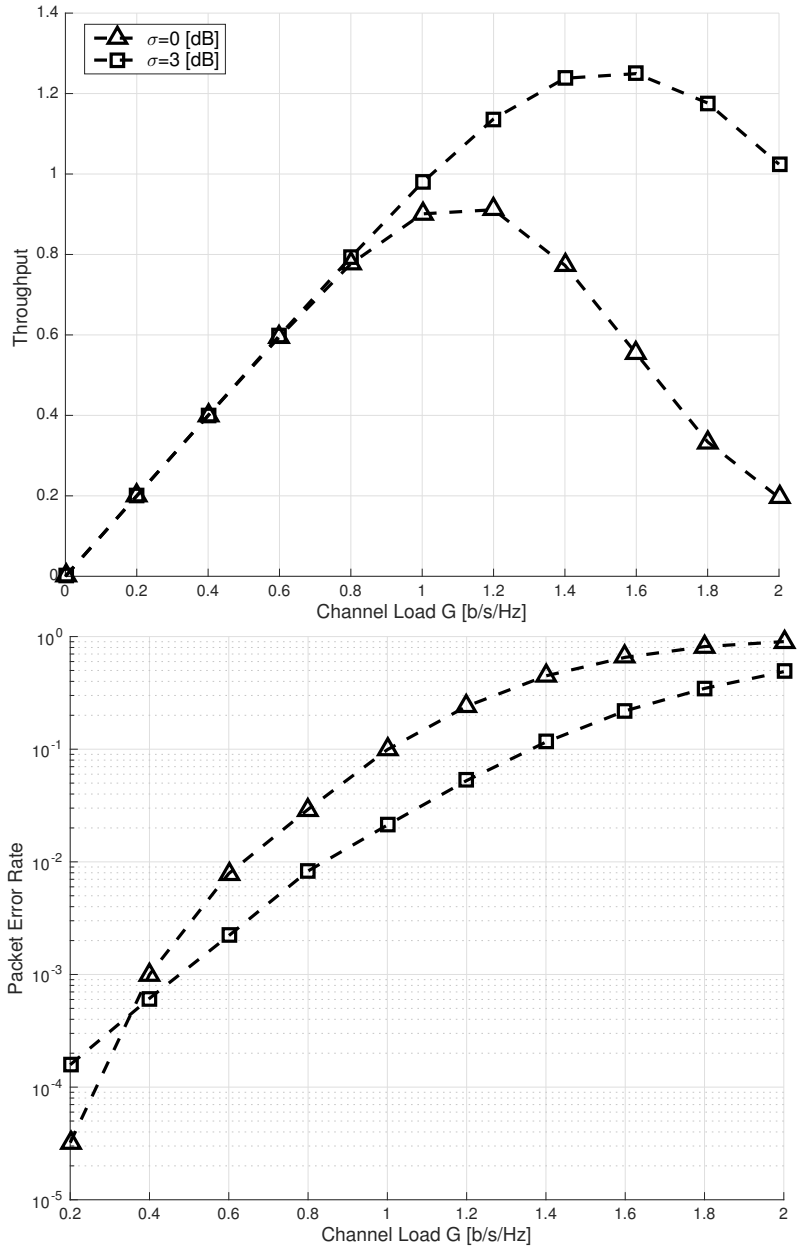


Figure 8.3 – Throughput and PER performance of CR-TFAA without time-frequency diversity ($W/B = 10$, $N_{\text{rep}} = 1$, $N_{\text{slots}} = 1$, BPSK modulation, $E_s/N_0 = 10$ dB and turbo FEC $\rho = 1/3$).

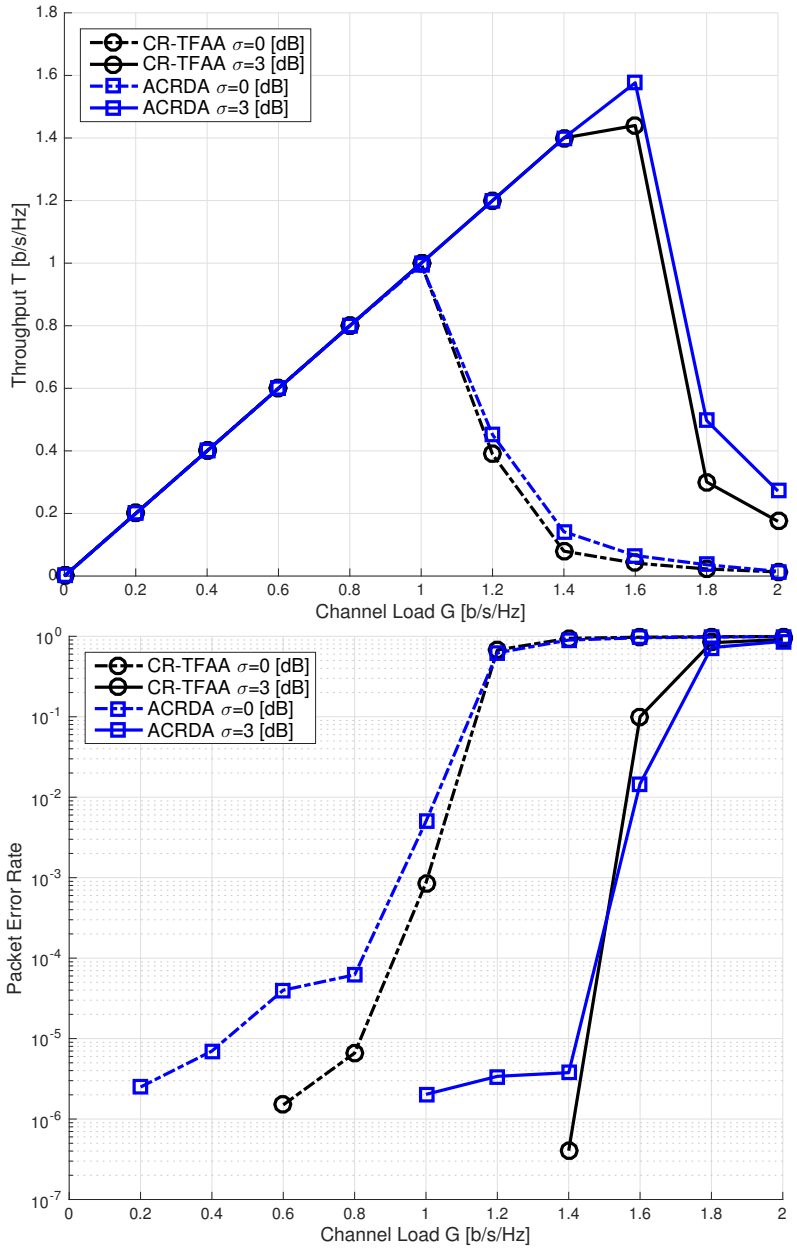


Figure 8.4 – Throughput and PER performance of CR-TFAA with time-frequency diversity ($W/B = 5$, $N_{\text{rep}} = 2$, $N_{\text{slots}} = 20$, QPSK modulation, $E_s/N_0 = 10$ [dB] and turbo FEC $\rho = 1/3$) and ACRDA (only two parameters are changed in this case, namely $W/B = 1$ and $N_{\text{slots}} = 100$).

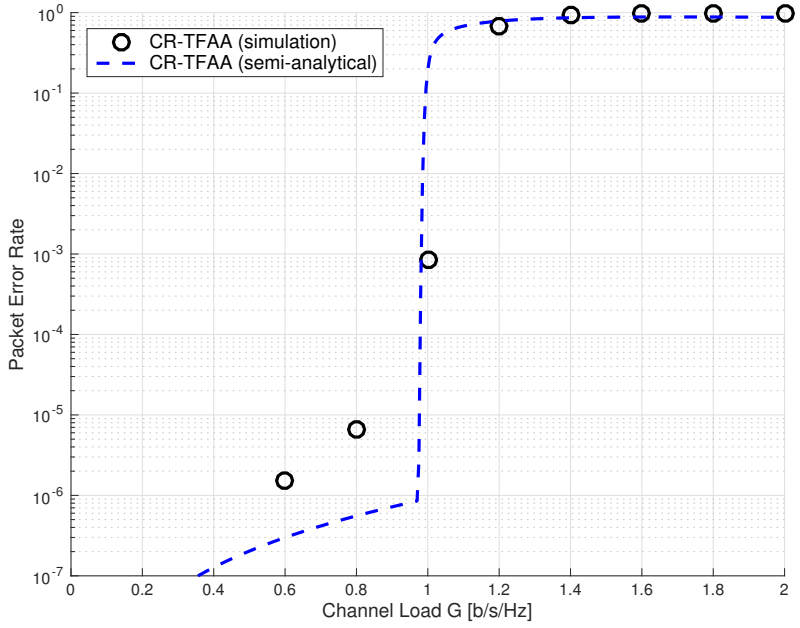


Figure 8.5 – Simulation vs semi-analytical performance of CR-TFAA with time-frequency diversity ($W/B = 5$, $N_{\text{rep}} = 2$, $N_{\text{slots}} = 20$, QPSK modulation, $E_s/N_0 = 10$ [dB] and turbo FEC $\rho = 1/3$).

"Such a chimerical idea as telegraphing vocal sounds would indeed, to most minds, seem scarcely feasible enough to spend time in working over."

Alexander Graham Bell (1847 – 1922)

9 Summary and Future Work

The subjects treated on this thesis were the result of an overall study on small satellite communication systems engineering, specially regarding store-and-forward data collection applications. Further, we focused on a particular class of services oriented to machine-type communications, which imposes additional constraints in the design of both the node terminals and the satellite communications payload.

We have seen that these systems are rather designed following very simple architectures, due to economical and physical constraints. As we move towards more advanced nanosatellite missions aiming to provide commercial services that can compete with traditional space systems, the need for novel and more efficient communications solutions becomes apparent. In this context, current advances in the field of Low Power Wide Area Networks (LPWAN) may bring new insight into this issue.

The contributions presented in this thesis hopefully represent a small step forward in this direction. The main products of this research are the performance evaluation of time- and frequency-asynchronous ALOHA (TFAA) and the introduction of contention resolution time- and frequency-asynchronous ALOHA—two random access schemes that are suitable for LPWANs based on ultra narrowband (UNB) modulation. In addition, a simple link-layer architecture wherein the packet collisions are partially resolved at the gateway ground stations has been proposed and evaluated, which can lead to interesting improvements.

TFAA is potentially a good solution for providing a simple access scheme for the transmission of short packets in a LEO satellite data collection network. In particular, adopting a UNB/TFAA

link design allows for long range communications, flat fading operation, improved energy efficiency, cheaper transmitters, among other interesting features. Even though preliminary analysis show that TFAA provides low-throughput performance, we have shown through a more detailed analysis that the performance gap with respect to the ALOHA protocol is rather small. Further, we have seen that it can be more easily enhanced with the help of low-rate FEC, leading to even higher performances. Perhaps the more interesting contribution regarding UNB/TFAA was the study of the impact of PHY-layer parameters such as modulation, channel coding, pulse shaping filtering and channel gain variations. Also, a distinctive characteristic of this study with respect to related works is that it is not based on the so-called Gaussian assumption for modelling the interference process, which may lead to more accurate results. Overall, we expect that our study of TFAA would be useful for: 1) a better design and dimensioning of narrowband and UNB networks using this access approach, 2) optimising the MAC performance of TFAA-based networks through improved PHY designs (for instance, by adopting an optimal pulse shaping filter) and 3) getting a better description of its performance, which is in general underestimated, as simple PHY-independent models (such as the collision channel model) are mostly adopted.

CR-TFAA was proposed as an enhancement of TFAA. The design follows current trends of high-throughput random access schemes: it combines a successive interference cancellation (SIC) receiver architecture, low-rate FEC, time–frequency diversity and the basic channel access approach of TFAA. It was shown that CR-TFAA allows to increase the throughput and packet error rate (PER) performance of TFAA significantly. In particular, packet error rates in the order of 10^{-7}–10^{-6} can be achieved. This characteristic is essential for providing reliability in the context of machine-type communications.

A multiple access scheme suitable for non-real-time data collection networks based on power-constrained nanosatellite platforms was proposed. The main advantage of this approach is that it allows to improve the MAC throughput without involving additional protocol logic/signal processing at the satellite side. Thus, this solution could be easily integrated in systems where the satellite is highly limited in terms of processing power. While it was shown that this approach allows to improve the throughput–PER performance, it comes at the cost of transmitter energy efficiency.

We conclude this thesis by proposing some directions for future work.

9.1 Future Work

9.1.1 UNB/TFAA Issues in the LEO Environment

The use of UNB modulation with TFAA in a LEO satellite communication system involves several practical issues that have not been considered in this thesis. While some of these are in general well-known problems in communications engineering, the high Doppler effects observed in the LEO

channel may bring additional difficulties.

Signal Parameter Estimation

Our analysis of TFAA for the PHY reception model was based on the hypothesis that the receiver was able to estimate the signal parameters (*i.e.*, frequency, initial phase and timing) of each incoming packet perfectly. While some justification for this hypothesis was provided, the actual impact of parameter estimation errors on the MAC performance was not evaluated. The reason for not having considered this issue is mainly that the performance of the signal parameter estimation module depends heavily on the actual system implementation (*e.g.*, the design of the preamble sequences and the estimation algorithm). Our intention was to provide a more generic model framework that could be adapted to more specific scenarios and implementations. In mobile terrestrial networks where the transmitter–receiver relative velocity is low, the overall PER can actually be dominated by decoding errors, as we assumed in our analysis. However, in LEO satellite networks the Doppler effect can be significant, specially under UNB operation and high frequency bands. In this scenario, the following issues should be analysed in more detail in future work:

- ***Frequency tracking***. The frequency of a TFAA signal transmitted through an ideal band-pass channel is $f_0 = f_k + f_c$, where f_k represents the packet frequency location within the baseband bandwidth W and f_c is the carrier frequency. In the LEO channel, the signal frequency observed by the receiver is actually $f_0(t) = f_k + f_c + f_D(t) + \delta_f$, where $f_D(t)$ is the instantaneous Doppler shift and δ_f is a term that accounts for the oscillator instability. If δ_f is high with respect to the packet bandwidth B, there is high uncertainty in the output of the frequency tracking module, which may result in performance degradation. In addition, for a UNB LEO system operating at UHF or above, the frequency variation during the packet transmission is comparable to B. An accurate, adaptive frequency tracking system is therefore required in order to demodulate the signal correctly. An alternative to developing a powerful frequency tracking system consist on using a Doppler precompensation mechanism at the transmitter side, which is also an interesting subject for future work.

- ***Timing Estimation***. The Doppler effect not only affects the high-frequency component of the signal but also all its other components. As a result, the information-bearing signal component is also distorted. Due to the relatively long transmission time T_p of UNB packets (*e.g.*, $T_p = 1$ [s] for a 100-bits packet and a bit rate of 100 [bps]), the variation of the symbol length during T_p can be significant. Hence, an accurate timing estimation algorithm must also be conceived in order to avoid sampling errors.

Doppler Effects on the MAI level and the Collision Probability

From the discussion above, it is clear that a long packet transmission time results in a significant net frequency change when high Doppler effects are observed. As a consequence, the time–frequency space occupied by a packet is no longer rectangular—as we assumed in our analysis—but rather trapezoidal. Note that this also means that the packet vulnerability has a trapezoidal form. This effect is studied in [Ant+15] for the collision channel model. However, as we have seen, this model does not describes well the actual MAI characteristics, which leads to pessimistic results in general. A more detailed analysis of this issue is therefore required for an accurate evaluation of the MAC performance of UNB/TFAA in the LEO environment.

9.1.2 Optimal Data Rate for Energy-Efficient Machine-Type TFAA Networks

Our analysis of TFAA is not only valid for UNB networks but also for narrowband networks–as long as the packet bandwidth is small compared to the available system bandwidth and assuming that the Doppler effects can be compensated. In a scenario wherein these hypothesis are valid, the MAC performance of TFAA does not depend on the actual signal bandwidth. Hence, the transmission data rate can be adjusted following different design goals, such as optimising the terminal energy efficiency. The optimal data rate can be found by taking into account both the transmission and the circuit power on the one hand, and the modulation scheme on the other [Mia+09]. In systems where the Doppler effect is important, the data rate can also be used as an optimisation variable to minimise its impact (since shorter packets are less distorted). In any case, adjusting the data rate involves a trade-off with transmission power and communication range.

9.1.3 TFAA Decoder Design

The transmission of packets in random frequency locations has also an impact on the decoder complexity as packets must be sought inside the whole system bandwidth. This issue brings several research opportunities. For instance, we may mention the design of an optimized TFAA packet detector and the design of optimized pilot sequences allowing to detect packets even in the presence of collisions.

9.1.4 Channel Estimation Error in CR-TFAA

Naturally, the same recommendations for future work related to TFAA are also useful for a more detailed analysis of CR-TFAA. In addition, channel estimation errors have also an impact on the SIC process. For instance, when a replica is decoded, the receiver must then estimate the signal parameters of its twin, which might be affected severely by interference. If the signal parameters

are not correctly estimated, the replica cannot be entirely cancelled and residual signal components will remain in the receiver buffer. Although this is a well-known problem in random access schemes combining SIC and diversity, the specificities of TFAA may involve additional issues. In particular, for UNB signals the uncertainty on the carrier frequency is an aspect to take into account.

"La dernière chose que nous attendions de vous, Général, est une leçon de géométrie !"

Pierre-Simon Laplace (1749 – 1827)

10 Résumé de la Thèse

Le nombre de petits satellites lancés dans l'espace chaque année a grandi de manière exponentielle au cours de ces deux dernières décennies. Ceci a été possible en grand partie grâce à l'introduction des plate-formes *CubeSats*. Ce standard a notamment permis de diminuer le coût minimum nécessaire de l'accès à l'espace. Ainsi, de nombreuses missions ont pu être développées par plusieurs organisations dans le monde entier, souvent au travers de projets universitaires à petit budget. Cette nouvelle tendance constitue un changement de paradigme important; en effet, le développement de missions spatiales était traditionnellement et exclusivement mené par des agences gouvernementales, à des coûts très importants.

Si les premières missions CubeSat étaient consacrées à des simples expériences académiques ou à tester de nouveaux composants électroniques, il est aujourd'hui possible d'envisager de nouvelles applications et services avec une plus forte valeur ajoutée. Le succès du standard CubeSat a certainement donné lieu à la création d'un écosystème technologique assez dynamique reliant des start-ups, des universités, des centres de recherche et des agences spatiales. À long terme, il est fort probable que les CubeSats (ou éventuellement d'autres nouveaux formats de petits satellites) joueront un rôle actif dans différents domaines du secteur spatial.

Cette thèse fait partie des projets de recherche et de développement menés au sein du Centre Spatial Universitaire Montpellier-Nîmes ayant pour objectif de concevoir de nouvelles applications basées sur des nanosatellites de type CubeSat. Plus spécifiquement, notre travail porte sur les réseaux globaux de collecte de données par petits satellites à défilement. De plus, nous sommes intéressés par une sous-classe de système orientée données de type *machine-to-machine* (M2M).

Dans ce contexte, les nœuds (appelés aussi *terminaux*) sont caractérisés par une petite taille, une puissance faible et un coût de fabrication extrêmement bas, ce qui rend difficile l'établissement d'un lien de communication robuste.

Bien que des microsatellites en orbite basse (LEO) de type *store-and-forward* aient été développés pour des applications similaires à partir des années 1990 et jusqu'au début des années 2000, la demande croissante de services de type M2M stimule actuellement le développement d'une nouvelle génération de réseaux de collecte de données basés sur des nanosatellites *low-cost*. Dans une perspective plus large, ce type de réseaux pourrait représenter une part importante des systèmes qui intégreront l'internet des objets (IoT, de l'anglais *internet of things*) dans le futur.

Toutefois, pour que ces attentes puissent être réalisées, plusieurs défis doivent être relevés. D'une part, les limitations en termes de masse et taille des nanosatellites ont un impact sur tous les sous-systèmes (contrôle d'attitude, puissance, antenne, etc.). Concernant le sous-système (ou charge utile) de communication, ces contraintes se traduisent notamment par un gain d'antenne réduit et une faible puissance de transmission et de traitement. De plus, des contraintes similaires sont observées dans les terminaux terrestres. D'autre part, le canal de transmission LEO comporte des difficultés supplémentaires qui doivent être surmontées pour la conception d'un lien de communication robuste. Enfin, en raison du grand nombre de terminaux partageant le lien montant de manière simultanée, celui-ci mérite une attention particulière.

Actuellement, la majorité des projets CubeSats utilise des composants dits *commercial-of-the-shelf* (en français, *composant pris sur étagère*), particulièrement pour le système de communications. Bien évidement, ce type de solutions n'est pas adapté aux systèmes orientés communications M2M. Il est donc nécessaire de concevoir de nouvelles solutions qui prennent en compte les caractéristiques et contraintes desdits systèmes.

Avec la prolifération des réseaux étendus à faible consommation énergétique (LPWAN, de l'anglais *low power wide area networks*) dans le secteur terrestre, des nouveaux paradigmes de conception ont été introduits. Au lieu de maximiser la capacité (*i.e.*, le débit binaire) réalisable par chaque terminal—ce qui a été le cas dans les réseaux terrestres conventionnels—ces réseaux sont orientés à garantir des liens à longue portée en optimisant l'efficacité énergétique des terminaux. Cette approche permet donc d'envisager des nouveaux types applications, notamment dans le contexte IoT/M2M. Il apparaît clairement que les réseaux terrestres LPWAN ont plusieurs points en commun avec les systèmes de collecte de données que nous étudions ici—en fait, certaines architectures LPWAN peuvent aussi être adaptées pour être utilisées dans l'environnement LEO. En conséquence, nous allons étudier dans cette thèse une technique LPWAN connue sous le nom de modulation à bande ultra étroite (UNB, de l'anglais *ultra narrow band*).

10.1 Présentation de la problématique

Les modems les plus utilisés dans des CubeSats sont basés sur le protocole de couche liaison AX.25 et sur des modulations binaires classiques, comme par exemple, la modulation par déplacement de fréquence (FSK) ou la modulation par déplacement de phase (PSK). Ces systèmes permettent d'établir des liens de transmission relativement robustes, avec un taux binaire de l'ordre de 1200–9600 [bps]. Par contre, ils ne sont pas optimaux dans le contexte M2M comprenant un nombre élevé d'utilisateurs que nous avons décrit plus haut. De plus, le trafic de données dans ces réseaux est souvent caractérisé par des petits messages arrivant au système de manière imprévisible, ce qui fait appel à l'utilisation de méthodes d'accès aléatoire. Dans ces conditions, le défi consiste à concevoir une architecture permettant de garantir un lien montant fiable avec une haute efficacité spectrale, tout en conservant un coût, une complexité et une consommation énergétique faible. Le travail développé dans cette thèse constitue un petit pas dans cette direction.

En raison des diverses contraintes mentionnées ci-dessus, la conception d'un système de communication optimal nécessite une approche *cross-layer* (CLD, de l'anglais *cross-layer design*); c'est-à-dire, une vision globale de l'ensemble des couches et sous-couches. Plus spécifiquement, la conception des couches physiques (PHY) et de contrôle d'accès au support (MAC, de l'anglais *Media Access Control*) doit être menée de manière conjointe.

C'est en prenant en compte ces principes que la technique UNB semble représenter une solution possible. En effet, elle offre, en principe, plusieurs des avantages recherchées : *e.g.*, un coût de terminal réduit, longue portée, efficacité énergétique et accès aléatoire. De plus, les réseaux UNB ont une caractéristique particulière: les paquets peuvent être transmis de manière complètement asynchrone tant dans le domaine temporel que dans le domaine fréquentiel. Malgré ces avantages, une analyse complète en termes de taux d'utilisation du canal (ou *throughput*) et de taux d'erreur de paquet au niveau MAC n'était pas disponible dans la littérature. Ainsi, l'évaluation de ces facteurs de mérite—essentiels pour un dimensionnement de réseau plus précis—constitue un des premiers objectifs de cette thèse.

Comme indiqué plus haut, l'accès au canal physique dans les réseaux UNB est de type aléatoire. Par conséquence, une grande partie des paquets transmis est perdue à cause des collisions (*i.e.*, lorsque deux ou plus paquets sont transmis de manière simultanée sur des fréquences assez proches), ce qui compromet le throughput et l'efficacité spectrale du système. Ceci correspond à notre deuxième problématique: l'optimisation des performances MAC dans des réseaux où les paquets sont transmis de manière asynchrone dans l'espace temps–fréquence.

Traditionnellement, la résolution des collisions dans les systèmes d'accès aléatoire est faite au travers de mécanismes au niveau couche MAC. Parmi quelques exemples, nous pouvons citer des algorithmes comme le *binary exponential backoff* ou le *tree algorithm*. Si ces solutions permettent de résoudre les collisions tout en garantissant une opération stable, elles ne sont pas capables d'atteindre un throughput élevé. De plus, la résolution des collisions au niveau MAC n'est pas

efficace en termes énergétiques en raison du haut nombre de paquets qui doivent être retransmis. Par conséquence, ces solutions ne sont pas adaptées pour les réseaux M2M. C'est dans ce contexte que l'approche CLD dévient pertinente; en effet, de nouvelles techniques permettant de résoudre les collisions dans la couche PHY ont été introduites récemment. L'évaluation des performances au niveau MAC qui peuvent être obtenues dans les réseaux UNB en suivant ce principe est aussi un problème que nous adressons dans cette thèse.

10.2 Principales contributions

10.2.1 Évaluation des performances du protocole TFAA

La méthode d'accès au canal adopté dans les réseaux UNB est ici désignée de manière plus générale sous le nom *time- and frequency-asynchronous ALOHA* (ALOHA). Une partie important du travail effectué au cours de cette thèse a été consacrée à l'étude de cette technique. Trois différents scénarios sont considérés: le modèle de collisions, le modèle avec *capture* basé sur le rapport signal à interférence plus bruit et un modèle plus détaillé qui prend en compte les paramètres de la couche PHY. Les deux premiers modèles sont donc basés sur une abstraction de la couche PHY et ont été traditionnellement utilisés dans l'étude des méthodes d'accès aléatoire en raison de leur simplicité. Si les performances obtenues pour ces modèles sont limitées, nous montrons que ces résultats sont plutôt pessimistes car ils ne prennent pas en compte les variations de l'interférence dans l'espace temps-fréquence—d'où l'importance du troisième modèle. Ce dernier est aussi utilisé pour étudier les effets de la modulation, du codage canal, du filtrage de mise en forme et des variations du gain du canal. Les principaux résultats obtenus dans l'étude de la méthode TFAA ont été publiés dans [AF17c].

10.2.2 CR-TFAA: une méthode à haut rendement basée sur des techniques SIC

Nous proposons et étudions une variation de la méthode TFAA appelée *Contention Resolution Time- and Frequency Asynchronous ALOHA* (CR-TFAA). L'idée centrale ici est d'améliorer les performances de TFAA au travers d'un mécanisme de résolution des collisions au niveau de la couche PHY. Ce mécanisme est basé sur des nouvelles méthodes d'accès (notamment sur les concepts introduits dans CRDSA [CDH07] et ensuite dans ACRDA [De +14]) et intègre des techniques de suppression successive d'interférences (SIC, de l'anglais *successive interference cancellation*) ainsi que des techniques de diversité en temps–fréquence. Nous évaluons les performances MAC de CR-TFAA par une méthode semi-analytique et par simulations. En comparant CR-TFAA avec ACRDA dans exactement les mêmes hypothèses, nous montrons que le premier peut atteindre notamment des taux d'erreur de paquet encore plus faibles—une caractéristique essentielle pour

concevoir des services à haute fiabilité dans le contexte M2M. Les résultats obtenus dans cet étude ont été publiés dans [AF17a].

10.2.3 Une architecture simple exploitant le compromis “performance–délai de bout-en-bout”

Les systèmes de collecte de données comprenant un nombre faible de satellites ne peuvent pas offrir de services de communications en temps réel. Les applications visées par ces systèmes sont donc tolérantes à de délais de transmission de bout-en-bout relativement élevés. Nous étudions des mécanismes simples permettant d'améliorer la performance du système en exploitant cette caractéristique et sans introduire de la complexité additionnelle dans le satellite. En particulière, nous étudions l'utilisation du codage paquet de bout-en-bout dans le contexte de l'accès aléatoire. Les discussions issues de cet approche ont été publiées dans [AF17b].

Bibliography

[3gp] *Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding*. Technical specification 36.212. 3GPP, 2017.

[AADH98] I. Ali, N. Al-Dhahir, and J.E. Hershey. "Doppler characterization for LEO satellites". In: *IEEE Transactions on Communications* 46.3 (Mar. 1998), pp. 309–313.

[AADH99] I. Ali, N. Al-Dhahir, and J.E. Hershey. "Predicting the visibility of LEO satellites". In: *IEEE Transactions on Aerospace and Electronic Systems* 35.4 (1999), pp. 1183–1190.

[AB87] J. Arnbak and W. van Blitterswijk. "Capacity of Slotted ALOHA in Rayleigh-Fading Channels". In: *IEEE Journal on Selected Areas in Communications* 5.2 (Feb. 1987), pp. 261–269.

[Abi] *The Internet of Things Will Drive Wireless Connected Devices to 40.9 Billion in 2020*. Available: `https://www.abiresearch.com/press/the-internet-of-things-will-drive-wireless-connect/`. Accessed: 2015-12-12. 2014.

[Abr70] Norman Abramson. "The ALOHA System – Another Alternative for Computer Communication". In: *Proc. the AFIPS Fall Joint Computer Conference*. Vol. 37. 1970, pp. 281–285.

[Abr77] N. Abramson. "The Throughput of Packet Broadcasting Channels". In: *IEEE Transactions on Communications* 25.1 (Jan. 1977), pp. 117–128.

[Abr90] N. Abramson. "VSAT data networks". In: *Proceedings of the IEEE* 78.7 (1990), pp. 1267–1274.

[Abr94] N. Abramson. "Multiple access in wireless digital networks". In: *Proceedings of the IEEE* 82.9 (1994), pp. 1360–1370.

[AF16] V. Almonacid and L. Franck. "Extending the Coverage for the Internet of Things with Low-cost Nanosatellite Networks". In: *CUBSAT 2016 : 2nd IAA Latin American CubeSat Workshop*. Flioranopolis, Brazil, Feb. 2016.

[AF17a] V. Almonacid and L. Franck. “An asynchronous high-throughput random access protocol for low power wide area networks”. In: *IEEE International Conference on Communications*. 2017.

[AF17b] V. Almonacid and L. Franck. “Extending the coverage of the internet of things with low-cost nanosatellite networks”. In: *Acta Astronautica* 138 (2017).

[AF17c] V. Almonacid and L. Franck. “Throughput Performance of Time- and Frequency-Asynchronous ALOHA”. In: *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*. 2017, pp. 1–6.

[AKZ07] A. Addaim, A. Kherras, and B. Zantou. “Design of Store and Forward Data Collection Low-cost Nanosatellite”. In: *2007 IEEE Aerospace Conference*. IEEE, 2007, pp. 1–10.

[AKZ08] A. Addaim, Abdelhak Kherras, and E. B. Zantou. “Design and Analysis of Store-and-Forward Data Collection Network using Low-cost LEO Nanosatellite and Intelligent Terminals”. In: *Journal of Aerospace Computing, Information, and Communication* 5.2 (2008), pp. 35–46.

[Ali+98] I. Ali et al. “The value of Doppler information for multiple access and power control in LEO satellite systems”. In: *Conference Record of Thirty-Second Asilomar Conference on Signals, Systems and Computers (Cat. No.98CH36284)*. Vol. 2. IEEE, 1998, pp. 1788–1792.

[All95] M.N. Allery. “Low Earth orbit microsatellites for data communications using small terminals”. In: *10th International Conference on Digital Satellite Communications (ICDSC-10)*. Vol. 1995. IEE, 1995, pp. 457–465.

[Ant+15] Mehdi Anteur et al. “Ultra Narrow Band Technique for Low Power Wide Area Communications”. In: *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.

[Axf92] R. A. Axford. *Effects of CW- and BPSK-Signal Interference on a Standard BPSK Digital Communications System*. Tech. rep. 1510. San Diego, CA: NCCOSC, 1992, p. 48.

[Bab+13] Alessandra Babuscia et al. “CommCube 1 and 2: A CubeSat series of missions to enhance communication capabilities for CubeSat”. In: *2013 IEEE Aerospace Conference*. IEEE, 2013, pp. 1–19.

[Bed13] Hector Bedon. “A DTN System for Nanosatellite-based Sensor Networks using a New ALOHA Multiple Access with Gateway Priority”. In: *The Smart Computing Review* 3.5 (2013).

[BG92] Dimitri P. Bertsekas and Robert G. Gallager. *Data networks*. Prentice Hall, 1992, p. 556.

[BH] R.L. Borchardt and T.T. Ha. “Power capture ALOHA”. In: *MILCOM 88, 21st Century Military Communications - What's Possible?'. Conference record. Military Communications Conference*. IEEE, pp. 703–707.

[Bra73] William T. Brandon. “A data courier satellite system concept”. In: *International Journal of Satellite Communications* 12.6 (1973), pp. 569–578.

[Bur13] Roger Chapman Burk. “Closed-Form Approximation of Revisit Rate for Low-Altitude Satellites”. en. In: *Journal of Spacecraft and Rockets* 50.4 (2013), pp. 872–883.

[Ccs] *CCSDS File Delivery Protocol*. Blue Book 4. CCSDS, 2007.

[CDH07] E. Casini, R. De Gaudenzi, and Od.R. Herrero. “Contention Resolution Diversity Slotted ALOHA (CRDSA): An Enhanced Random Access Schemefor Satellite Access Packet Networks”. In: *IEEE Transactions on Wireless Communications* 6.4 (Apr. 2007), pp. 1408–1419.

[Che+11] Yan Chen et al. “Fundamental trade-offs on green wireless networks”. English. In: *IEEE Communications Magazine* 49.6 (2011), pp. 30–37.

[Che89] D.K. Cheng. *Field and wave electromagnetics*. The Addison-Wesley series in electrical engineering. Addison-Wesley Publishing Company, 1989.

[Che99] Baozhong Cheng. “An optimised network protocol (LEONET) for efficient data store-&-forward communications using a LEO satellite”. Thesis submitted for the Degree of Doctor of Philosophy, University of Surrey. Copyright remains with the author. PhD thesis. University of Surrey, 1999.

[Chi] M. Chiani. “Performance of BPSK and GMSK with multiple cochannel interferers”. In: *Proceedings of PIMRC '96 - 7th International Symposium on Personal, Indoor, and Mobile Communications*. Vol. 3. IEEE, pp. 833–837.

[Chi97] M. Chiani. “Analytical distribution of linearly modulated cochannel interferers”. In: *IEEE Transactions on Communications* 45.1 (1997), pp. 73–79.

[CKM16] Federico Clazzer, Christian Kissling, and Mario Marchese. “Exploiting Combination Techniques in Random Access MAC Protocols: Enhanced Contention Resolution ALOHA”. In: (2016). arXiv: `1602.07636`.

[Cla+09] Christopher Clark et al. “CubeSat communications transceiver for increased data throughput”. English. In: *2009 IEEE Aerospace conference*. IEEE, Mar. 2009, pp. 1–5.

[Cla89] D.D. Clark. “Overview Of The Argos System”. In: *OCEANS '89. Proceedings*. Vol. 3. IEEE, 1989, pp. 934–939.

[Cor07] Giovanni E. Corazza. *Digital satellite communications*. Springer, 2007, p. 535.

[CR83] G. Choudhury and S. Rappaport. “Diversity ALOHA–A Random Access Scheme for Satellite Communications”. In: *IEEE Transactions on Communications* 31.3 (Mar. 1983), pp. 450–457.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[CV94] G.E. Corazza and F. Vatalaro. “A statistical model for land mobile satellite channels and its application to nongeostationary orbit systems”. In: *IEEE Transactions on Vehicular Technology* 43.3 (1994), pp. 738–742.

[DD12] Oscar Del Rio Herrero and Riccardo De Gaudenzi. “High Efficiency Satellite Multiple Access Scheme for Machine-to-Machine Communications”. In: *IEEE Transactions on Aerospace and Electronic Systems* 48.4 (Oct. 2012), pp. 2961–2989.

[dD14] Oscar del Rio Herrero and Riccardo De Gaudenzi. “Generalized Analytical Framework for the Performance Assessment of Slotted Random Access Protocols”. In: *IEEE Transactions on Wireless Communications* 13.2 (2014), pp. 809–821.

[DGG14a] Minh-Tien Do, Claire Goursaud, and Jean-Marie Gorce. “Interference Modelling and Analysis of Random FDMA scheme in Ultra Narrowband Networks”. en. In: *AICT 2014*. 2014.

[DGG14b] Minh-Tien Do, Claire Goursaud, and Jean-Marie Gorce. “On the benefits of random FDMA schemes in ultra narrow band networks”. English. In: *2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*. IEEE, 2014, pp. 672–677.

[DS04] G. Dimic and N.D. Sidiropoulos. “Multiaccess queuing from a signal processing perspective - Medium access control-physical cross-layer design”. In: *IEEE Signal Processing Magazine* 21.5 (Sept. 2004), pp. 40–50.

[Dvb] *DVB-RCS2 Lower Layer Satellite Specification*. Available: `https://www.dvb.org/standards/dvb-rcs2`. Accessed: 2017-04-18.

[EV89] A. Ephremides and S. Verdu. “Control and optimization methods in communication network problems”. In: *IEEE Transactions on Automatic Control* 34.9 (1989), pp. 930–942.

[Far11] Fares Fares. “Traitement des signaux Argos 4”. PhD thesis. University of Toulouse, 2011.

[Fon+01] F.P. Fontan et al. “Statistical modeling of the LMS channel”. In: *IEEE Transactions on Vehicular Technology* 50.6 (2001), pp. 1549–1567.

[GC05] A. Giorgetti and M. Chiani. “Influence of fading on the Gaussian approximation for BPSK and QPSK with asynchronous cochannel interference”. In: *IEEE Transactions on Wireless Communications* 4.2 (2005), pp. 384–389.

[GG15] C. Goursaud and J. M. Gorce. “Dedicated networks for IoT: PHY / MAC state of the art and challenges”. In: *EAI Endorsed Transactions on Internet of Things* 1.1 (2015), p. 150597.

[GK00] P. Gupta and P.R. Kumar. “The capacity of wireless networks”. In: *IEEE Transactions on Information Theory* 46.2 (2000), pp. 388–404.

[GL92] A. Ginari and G. Lehmann. “SAFlR: A Small-Satellite Program for Two-Way Environmental Data Collection and Distribution”. In: *AIAA/USU Conference on Small Satellites*. 1992.

[GM16] Claire Goursaud and Yuqi Mo. “Random unslotted time-frequency ALOHA: Theory and application to IoT UNB networks”. In: *2016 23rd International Conference on Telecommunications (ICT)*. IEEE, 2016, pp. 1–5.

[Gol97] Wolfhard J. Goldhirsh Julius; Vogel. *An Overview of the Revised Mobile Satellite Handbook. ’Propagation Effects for Land Mobile Satellite Systems: Overview of Experimental and Modeling Results’*. Tech. rep. NASA, 1997.

[Had98] M.H. Hadjitheodosiou. “Store-and-forward data communications using small terminals and microsatellites”. In: *Proceedings Third IEEE Symposium on Computers and Communications. ISCC’98. (Cat. No.98EX166)*. IEEE Comput. Soc, 1998, pp. 79–83.

[Har33] G. H. Hardy. “A Theorem Concerning Fourier Transforms”. In: *Journal of the London Mathematical Society* s1-8.3 (1933), pp. 227–231.

[HMR95] J.P. Havlicek, J.C. McKeeman, and P.W. Remaklus. “Networks of low-Earth orbit store-and-forward satellites”. In: *IEEE Transactions on Aerospace and Electronic Systems* 31.2 (Apr. 1995), pp. 543–554.

[Itu] *Ionospheric propagation data and prediction methods required for the design of satellite services and systems*. Recommendation ITU-R P.531-12. International Telecommunication Union, 2013.

[JB93] A. Jahn and A. Bottcher. “Improvement of slotted Aloha for land-mobile satellite communications, using channel state information”. English. In: *IEEE 43rd Vehicular Technology Conference*. IEEE, 1993, pp. 339–342.

[Jon13] Stephanie Jones. “Small Satellite Capability Analysis: Communications Subsystem”. In: *AIAA/USU Conference on Small Satellites*. 2013.

[Kis11] Christian Kissling. "Performance enhancements for asynchronous random access protocols over satellite". In: *IEEE International Conference on Communications*. IEEE, 2011, pp. 1–6.

[KK05] V. Kawadia and P.R. Kumar. "A cautionary perspective on cross-layer design". In: *IEEE Wireless Communications* 12.1 (2005), pp. 3–11.

[KL13] Bryan Klofas and Kyle Leveque. "A Survey of CubeSat Communication Systems: 2009-2012". In: *10th Annual CubeSat Developers' Workshop*. 2013.

[Kne+12] Edward Kneller et al. "Cadet: A High Data Rate Software Defined Radio for SmallSat Applications". In: *AIAA/USU Conference on Small Satellites*. 2012.

[KR12] James F. Kurose and Keith W. Ross. *Computer Networking: A Top-Down Approach (6th Edition)*. 6th. Pearson, 2012.

[LHV92] J.-P.M.G. Linnartz, R. Hekmat, and R.-J. Venema. "Near-far effects in land mobile random access networks with narrow-band Rayleigh fading channels". In: *IEEE Transactions on Vehicular Technology* 41.1 (1992), pp. 77–90.

[Li+16] Zhuocheng Li et al. "2D Time-frequency interference modelling using stochastic geometry for performance evaluation in Low-Power Wide-Area Networks". In: (2016). arXiv: `1606.04791`.

[Lic] *Licklider Transmission Protocol - Specification*. Request for Comments 5326. IETF, 2008.

[Liv11] Gianluigi Liva. "Graph-Based Analysis and Optimization of Contention Resolution Diversity Slotted ALOHA". In: *IEEE Transactions on Communications* 59.2 (2011), pp. 477–487.

[LL92] C.T. Lau and C. Leung. "Capture models for mobile packet radio networks". In: *IEEE Transactions on Communications* 40.5 (1992), pp. 917–925.

[Loo85] C. Loo. "A statistical model for a land mobile satellite link". In: *IEEE Transactions on Vehicular Technology* 34.3 (Aug. 1985), pp. 122–127.

[LPST07] Kyle Leveque, Jordi Puig-Suari, and Clark Turner. "Global Educational Network for Satellite Operations (GENSO)". In: *AIAA/USU Conference on Small Satellites*. 2007.

[LS90] A.W. Lam and D.V. Sarwate. "Time-hopping and frequency-hopping multiple-access packet communications". In: *IEEE Transactions on Communications* 38.6 (1990), pp. 875–888.

[Lut+91] E. Lutz et al. "The land mobile satellite communication channel-recording, statistics, and channel model". English. In: *IEEE Transactions on Vehicular Technology* 40.2 (May 1991), pp. 375–386.

[Mas88] James L. Massey. “Some New Approaches to Random-Access Communication”. In: *Proceedings of the 12th IFIP WG 7.3 International Symposium on Computer Performance Modelling, Measurement and Evaluation*. Performance ’87. Amsterdam, The Netherlands, The Netherlands: North-Holland Publishing Co., 1988, pp. 551–569.

[MB09] G. Maral and M. Bousquet. *Satellite communications systems systems, techniques and technology*. John Wiley, 2009.

[Met76] J. Metzner. “On Improving Utilization in ALOHA Networks”. In: *IEEE Transactions on Communications* 24.4 (1976), pp. 447–448.

[Mia+09] Guowang Miao et al. “Cross-layer optimization for energy-efficient wireless communications: a survey”. In: *Wireless Communications and Mobile Computing* 9.4 (2009), pp. 529–542.

[MV10] Martijn de Milliano and Chris Verhoeven. “Towards the next generation of nanosatellite communication systems”. In: *Acta Astronautica* 66.9 (2010), pp. 1425–1433.

[Nam84] C. Namislo. “Analysis of Mobile Radio Slotted ALOHA Networks”. In: *IEEE Journal on Selected Areas in Communications* 2.4 (July 1984), pp. 583–588.

[Nan] *Nanosatellite and CubeSat Database*. Available: `http://www.nanosats.eu/`. Accessed: 2017-05-05.

[NV04] V. Naware and P. Venkitasubramaniam. “A cross-layer perspective in an uncharted path - Signal processing in random access”. In: *IEEE Signal Processing Magazine* 21.5 (Sept. 2004), pp. 29–39.

[Oli+12] Steven J. Olivieri et al. “Modular FPGA-based software defined radio for CubeSats”. In: *2012 IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 3229–3233.

[Osc] *OSCAR-11 30th Birthday Report*. Available: `https://amsat-uk.org/2014/03/10/oscar-11-30th-birthday-report/`. Accessed: 2017-04-18.

[Pal+14] Scott Palo et al. “Expanding CubeSat Capabilities with a Low Cost Transceiver”. In: *AIAA/USU Conference on Small Satellites*. 2014.

[PLC11] Enrico Paolini, Gianluigi Liva, and Marco Chiani. “High Throughput Random Access via Codes on Graphs: Coded Slotted ALOHA”. In: *2011 IEEE International Conference on Communications (ICC)*. IEEE, 2011, pp. 1–6.

[Pro07] John Proakis. *Digital Communications 5th Edition*. McGraw Hill, 2007.

[PSTA01] J. Puig-Suari, C. Turner, and W. Ahlgren. “Development of the standard CubeSat deployer and a CubeSat class PicoSatellite”. In: *2001 IEEE Aerospace Conference Proceedings (Cat. No.01TH8542)*. Vol. 1. IEEE, 2001, pp. 1/347–1/353.

[Ray84] D. Raychaudhuri. “ALOHA with Multipacket Messages and ARQ-Type Retransmission Protocols–Throughput Analysis”. In: *IEEE Transactions on Communications* 32.2 (Feb. 1984), pp. 148–154.

[RBB16] Christian Rodriguez, Henric Boiardt, and Sasan Bolooki. “CubeSat to commercial intersatellite communications: Past, present and future”. In: *2016 IEEE Aerospace Conference*. IEEE, 2016, pp. 1–15.

[RE88] R.R. Rao and A. Ephremides. “On the stability of interacting queues in a multiple-access system”. In: *IEEE Transactions on Information Theory* 34.5 (1988), pp. 918–930.

[Ren+00] W. Ren et al. “Modelling of capture effect and analysis in LEO satellite channel”. In: *2000 IEEE International Conference on Communications. ICC 2000. Global Convergence Through Communications. Conference Record*. Vol. 1. IEEE, 2000, pp. 124–128.

[Ren01] Weili Ren. “A control-centralised multiaccess protocol exploiting non-selective fading for LEO satellite communications.” PhD thesis. University of Surrey, 2001.

[Rob75] Lawrence G. Roberts. “ALOHA packet system with and without slots and capture”. In: *ACM SIGCOMM Computer Communication Review* 5.2 (Apr. 1975), pp. 28–42.

[RWS99] W. Ren, J.W. Ward, and M.N. Sweeting. “Evaluation of capture effect in LEO satellite channel”. In: *Electronics Letters* 35.25 (1999), p. 2171.

[Sad05] B.M. Sadler. “Fundamentals of energy-constrained sensor network systems”. In: *IEEE Aerospace and Electronic Systems Magazine* 20.8 (2005), pp. 17–35.

[SAM13] Scott Schaire, Serhat Altunc, and Benjamin Malphrus. “CubeSat Communication Direction and Capabilities at Morehead State University and NASA Goddard Space Flight Center, Wallops Flight Facility”. In: *AIAA/USU Conference on Small Satellites*. 2013, pp. 1–13.

[Sar] *Saratoga: A Scalable Data Transfer Protocol*. Internet-Draft, work in progress. IETF, 2017.

[Sar15] M. Sarthou. “Argos-3 to Argos-4”. In: *NOAA Satellite Conference*. 2015.

[Sat] *The SatNet Project*. Available: `https://satnet.aero.calpoly.edu/`. Accessed: 2017-05-05.

[Sca+13] Sandro Scalise et al. “S-MIM: a novel radio interface for efficient messaging services over satellite”. In: *IEEE Communications Magazine* 51.3 (Mar. 2013), pp. 119–125.

[Sfo93] Sergio Sforza Mario; Buonomo. *Characterisation of the LMS propagation channel at L- and S-bands: Narrowband experimental data and channel modelling*. Tech. rep. ESA, 1993.

[Sie13] Thomas Sielicki. "Overcoming CubeSat downlink limits with VITAMIN: a new variable coded modulation protocol". MA thesis. University of Alaska Fairbanks, 2013.

[SM05] V. Srivastava and M. Motani. "Cross-layer design: a survey and the road ahead". In: *IEEE Communications Magazine* 43.12 (2005), pp. 112–119.

[Spa+12] S. C. Spangelo et al. "Models and Tools to Evaluate Space Communication Network Capacity". In: *IEEE Transactions on Aerospace and Electronic Systems* 48.3 (2012), pp. 2387–2404.

[Swe87] M.N. Sweeting. "The University of Surrey UoSAT-2 spacecraft mission". In: *Journal of the Institution of Electronic and Radio Engineers* 57.5S (1987), S99.

[SYW90] A.U.H. Sheikh, Y.-D. Yao, and X. Wu. "The ALOHA systems in shadowed mobile radio channels with slow or fast fading". In: *IEEE Transactions on Vehicular Technology* 39.4 (1990), pp. 289–298.

[TB00] M.K. Tsatsanis and S. Banerjee. "Network-assisted diversity for random access wireless networks". In: *IEEE Transactions on Signal Processing* 48.3 (Mar. 2000), pp. 702–711.

[TM79] B.S. Tsybakov and V.A. Mikhajlov. "Ergodicity of a slotted ALOHA system." In: *Probl. Inf. Transm.* 15 (1979), pp. 301–312.

[TP+14] R. Tubío-Pardavila et al. "The HUMSAT System: a CubeSat-based Constellation for In-situ and Inexpensive Environmental Measurements". In: *American Geophysical Union, Fall Meeting 2014*. 2014.

[TV99] V. Tralli and R. Verdone. "Performance characterization of digital transmission systems with cochannel interference". In: *IEEE Transactions on Vehicular Technology* 48.3 (1999), pp. 733–745.

[Var+15] Dimitris Vardalis et al. "Decentralized space-data dissemination for low-cost, dense satellite networks". In: *IEEE Transactions on Aerospace and Electronic Systems* 51.4 (2015), pp. 3071–3084.

[VG88] W.J. Vogel and J. Goldhirsh. "Fade measurements at L-band and UHF in mountainous terrain for land mobile satellite systems". In: *IEEE Transactions on Antennas and Propagation* 36.1 (1988), pp. 104–113.

[Wal91] Philip Walker. "A Global Enviromental Data Distribution System". In: *AIAA/USU Conference on Small Satellites*. 1991.

[War93] J. Ward. “Design, Implementation and In-Orbit Demonstration of a Store-and-Forward Digital Communication System for Low Earth Orbit Microsatellites”. PhD thesis. University of Surrey, Oct. 1993.

[WE02] Wei Luo and A. Ephremides. “Power levels and packet lengths in random multiple access”. In: *IEEE Transactions on Information Theory* 48.1 (2002), pp. 46–58.

[Wer01] James Richard. Wertz. *Mission geometry : orbit and constellation design and management : spacecraft orbit and attitude systems*. Microcosm Press, 2001, p. 934.

[WP91] J.W. Ward and H.E. Price. “Protocols for Store-and-Forward Message Switching via Microsatellites”. In: *AIAA/USU Conference on Small Satellites*. 1991.

[WS00] J. Ward and M. Sweeting. “Capture effect and its enhancement in LEO satellite channel”. In: *IEEE/AFCEA EUROCOMM 2000. Information Systems for Enhanced Public Safety and Security (Cat. No.00EX405)*. IEEE, 2000, pp. 184–188.

[Yoo+10] Do-Sik Yoo et al. “Coding and Modulation for Short Packet Transmission”. In: *IEEE Transactions on Vehicular Technology* 59.4 (2010), pp. 2104–2109.

[Zan+13] Andrea Zanella et al. “M2M massive wireless access: Challenges, research issues, and ways forward”. In: *2013 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2013, pp. 151–156.

[ZC05] D.B. Zeleke and M.A.V. Castro. “ALOHA versus Single Code Spread ALOHA for Satellite Systems”. In: *2005 IEEE 61st Vehicular Technology Conference*. Vol. 4. IEEE, 2005, pp. 2692–2696.

[ZP92] K. Zhang and K. Pahlavan. “Relation between transmission and throughput of slotted ALOHA local packet radio networks”. In: *IEEE Transactions on Communications* 40.3 (Mar. 1992), pp. 577–583.

[De +14] Riccardo De Gaudenzi et al. “Asynchronous Contention Resolution Diversity ALOHA: Making CRDSA Truly Asynchronous”. In: *IEEE Transactions on Wireless Communications* 13.11 (2014), pp. 6193–6206.

[De +16] Mauro De Sanctis et al. “Satellite Communications Supporting Internet of Remote Things”. In: *IEEE Internet of Things Journal* 3.1 (2016), pp. 113–123.