Università
degli Studi
di Padova

Sede Amministrativa: Universit degli Studi di Padova
Dipartimento di Ingegneria dell'Informazione

DOTTORATO DI RICERCA IN : Ingegneria dell'Informazione
CURRICOLO: Scienza e Tecnologia dell'Informazione
CICLO: XXIX

Energy Efficient Communication in Wireless Networks

Coordinatore: Ch.mo Prof. Matteo Bertocco
Supervisore: Ch.mo Prof. Michele Zorzi

Dottorando: Waqas Bin Abbas

Academic Year 2016-2017

Abstract

This thesis investigates energy efficient communication in wireless networks. The work considers energy efficiency from both a device and a network perspective.

Firstly, we investigate the energy efficiency (EE) of a receiver equipped with large antenna arrays. In a receiver architecture, operating either at millimeter wave (mmWave) or at microwave (massive multiple input multiple output (MIMO)), the use of a high resolution analog-to-digital converter (ADC) is traditionally considered as a main bottleneck to achieve an energy efficient design. To address this issue, in contrast to previous works, we compared the performance of analog, hybrid and digital beamforming (ABF, HBF and DBF) based receiver architectures with low resolution ADCs, and show that when power consumption of all receiver components is taken into account DBF results in a better spectral, power and energy efficiency than ABF and HBF in most scenarios for both control plane and data plane communication. Results also show that based on the ADC power consumption model there is an optimal number of ADC bits that results in a maximum EE (defined as capacity/total power consumption). To further reduce the power consumption of DBF, a novel variable resolution ADC architecture is proposed, which motivates the use of variable resolution ADCs instead of conventional fixed resolution ADCs. To address the issue of energy consumption related to the directional cell search associated with a large antenna array system, we propose to exploit the availability of context information to reduce the search delay associated with ABF and propose a low power phase shifters network based receiver architecture which compares and selects the best beam in analog hardware and requires only a single RF chain. We also propose a higher sub-carrier bandwidth based control plane structure to reduce the energy consumption of the directional cell discovery process.

Secondly, we study energy efficient communication in a multi-hop flooding network. Particularly, we study and analyze how in a multi-hop network a hybrid automated repeat request (HARQ) scheme based on fountain codes can reduce the probability of failure at the destination. The analysis is performed for both unrestricted and restricted flooding scenarios, and results verify that below a certain minimum link failure probability restricted flooding performs similar to unrestricted flooding but with fewer transmissions compared to unrestricted flooding. This reduction in number of transmissions is directly related to an improvement in energy efficiency of an overall network. Moreover, we also proposed two practical restricted flooding policies; 1) Predetermined restriction, which performs similar to restricted flooding, and 2) Adaptive restriction, which adapts its number of transmissions according to the link failure probability and performs similar to unrestricted flooding. Furthermore, we also study the performance of a network in the presence of helping nodes which can decode and generate extra redundancy packets in the network. Results show that the inclusion of helping nodes can further reduce the number of transmissions compared to a case where only the source is transmitting.

Acknowledgments

I would like to thank my family for their moral support. I would also like to thank Professor Michele Zorzi and all others who supported me in the fulfillment of this work.

Contents

List of Acronyms

5G	Fifth Generation
ABF	Analog Beamforming
AC	Analog Combining
ACK	Acknowledgment
ADC	Analog to Digital Converter
AoA	Angle of Arrival
AoD	Angle of Departure
AR	Adaptive Restriction
ARQ	Automatic Repeat Request
AQNM	Additive Quantization Noise Model
B	Bandwidth
BS	Base Station
CI	Context Information
CP	Control Plane
CSIR	Channel State Information at the Receiver
CSIT	Channel State Information at the transmitter

DBF	Digital Beamforming
DC	Digital Combining
DP	Data Plane
EE	Energy Efficiency
GASBA	Greedy Antenna Selection and Bit Allocation
GBA	Greedy Bit Allocation
HARQ	Hybrid Automatic Repeat Request
HBF	Hybrid Beamforming
HC	Hybrid Combining
ICD	Initial Cell Discovery
i.i.d	Independent and Identically Distributed
LTE	Long Term Evolution
MIMO	Multiple Input Multiple Output
mmWave	Millimeter Wave
MS	Mobile Station
PR	Predetermined Restriction
PS	Phase Shifters
PSN	Phase Shifters Network
RF	Radio Frequency
SE	Spectral Efficiency
SNR	Signal-to-Noise Ratio
ULA	Uniform Linear Array

List of Figures

List of Tables

1

Introduction

The work presented in this thesis is focused on energy efficient communication in wireless networks. Two important aspects of wireless communication have been studied. The first part focuses on fifth generation (5G) cellular communication, where the efficiency of a millimeter wave (mmWave) receiver (equipped with large antenna arrays) using low resolution analog-to-digital converters (ADC) has been studied and analyzed. Receiver energy efficiency has been examined considering both data plane (DP) and control plane (CP) communication. The second part focuses on energy efficient communication in a multi-hop network, where focusing on the controlled flooding of fountain codes, a trade-off between the number of transmissions and the amount of redundancy injected into the network has been studied and analyzed. The energy efficiency is directly related to the total number of transmissions in a multi-hop network, and therefore, the goal is to achieve a desired success probability at the destination with a minimum number of transmissions, which improves the energy efficiency of the network.

Energy efficiency is already a key concern, and will be even more critical for future wireless communication. This is important from both a device and a network perspective [1]. Firstly, future transmitters/receivers (especially operating at mmWave) are expected to be equipped with large antenna arrays that can result in an increase in power consumption compared to current single antenna devices. This makes energy efficient communication a critical issue especially at the receiver where the power consumption of the ADC is considered as a main bottleneck. Secondly, an increase in the number of connected devices and in the size of multi-hop networks suggests to incorporate transmission efficient routing protocols to achieve better energy efficiency at the network level by reducing the number of transmissions. In this thesis, we address both these aspects, and study the regimes which can be beneficial to improve energy efficiency.

In this chapter we introduce those topics and provide the outline of the thesis. Particularly, in Section 1.1 we introduce 5G and highlight the importance of energy efficiency, while in Section 1.2 we introduce the topic of controlled flooding of fountains codes. Finally, in Section 1.3 we provide a brief outline of rest of this thesis.

1.1 Receiver Efficiency in 5G Cellular Communication

Each generation of cellular or mobile communication has emerged with certain key technological advancements [2]. For instance, the first generation (1G) of mobile communication started with analog voice communication, which was improved to digital voice communication and messaging in the second generation (2G). The third generation (3G) extended 2G by adding data communication. Finally, the fourth generation (4G) incorporated orthogonal frequency division multiplexing and all IP based communication, which allows mobile users to experience high speed broadband internet with very low latency.

Currently, 4G technology is successfully fulfilling the requirements of the mobile users. However, the future market trends (2020 and beyond), such as ubiquitous connectivity, internet of things (IoT), smart cities, machine-to-machine communication, tactile internet, etc, suggest a need for a new generation of mobile communication [3, 4]. Moreover, Cisco Visual Networking Index (VNI) forecasted an exponential increase in the Compound Annual Growth Rate, reaching 24.3 exabytes per month by 2019, which cannot be accommodated with the current Long Term Evolution (LTE) standard [5].

Therefore, fifth generation of mobile communication will be necessary to address the demands of future communication. The key requirements for 5G to enable the above mentioned (and many others) objectives are summarized below [3, 6–9]:

- **Higher System Capacity:** 5G requires to manage large volumes of traffic, and therefore requires many orders of magnitude higher capacity than the current 4G standard. The target is to achieve a 1000x increase in capacity compared to today's networks.

- **Higher Data Rate:** 5G has to support higher data rates, around 1 Gbps. The requirement is to have an increase of 100x in user throughput compared to LTE.

- **Support of Massive Connectivity:** To support IoT, a 100x increase in simultaneous device connectivity is mandatory.

- **Reduced Latency:** To provide faster access and to support technologies such as tactile internet, a latency of less than a millisecond is a key requirement.

- **Reduced Cost, Higher Energy Efficiency and Robustness:** In the context of green communication and of a massive increase in the number of communicating devices [10], energy consumption should be reduced by a factor 1000x. Moreover, robustness against emergencies is also a necessary requirement.

The first part of the thesis mainly focuses on the last point i.e., hoq to improve energy efficiency while achieving the required data rates and reduced cell search latency. There are two main competing technologies to support the data rate requirements of 5G; 1) communication at millimeter wave (mmWave), and 2) massive multiple input multiple output (MIMO). Both technologies rely on large antenna arrays and therefore require a much higher number of additional analog blocks, which significantly increases the power consumption, especially at the receiver. In this thesis, we study the energy efficiency of a mmWave receiver, and identify the choice of an appropriate beamforming scheme for both control and data plane communication.

Large antenna arrays are inevitable for future wireless communication networks. For instance, massive MIMO and mmWave which are considered as prime candidates for 5G wireless communication will be equipped with large antenna arrays [11–15]. Firstly, in massive MIMO, large antenna arrays are required to exploit the spatial degrees of freedom which eventually result in an increased capacity. Secondly, at mmWave huge bandwidth is available to fulfill the data rate requirements of 5G cellular networks. However, frequencies at these wavelengths suffers from high path-loss. To reduce this path-loss, and therefore to improve coverage, beamforming gain using large antenna arrays is an essential requirement. [1].

Beamforming is a key technology for both massive MIMO and mmWave communication. Analog, Hybrid and Digital are the available beamforming/combining schemes for MIMO communication systems [16]. Although a fully digital architecture, which requires a separate

---

[1] Although the main focus of this work is on mmWave, the results related to the architecture of variable resolution ADCs are in general valid for receivers with large antenna arrays (i.e., for both mmWave and massive MIMO).

RF chain per antenna element, is a popular choice in classical systems, it is not generally considered as a viable option for large antenna arrays due to the high power consumption of analog-digital signal processing components [17]. This is even more critical at mmWave frequencies where the power consumption of an Analog to Digital Converter (ADC) in a receiver grows linearly with the system bandwidth. However, multiple options exist to reduce power consumption, such as the use of 1) analog combining (AC) [18], 2) hybrid combining (HC) [19], and 3) low-resolution ADCs [20, 21].

This thesis studies the feasibility of all these options. Particularly, we identify, 1) the scenarios where AC can be advantageous, 2) whether HC always consumes lower power, especially when the power consumption of all the receiver components is taken into account, and 3) what is the minimum number of ADC bits that would limit the power of a DC architecture but still manage to provide an acceptable system performance.

Specifically, in contrast to previous works, a detailed analysis and comparison of power, energy and spectral efficiency of the mentioned combining schemes with low resolution ADCs will be presented. We also show that there is an optimal number of bits (4-5) that maximizes the energy efficiency. Then, a variable resolution ADC architecture will be proposed which provides a power efficient solution compared to fixed resolution ADC based receivers. We then study the directional initial cell discovery problem and show how the availability of context information can be useful and propose a low power phase shifters network based architecture to mitigate the errors in the available context information. We also compare different beamforming schemes in terms of power, latency and energy consumption. The outcome of this part of the thesis is the following:

"A fully digital architecture with low resolution ADCs and while taking into account power consumption of all receiver components may actually result in a better spectral, energy and power efficiency than other available beamforming architectures in many use-cases."

1.2 Energy Efficient Communication in Multi-hop Networks

The second part of this thesis is related to the study of energy efficient communication in a multi-hop flooding wireless network. Since the last decade, energy efficient communication in a multi-hop flooding network has been a key topic of interest [22–24]. In addition to energy efficiency, reliable communication is also another important aspect. Flooding is one of the options to improve the reliability of communication in a multi-hop network. However, this increases the number of nodes and hence the number of transmissions required to forward a packet towards its destination. Therefore, uncontrolled flooding may results in a higher number of transmissions which is directly related to an increase in the energy consumption (and therefore reduced energy efficiency).

The focus of this part of the thesis is on delay tolerant networks and on erasure channels (where the packet loss probability may be unknown a-priori). Typical examples of such scenarios are file transfer and over the air software updates. In such scenarios the main priority is to transfer data successfully, while delay may be less important. To achieve high reliability, most recent wireless standards use a combination of automatic repeat request (ARQ) and forward error correction (FEC), i.e., hybrid ARQ (HARQ). In ARQ, the receiver detects whether or not there is an error in the received message. In case of an error, a negative acknowledgement (NAK) is sent back to the transmitter, and the packet is retransmitted, while in case of successful reception, the receiver sends back an acknowledgement (ACK) to the transmitter. To further improve the reliability, ARQ is combined with FEC and termed as hybrid automatic repeat request. The use of rate-less codes as FEC in case of an erasure channel is a popular choice. Rate-less codes allow a near-instantaneous adaptation to the channel quality and are well suited for communication in wireless networks. An important class of rate-less codes is fountain codes, which are well known to improve the reliability of successful communication [25–27]. In this work we evaluate a network performance where the source transmits packets using fountain codes.

We will consider a multi-hop network, where the source transmits fountain coded packets to the destination separated by r relay stages, and each relay stage consists of N_i nodes. In this flooding based scenario, we will investigate and analyze the performance in terms of probability of successful reception at the destination. Although the combination of flooding and fountain codes in a multi-hop network achieves very high reliability, it may result in unnecessary transmissions and therefore in an overall reduction of the energy efficiency of the network. To address this issue, we will study a trade-off between the amount of redundancy added by the fountain codes and the required number of transmissions while achieving the required probability of success at the destination.

Specifically, in contrast to previous works, we will devise a probabilistic model to evaluate the probability of successful decoding (of a fountain coded packet) at the destination in a multi-hop flooding network, as a function of the number of forwarding nodes, the link error probability and the amount of redundancy injected into the network. The model will consider both restricted and unrestricted flooding scenarios, and will allow to evaluate the number of transmissions required to achieve a certain probability of success at the destination as a function of the link error probability. Two practical routing protocols will be proposed to model the restricted flooding case. The model will then be extended to study the impact of helper nodes (which can add extra redundancy in the network) on the reduction of required transmissions. We will also show that our proposed model matches the simulations very well. The outcome of this part is the following:

"In comparison to unrestricted flooding, restricted flooding of fountain coded packets can achieve better energy efficiency without any performance degradation."

1.3 Outline

As mentioned earlier, this work studies energy efficiency from a device perspective and a network perspective. Particularly, in Chapters 2-4, we will discuss mmWave receiver energy efficiency for 5G, while in Chapter 5 we will study energy efficient transmission in a multi-hop wireless network. Moreover, due to wide span of topics addressed in this work, each chapter will include its own separate introduction/related work sections. The rest of the thesis is organized as follows:

- **Chapter 2** discusses the power and energy efficiency of a mmWave receiver architecture. A comprehensive comparison of different beamforming schemes is presented while considering low resolution ADCs.

- **Chapter 3** highlights how the availability of variable resolution ADCs can further reduce the power consumption of a fully digital architecture.

- **Chapter 4** discusses directional cell discovery and provides a comprehensive comparison based on access error probability, time vs power consumption trade-off and energy consumption of different beamforming schemes.

- **Chapter 5** focuses on controlled flooding of fountain codes. Considering a multi-hop scenario a trade-off between the amount of redundancy injected in the network and the required number of transmissions has been analyzed.

- **Chapter 6** provides the conclusions.

2

Millimeter Wave Receiver Efficiency

This chapter introduces the popular millimeter wave receiver architectures for Analog, Digital and Hybrid combining (AC, DC and HC) schemes. Then, primarily focusing on low resolution Analog-to-Digital Converters (ADCs), we analyze and compare the power consumption (considering all receiver components) of Analog, Digital and Hybrid combining architectures. We show how the power consumption of these combining schemes varies with a change in the number of antennas, the number of ADC bits (b) and the bandwidth (B). Moreover, we compare different ADC power consumption models, and show that for a certain range of number of antennas, b and B, DC may actually have a comparable and lower power consumption than AC and HC, respectively. In addition, we also show how the choice of an appropriate beamforming scheme depends on the signal-to-noise ratio regime.

We then extend our analysis one step further and study the capacity and the energy efficiency of Analog, Hybrid and Digital Combining (AC, HC and DC) for millimeter wave (mmWave) receivers. We take into account the power consumption of all receiver components, not just ADCs, determine some practical limitations of beamforming in each architecture, and develop a performance chart that enables a comparison of different architectures at a glance. We consider an Additive Quantization Noise Model (AQNM) to evaluate the achievable rates with low resolution ADCs. We show that there is an optimal number of ADC bits that maximizes the energy efficiency. Our analysis also shows that AC is only advantageous if the channel rank is strictly one or the power constraint is very stringent. Otherwise, we show that the usual claim that DC requires the highest power stems from a quite conservative focus only on ADCs. Rather, considering the power consumed by all mmWave receiver components, DC may result in a better efficiency than HC in many use-cases. Therefore, DC due to its versatility and the uncertainty on future mmWave conditions, seems to be the most future-proof architecture.

2.1 Introduction

The millimeter wave spectrum (30-300 GHz), where a very large bandwidth is available, is considered as a prime candidate to fulfill the data rate requirements of future broadband communication [11–14, 28]. However, communication at these frequency bands exhibits high pathloss [11]. To overcome this high pathloss, spatial beamforming/combining using large antenna arrays is considered as an essential part of a mmWave communication system.

Analog, Hybrid and Digital are the available beamforming/combining schemes for MIMO communication systems [16]. Although a fully digital architecture, which requires a separate RF chain per antenna element, is a popular choice in classical systems, it is not generally considered as a viable option for large antenna arrays due to the high power consumption of analog-digital signal processing components [17]. This is even more critical at mmWave frequencies where the power consumption of an ADC in a receiver grows linearly with the system bandwidth. However, multiple options exist to reduce the power consumption

1. The use of AC, which requires a single Radio-Frequency (RF) plus ADC chain, consumes least power and is an attractive choice whenever the advantages of digital processing techniques (mainly spatial multiplexing) are not required [18].

2. The use of HC, which performs combining in both the analog and the digital domain with a reduced number of RF chains, at the cost of lower flexibility than fully digital architectures [19].

3. The use of low-resolution ADCs (for example, 1 bit) while maintaining the Digital Combining (DC) MIMO architecture [20, 21].

In this chapter, we show in Section 2.2 that the general perception regarding the high power consumption of DC is not always justified when the power consumption of all receiver components is considered. Rather, there is a certain range of system bandwidth (suitable to fulfill the requirements of certain key functionalities of future wireless communication) and ADC resolution (to avoid any significant performance loss) for which DC results in a power consumption lower than HC and comparable to AC while providing the flexibility of digital processing, which makes DC an attractive candidate for mmWave receiver design.

We then extend our analysis and in Section 2.3 we provide a comprehensive characterization of the Spectral Efficiency (SE, defined as capacity/bandwidth) versus the Energy Efficiency (EE, defined as capacity/total power consumption) of AC, HC and DC architectures for a wide variety of possible scenarios. Our results show that in a wide range of use-cases DC may be a more attractive choice than HC in terms of both SE and EE. Primarily, we argue that the popular notion that DC has the highest power consumption, based solely on ADC power, is quite narrow. Rather, DC with low resolution ADCs may result in a lower power consumption than HC when the power consumption of all receiver components is taken into consideration.

2.1.1 Related Work

Recent works study energy efficient designs, particularly focusing on how the system capacity varies as a function of the ADC resolution. In [29], the capacity of a quantized MIMO system along with coding has been studied, while in [30] the capacity and bit error rate for a quantized MIMO system was analyzed. In [31] a capacity lower bound for a quantized MIMO system with Gaussian input were analyzed. However, the quantized capacity model is only optimal at low SNR regimes. In [32], an exact nonlinear quantizer model is utilized to evaluate the optimal capacity for a 1-bit ADC. In [33], considering a MIMO channel and an additive quantization noise model (AQNM, an approximate model for ADCs), a joint optimization of ADC resolution and number of antennas is studied. In a recent work [34], the authors studied how the number of ADC bits b and the bandwidth (sampling rate) B of ADCs affect the total power consumption for AC and DC based receivers with a stringent power constraint such as a mobile station. They studied the optimal b and B which maximize the capacity for AC and DC only for low power receiver design while also showing that DC with similar power budget to AC may achieve a higher rate than AC when the channel state information is available at the transmitter. In another recent work [35], the EE and SE for low resolution ADC HC architecture is studied. The authors only show the advantages of a low resolution HC with few RF chains over an infinite resolution ADC for DC and HC.

Recently, energy efficient architectures for HC are proposed [36, 37]. In [36], an energy efficient HC architecture has been proposed where each RF chain is only connected to a subset of antennas. In [37], to further reduce the power consumption of a conventional HC with phase shifters, switch based architectures are proposed, where at a particular instant only a reduced set of antennas (equal to the number of RF chains) is selected and connected to the RF chains. However, in both works the proposed architectures result in a lower SE than the fully connected phase shifter architecture.

In [38,39], the SE of uplink massive MIMO with low resolution ADCs is studied, and it is shown that few ADC bits are enough to achieve almost the same SE of unquantized MIMO, and in [38] it is also shown that a 2-bit ADC achieves good performance for a small Rician k-factor.

To the best of our knowledge, there is no previous work that compares the *total* power consumption of AC, HC and DC based receiver design for different values of the number of receive antennas N_r, the number of ADC bits b, and the bandwidth B. In Section 2.2, we provide a comprehensive comparison of total power consumption of different beamforming schemes while considering ADC power models. We also discuss the relationship of the signal-to-noise ratio (SNR) with the ADC resolution and how it affects the choice of an appropriate beamforming scheme.

Moreover, there are no previous works that analyze and compare AC, DC and HC for a finite resolution ADC (i.e., quantized MIMO system), and study their trade-off in terms of both EE and SE. In Section 2.3, we would like to highlight that our model identifies and overcomes some limitations of previous frameworks on this topic.

- First, some works on HC benchmark their proposal against ideal unquantized DC [19, 37,48]; thus, rather than evaluating which architecture requires less power, these works focus on how close HC performance is to ideal DC. This framework inherently assumes that HC is preferable over low-resolution DC, whereas we show that this may not be true in many cases.

- Second, some works do not account for some components in the HC architecture, such as the signal splitters (SP) or phase shifters (PS) [17,19]. These components, too, have a non-negligible power signature. Our analysis takes into account the power consumption of *all* the components of the receivers as considered in the recent literature [34,37].

- Third, we take into account that system designers may express different preference for SE and EE. We express such preference as a multi-objective utility maximization problem where the relative weight of SE vs EE is a free parameter. This gives new insights on the choice between HC and DC that complements their separate analysis in [35].

- Fourth, we consider point-to-point Downlink and Uplink scenarios to study the effect of the number of receive antennas. The results show that for the Downlink (i.e., the case of a mobile receiver with fewer antennas), DC outperforms HC for all levels of preference between SE and EE. On the other hand, in the Uplink case (i.e., a base station receiver with many antennas) DC is preferable if SE and EE have similar importance, but HC is preferable if the operator only desires to maximize EE while accepting a degradation in SE.

- Fifth, we observe that the power consumption figure of merit for ADCs varies 3 orders of magnitude between different references in the literature. Likewise, the power consumption of phase shifters varies 2 orders of magnitude in references. We discuss the origin of this variation, and evaluate the three receiver architectures with two types of component models: Firstly, we consider existing state-of-the-art devices, and secondly, we study how the results would change with a plausible future projection of technology improvements, based on existing hardware surveys.

- Sixth, we emphasize that the trade-off between EE and SE is critically dependent on the component characteristics. It is incomplete to study power consumption with only one typical set of values, even if these are very well documented in industry feedback as in [35]. We encourage independent researchers to reproduce our analysis for their own set of component parameters using a web tool we have made available [49]. In the tool, our proposed analysis can be extended by the user through the specification of any desired power consumption value for each component.

2.2 Receiver Power Consumption Comparison

In mmWave receiver design, the ADC is usually considered to be the most power hungry block. In general, DC, which requires a number of ADCs (N_{ADC}) equal to twice the number of receive antennas (separate ADCs for each inphase and quadrature phase signal) is typically

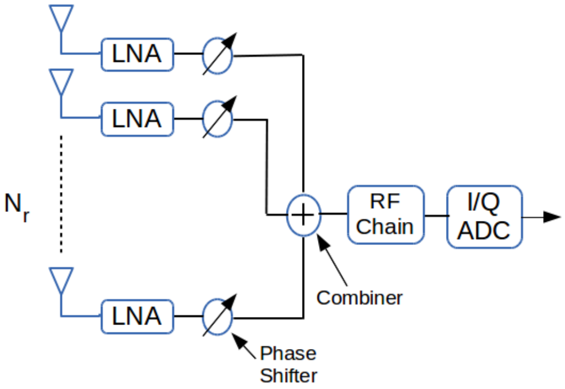

Fig. 2.1. *Analog Combiner.*

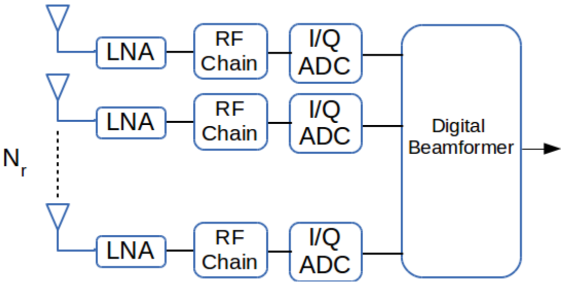

Fig. 2.2. *Digital Combiner.*

assumed to result in maximum power consumption, while AC with $N_{ADC} = 2$ is the least power consuming scheme. On the other hand, Hc requires $N_{ADC} = 2 \times N_{RF}$ where $N_{RF} \leqslant N_r$ is the number of RF chains, and generally is assumed to have a lower power consumption than DC.

The commonly accepted conclusion that DC suffers from high power consumption is the result of implicitly assuming the use of high resolution and wide band ADCs, that therefore dominate the overall power budget. However, the power consumption of an ADC is directly proportional to the number of quantization levels and to the sampling rate. In addition, for different beamforming schemes, a power consumption comparison only based on ADCs can result in a different outcome with respect to what would be obtained when considering the total power consumption, especially when the resolution and/or the bandwidth of the ADC are not large. In this section, we compare the total power consumption (P_{Tot}) of AC, DC and HC by considering a low power ADC (LPADC) and a high power ADC (HPADC) models, and for different values of N_r, B and b. Our results show that

- P_{Tot} for all beamforming schemes increases with an increase in N_r, b or B;
- for fixed N_r and B, there is a maximum number of bits (b^*) up to which DC is more energy efficient than HC;
- for fixed N_r and b, there is a maximum bandwidth B^* up to which DC is more energy efficient than HC;
- DC always has higher power consumption than AC, however, for small b and B the difference is relatively small and therefore, in those configurations, DC may still be

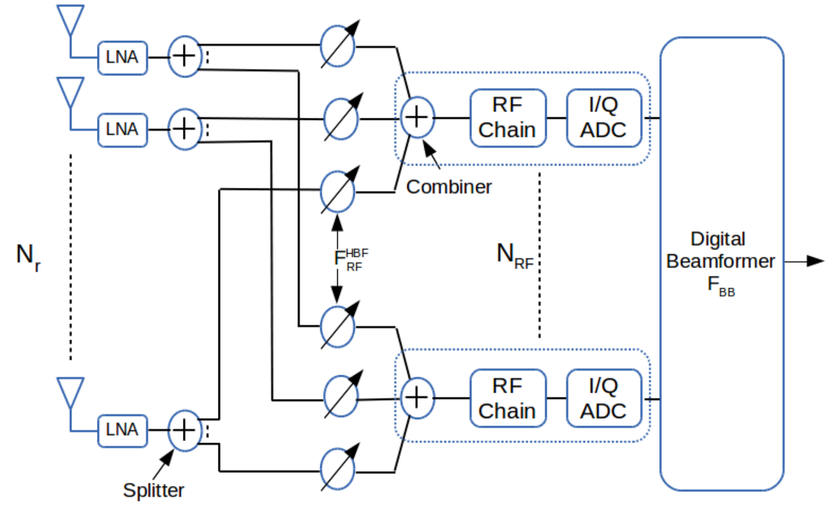

Fig. 2.3. *Hybrid Combiner.*

an attractive option, also in view of the much greater flexibility provided by digital processing;

- in comparison to HC, if the ratio of N_{RF} and N_r remains constant, b^* and B^* for DC both increase with an increase in N_r;
- the variation in b has a more significant effect at high SNR than at low SNR.

Finally, the choice of the appropriate beamforming scheme depends not only on the total power consumption but also on the SNR regime. Although the qualitative trends among different beamforming schemes are rather predictable, the precise quantification of these relationships and the results presented in this section are useful to precisely characterize the regimes where the various beamforming options are to be preferred.

2.2.1 System Model

Common mmWave receiver architectures for AC, DC and HC are shown in Figures 2.1, 2.2 and 2.3, respectively. For all schemes, it is assumed that mixing is done in RF. The total power consumption P_{Tot} of these beamforming schemes can be evaluated as[1]

$$P_{Tot}^{AC} = N_r(P_{LNA} + P_{PS}) + P_{RF} + P_C + 2P_{ADC} \tag{2.1}$$

---

[1] Note that our power consumption model for analog, digital and hybrid combining schemes does not account for the digital interface.

$$\begin{aligned} P_{Tot}^{HC} &= N_r(P_{LNA} + P_{SP} + N_{RF}P_{PS}) \\ &\quad + N_{RF}(P_{RF} + P_C + 2P_{ADC}) \end{aligned} \tag{2.2}$$

$$P_{Tot}^{DC} = N_r(P_{LNA} + P_{RF} + 2P_{ADC}) \tag{2.3}$$

where P_{RF} represents the power consumption of the RF chain and is given by

$$P_{RF} = P_M + P_{LO} + P_{LPF} + P_{BB_{amp}} \tag{2.4}$$

and P_{LNA}, P_{PS}, P_C, P_M, P_{LO}, P_{LPF}, $P_{BB_{amp}}$, P_{ADC}, and P_{SP} represent the power consumption of low noise amplifier (LNA), phase shifter, combiner, mixer, local oscillator[2], low pass filter, baseband amplifier, ADC, and splitter, respectively. In our analysis, the power consumption of all other components except the ADC is considered independent of the system bandwidth, whereas P_{ADC} increases linearly with B and exponentially with b [43]. Therefore, considering Nyquist sampling rate, P_{ADC} in terms of B and b is given by

$$P_{ADC} = cB2^b = cBR \tag{2.5}$$

where c is the energy consumption per conversion step, and $R = 2^b$ is the number of quantization levels of the ADC. The value for c is 12.5 pJ [42] and 474 fJ [44] for HPADC and LPADC, respectively.

2.2.1.1 P_{Tot} Comparison

A comparison of P_{Tot}^{AC}, P_{Tot}^{HC} and P_{Tot}^{DC} is shown in Figures 2.4 and 2.5 for B equal to 100 MHz and 1 GHz, respectively. In these plots, N_r is set to 16 and 64, b is varied from 1 to 10, and $N_{RF} = 4$ for HC. Moreover, $P_{LNA} = 39$ mW, $P_{PS} = 19.5$ mW, $P_M = 16.8$ mW, [41], [45] $c = 494$ fJ [44], $P_{LO} = 5$ mW, $P_{LPF} = 14$ mW, $P_{BB_{amp}} = 5$ mW [42] and $P_{SP} = 19.5$ mW. Note that the results shown in Figures 2.4 and 2.5 are for the LPADC considered in [44][3].

In Figures 2.4 and 2.5, results show that P_{Tot} increases with an increase in N_r, B or b, as expected. Firstly, note that AC consumes the least power for every configuration. Secondly, DC always has some configuration for which it has a lower power consumption than HC. This

---

[2] We considered a separate local oscillitor for each RF chain as in [42].

[3] Similar results can be obtained by considering HPADC (with $c \approx 12.5$ pJ) as in [42], which results in a reduced range of b or B for which DC has a lower power consumption than AC or HC. We will mention the range of b and B for HPADC whenever necessary.

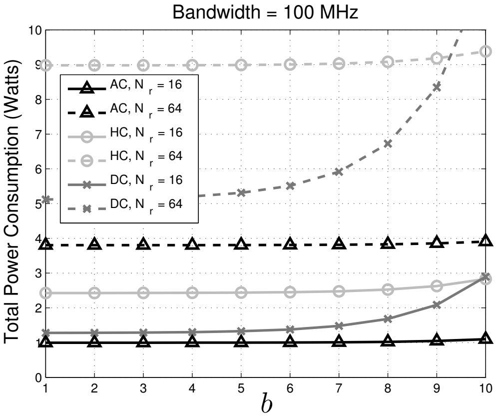

Fig. 2.4. *P_{Tot} for different beamforming schemes vs b for $B = 100$ MHz and $N_r = 16, 64$.*

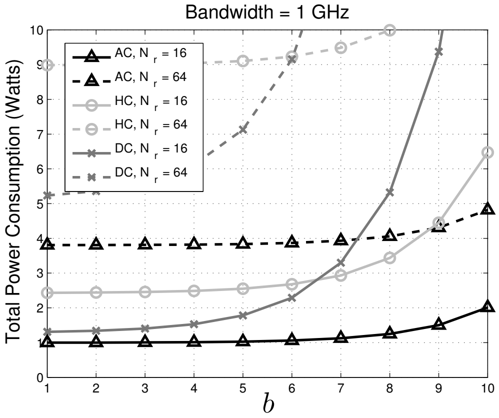

Fig. 2.5. *P_{Tot} for different beamforming schemes vs b for $B = 1$ GHz and $N_r = 16, 64$.*

is because P_{ADC} increases exponentially with b, and therefore for small b there is no significant power consumption due to P_{ADC} with respect to the other components in Eq. (2.3). For instance, an ADC with 6 bits only consumes 31.6 mW whereas a single LNA consumes 39 mW of power. This suggests that in a low bit multi-antenna receiver, the power consumption of an LNA may be of more concern for an energy efficienct receiver design. Moreover, at low b, the power consumption of additional components in HC, e.g., phase shifters, becomes dominant and therefore HC may even result in a higher power consumption than DC. Note that the value of b which results in a lower P_{Tot}^{DC} in comparison to P_{Tot}^{HC} (for fixed N_{RF}) decreases with an increase in N_r and B. For instance, for $N_r = 64$ and with $B = 1$ GHz and $B = 100$ MHz, P_{Tot}^{DC} is less than P_{Tot}^{HC} up to 6 bits and 9 bits, respectively. Moreover, similar results obtained by considering an HPADC model [42] (not shown here), show that P_{Tot}^{DC} always results in a higher power consumption than P_{Tot}^{AC} for the configurations used in Figure 2.4 and 2.5. However, DC results in a lower power consumption than HC for $B = 100$ MHz and $B = 1$ GHz and with $N_r = 16$ only for a range of b up to 5 and 2, respectively. A further discussion on the impact of the number of bits is given in Section-2.2.2.

We next provide analytical formulas to identify B^* and b^* for which P_{Tot}^{DC} is similar to P_{Tot}^{HC}, for a general N_r. This is useful to properly characterize the regions in which DC is to be preferred over the HC alternative.

2.2.1.2 Evaluation of b^* and B^*

We now compare DC with HC, and evaluate the maximum number of bits b^* and the maximum bandwidth B^* which satisfy the condition that $P_{Tot}^{DC} \leqslant P_{Tot}^{HC}$[4].

To find the values of b^* and B^* that result in the same total power consumption for HC and DC we first evaluate the intersection point of Eqs. (2.2) and (2.3). This gives the following result

$$(N_r - N_{RF})P_{RF} + 2(N_r - N_{RF})P_{ADC} = N_r N_{RF} P_{PS} + N_{RF} P_C + N_r P_{SP} \tag{2.6}$$

and therefore b^* and B^* for HC and DC can be calculated as

$$\begin{aligned} R &= \frac{N_r(N_{RF}P_{PS} + P_{SP}) + N_{RF}P_C - (N_r - N_{RF})P_{RF}}{2(N_r - N_{RF})cB} \\ b^* &= \lfloor log_2(R) \rfloor \end{aligned} \tag{2.7}$$

$$B^* = \frac{N_r(N_{RF}P_{PS} + P_{SP}) + N_{RF}P_C - (N_r - N_{RF})P_{RF}}{2(N_r - N_{RF})cR} \tag{2.8}$$

where $\lfloor x \rfloor$ represents the floor of the variable x, i.e., the largest integer $\leqslant x$. Eqs. (2.7) and (2.8) hold for $N_{RF} < N_r$. Now if $N_r \to \infty$, b^* and B^* are given by

$$b^* = \left\lfloor log_2\left(\frac{N_{RF}P_{PS} + P_{SP} - P_{RF}}{2cB}\right) \right\rfloor \tag{2.9}$$

$$B^* = \frac{N_{RF}P_{PS} + P_{SP} - P_{RF}}{2cR} \tag{2.10}$$

Eqs. (2.9) and (2.10) show that, for a large number of antennas, the values of b^* and B^* for DC are inversely related to B and b, respectively, and directly related to N_{RF}. Moreover, for constant P_{PS}, P_{RF}, P_{SP}, c, N_{RF} and b or B, Eqs. (2.9) and (2.10) also provide a lower bound for b^* and B^*, respectively, for any N_r. In addition, note that if N_{RF} increases in proportion to N_r, then the values of b^* and B^* will increase with an increase in N_r.

A detailed analysis is shown in Figures 2.6 and 2.7, where we assume $N_{RF} = N_r/8$. This increase in N_{RF} in proportion to N_r ensures that the system performance comparison among HC and DC remains the same. In Figures 2.6 and 2.7, b^* and B^* are plotted, respectively, for different values of N_r while considering both LPADC and HPADC. As expected, b^* and B^* for HC and DC are higher for LPADC than for HPADC. Moreover, b^* and B^* increase

---

[4] Figures 2.4 and 2.5 show that DC consumes higher power than AC. Therefore, we are not further analyzing the condition $P_{Tot}^{DC} \leqslant P_{Tot}^{HC}$

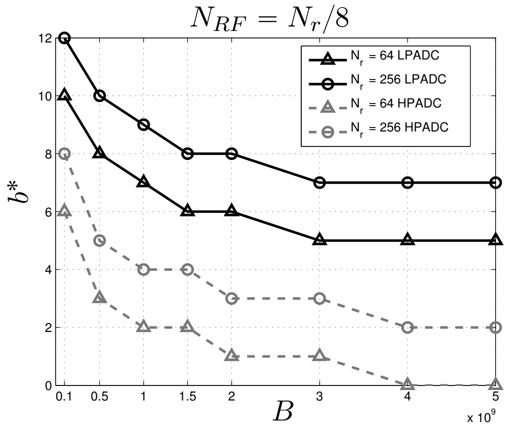

Fig. 2.6. *b^* vs B for HC vs DC considering both HPADC and LPADC.*

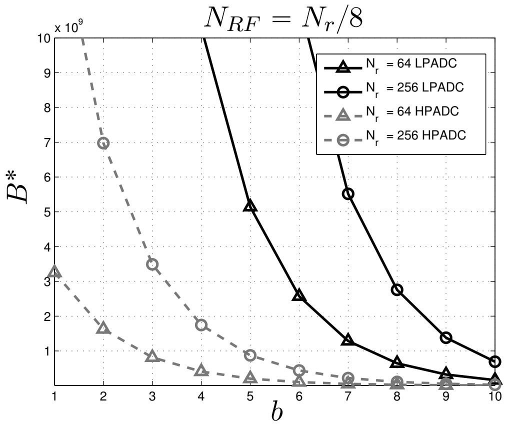

Fig. 2.7. *B^* vs b for HC vs DC considering both HPADC and LPADC.*

with an increase in N_r. This is due to the dependence of N_{RF} on N_r. This shows that large antenna systems with DC based receivers can take advantage of a large bandwidth and/or a higher number of ADC bits while keeping the power consumption similar to that of HC. Moreover, by increasing the N_r/N_{RF} ratio, the number of RF chains for a fixed N_r decreases and therefore P_{Tot}^{HC} decreases. This decrease in P_{Tot}^{HC} results in a reduction of b^* and B^*.

Moreover, note that for $B = 1.5$ GHz and $N_r = 256$, the DC receiver outperforms HC when using ADCs with up to $b = 4$ and $b = 8$ bits for HPADC and LPADC, respectively. These values of b are large enough not to result in any significant SNR loss compared to a high resolution ADC, as discussed in Section 2.2.2. With these configurations, DC may be a preferable option than HC for mmWave receiver design. It is also important to note that with an increase in N_r it is very difficult to acquire the complete channel state information with a fully digital architecture (i.e., DC) as it requires a very high complexity receiver design, whereas HC decreases this complexity but at the cost of lower flexibility. Therefore, the choice between DC and HC may also depend on flexibility and/or complexity, which are directly related to the application requirements. The detailed study of these complexity/flexibility issues is left as a future work. Finally, as B increases b^* decreases (Figure 2.6) and similarly B^* decreases with an increase in b (Figure 2.7). This shows a trade off between the choice of b and B, which means that for a high bandwidth receiver design b should be reduced to keep the total power consumption within the required budget and vice versa.

Table 2.1. *η for different values of b*

b	1	2	3	4	5
η	0.3634	0.1175	0.03454	0.009497	0.002499

2.2.2 ADC Bits vs SNR

We now extend our analysis to study how a change in the number of ADC bits affects the SNR. We consider an AQNM based model as in [44]. For this model, the effective SNR (γ_{ef}) is defined as [46]

$$\gamma_{ef} = \frac{(1-\eta)\gamma}{1+\eta\gamma} \tag{2.11}$$

where γ and γ_{ef} represent the SNR of a high resolution ADC and the effective SNR of a low resolution ADC, respectively, and η is the inverse of the signal-to-quantization-noise ratio of the ADC, which depends on the quantizer design, the input distribution and the number of bits b. For a gaussian input distribution, the values of η for $b \leqslant 5$ are listed in Table 2.1, and for $b > 5$ can be approximated by $\eta = \frac{\pi\sqrt{3}}{2}2^{-2b}$ [38].

Figure 2.8 shows how γ_{ef} varies with the number of ADC bits. Each curve corresponds to a different value of γ, where γ is varied from -10 to 20 dB. The results show that there is a number of bits b_m after which any further increase in b will not result in a significant increase in γ_{ef}, as $\gamma_{ef} \simeq \gamma$ for $b = b_m$. Moreover, they also show that b_m increases as we move from low to high SNR regime. For instance, b_m for $\eta = -10$ dB is 3 bits, whereas for $\eta = 20$ dB it goes up to 6 bits. Therefore, the SNR regime which identifies b_m is also directly related to the choice of the appropriate beamforming schemes.

2.2.2.1 P_{Tot} vs γ_{ef} Comparison:

To summarize the analysis and to better identify the appropriate configuration for DC, we now show a comparison between P_{Tot} and γ_{ef} for different values of b. In particular, we combine the results provided in Figure 2.8 with those of Figures 2.4 and 2.5, and plot the effective SNR performance γ_{ef} vs. the total power consumption, P_{Tot} (where the different points on the curves correspond to different values of b), thereby highlighting the tradeoff between the power spent and the performance achieved.

Figures 2.9 and 2.10 show a comparison between P_{Tot} and γ_{ef} for a low and high power consumption ADC model, respectively. Results are obtained for $\gamma = 10$ dB, $N_{RF} = 4$,

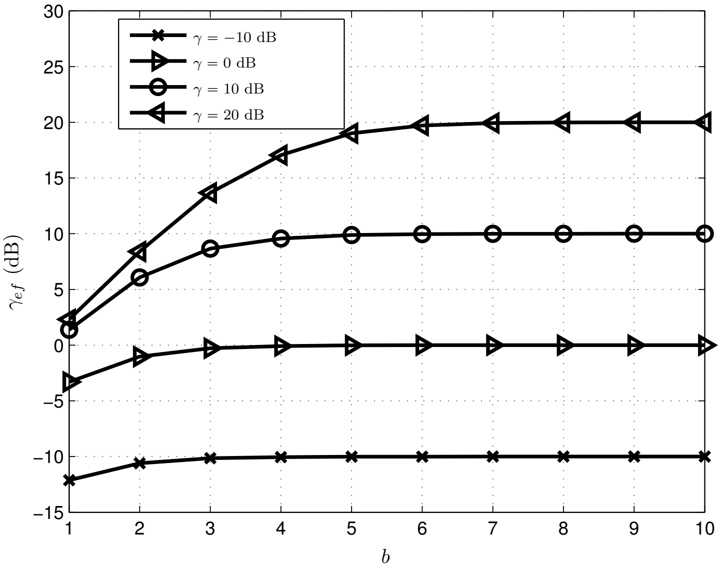

Fig. 2.8. *Change in effective SNR γ_{ef} with a variation in the number of ADC bits b.*

$N_r = 16$ and for $B = 1$ GHz and $B = 100$ MHz. Markers on each curve correspond to increasing values of b when going left to right, where b varies from 1 to 6 and from 1 to 5 in Figures 2.9 and 2.10, respectively. Results show that an appropriate configuration for DC is directly related to the ADC power consumption values. For instance, with LPADC (Figure 2.9), DC has lower power consumption than HC even up to 6 bits, for both $B = 1$ GHz and $B = 100$ MHz. However, with HPADC, DC has a similar power consumption to HC up to 5 and 2 bits for $B = 100$ MHz and $B = 1$ GHz, respectively, and rapidly becomes worse as the number of bits is increased. Note that, with 5 bits, γ_{ef} is almost equal to a value which it can attain with infinite b. Therefore, DC with LPADC is an attractive choice, and preferable to HC for both $B = 100$ MHz and $B = 1$ GHz, whereas with HPADC, DC is a feasible choice only for $B = 100$ MHz.

The results also show that AC is always a better option from a power consumption perspective. However, note that DC for both $B = 100$ MHz and $B = 1$ GHz with LPADC and for $B = 100$ MHz with HPADC results in approximately 30% more power consumption in comparison to AC, and this percentage increases with an increase in B or N_r. Therefore, for receivers with a relatively small number of antennas and low bandwidth requirements, DC may be a preferable choice as, for a limited increase in the total power consumption, it provides significant advantages in terms of flexibility, thanks to digital processing.

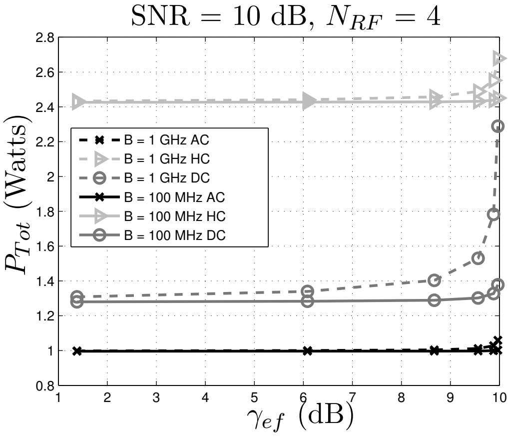

Fig. 2.9. *P_{Tot} vs γ_{ef} for LPADC model.*

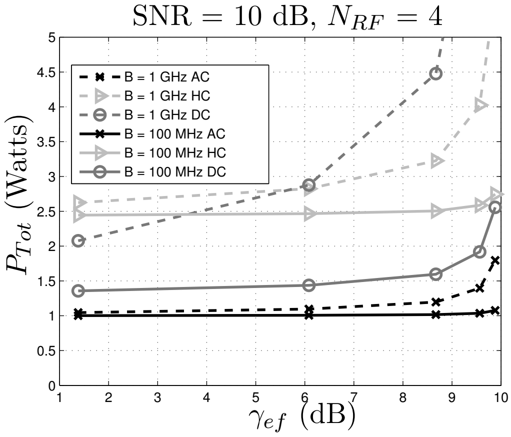

Fig. 2.10. *P_{Tot} vs γ_{ef} for HPADC model.*

2.2.3 Discussion

In this section, we discuss the choice of an appropriate beamforming scheme from the device (the MS or the BS) and the communication signal (control plane or data plane) perspective, as a function of the typical parameters of each configuration. We discuss the choice of the beamforming scheme for the device and the communication signal separately. In the former, we identify an appropriate beamforming scheme both at the MS and at the BS, focusing on their different form factor and application requirements. In the latter, we identify the preferable beamforming scheme focusing on the different bandwidth requirements for control plane (CP) and data plane (DP) communication[5].

2.2.3.1 Optimal Beamforming Scheme at the MS and at the BS

The MS and the BS can accommodate different numbers of antennas and have different application requirements. The MS can accommodate only a small number of antennas due to its small form factor and has simple application requirements, whereas the BS can accommodate a much higher number of antennas and has typically more advanced application requirements such as to ensure multi-user communication, etc.

2.2.3.1.1 Mobile Station To ensure the constraints of a limited power budget and a small form factor, we assume that the MS can have at most 16 antennas (as assumed in most

---

[5] Note that all these comparisons follow from the power consumption values of the receiver components mentioned in Section 2.2.1, and therefore the desirable configuration of any beamforming scheme may vary with the change in the component's power consumption values. However, the general trend among the different beamforming schemes would remain the same, and the corresponding numerical values can be easily derived from our general analysis in Section 2.2.1.

works). It can be seen from Figures 2.9 and 2.10 that with $N_r = 16$, the appropriate beamforming schemes are AC and DC. The choice between these schemes is application dependent. For instance, during initial cell search, where the MS has to look in different angular directions to receive the synchronization signals, the formation of multiple simultaneous beams can be advantageous and therefore justify the additional power consumption. In this case, DC may in fact be a preferable choice than AC as it allows to form multiple simultaneous beams which result in a lower search delay and in a reduced energy consumption, whereas with an AC based receiver the MS has to look in all angular directions sequentially to identify the desired BS, which will increase the initial cell search delay and the total energy consumption. However, if the desired beamforming direction is already available and the advantages of DC are not required (e.g., as in context information based schemes [47]), then AC can be a better option.

2.2.3.1.2 Base Station The base station has to simultaneously serve multiple MSs and, in contrast to the MS, can accommodate higher N_r and has a much higher power available. For analysis, we set the minimum $N_r = 64$ for the BS receiver design. Moreover, to fulfill the requirement of serving multiple MSs, we primarily focus on the comparison of HC and DC, as with AC at a particular instant a BS can communicate with only a single or a limited number of MSs.

As discussed in Section 2.2.1, the choice of the appropriate beamforming scheme between HC and DC depends not only on N_r, b and B, but also on N_{RF}. As shown in Figure 2.6, with LPADC model and low N_{RF}, DC has a similar power consumption to HC up to 6 bits even with $B = 2$ GHz, which makes DC an appropriate choice with LPADC model, whereas in case of HPADC the resultant b^* or B^* for DC is relatively low. Moreover, HC generally allows to simultaneously communicate in only N_{RF} different directions, whereas DC with the same number of antennas can cover a much higher angular space and therefore can communicate with a larger number of MSs. With this difference, and considering an interference free scenario, DC can result in much higher capacity as compared to HC. However, to get the capacity with HC similar to DC and keeping N_r fixed, we need to increase the number of RF chains which then results in an increase in P_{Tot}^{HC}, which corresponds to an increase in B^* and b^*, which may make DC a more preferable design choice than HC even with HPADC.

2.2.3.2 Optimal Beamforming Scheme for Control and Data

Typically, the control plane (CP) and the data plane (DP) have different data rate requirements. The CP has a low data rate requirement, which corresponds to a lower bandwidth, whereas the DP requires higher B to support high data rate. From the MS perspective, a power efficient receiver may require separate beamforming schemes for CP and DP signaling (Figures 2.9 and 2.10). For instance, to reduce the initial cell search delay during CP signaling and based on the low B requirement for CP, DC can be a preferable choice even with HPADC. However, when the beamforming direction is already established and under the high data rate requirements of the DP, AC may be a valid choice, especially at the MS side. On the other hand, for BSs that have to support more advanced applications and to support many users simultaneously, DC may be a preferable choice for both CP and DP signaling.

In the next section, we will study how the capacity and the energy efficiency of different beamforming schemes vary with a change in the number of ADC bits.

2.3 Receiver Spectral and Energy Efficiency

In this section we study and analyze spectral and energy efficiency of a mmWave based MIMO system. Particularly, we study the capacity and energy efficiency of analog, hybrid and digital Combining for a millimeter wave receiver with low resolution ADCs. In our analysis we show that

- SE in a quantized channel approaches that of an ideal unquantized channel as the number of ADC bits grows. Moreover, as SNR increases the number of bits required to approximate the unquantized performance also increases.
- EE is convex in the number of bits and achieves a maximum at a certain optimal resolution that is directly related to the SNR and inversely related to the ADC power figure of merit.
- AC achieves both the best SE and EE when the mmW channel has rank 1 and/or in very low SNR links, and is the only viable architecture under a very stringent power constraint.
- DC always outperforms HC in terms of SE. In some cases it also has better EE, while

in other cases DC with 3-5 bits offers a trade-off with some more SE and only slightly worse EE than HC. This depends critically on how pessimistic are the ADC and Phase Shifter power models, the channel gain impact on SNR, the number of receiver antennas, and the number of parallel RF lines in the HC architecture. There is no one-size-fits-all universally best scheme.

2.3.1 System Model

In 5G mmWave cellular systems, a distinction is usually made between the Base Station (BS) and the User Equipment (UE). Typical UE values are 16 antennas and 1 W power, while typical BS values can be 64 or more antennas and 5 W power in small pico-cells, or 50 W in large macro-cells. Moreover, the system bandwidth varies from 500 MHz up to 7 GHz [34].

For our channel model, however, we simply distinguish between transmitter and receiver roles, where each role may be played by either a BS or a UE depending on the appropriate choice of the parameters for transmitted power and number of transmit and receive antennas. We obtain analytical expressions for SE and EE as a direct function of the number of antennas and of the SNR of the link, obtaining results that apply to all mmWave devices. Note that we considered hybrid architectures and finite quantization only at the receiver, whereas we assume that the transmitter implements ideal Digital Preprocessing (DP). This assumption is used in related works and is the most relevant case for this work in which we study whether AC, HC or DC perform better, due to the fact that 1) if HC outperforms DC using DP, then the use of Hybrid Preprocessing can only make DC even worse; and 2) if DC outperforms HC with DP then switching to HP+DC (where HP stands for Hybrid Processing) or HP+HC to reduce Digital-Analog Converter (DAC) power consumption is never necessary, because the power consumed by the receiver's ADC is of more concern than that of the transmitter's DAC, and therefore if DC is preferable so will be DP.

We consider a point-to-point multiple input multiple output (MIMO) mmWave channel where the transmitter is equipped with N_t antennas and the receiver with N_r antennas. The channel has bandwidth B and a delay spread much smaller than the transmission frame duration, so that inter-symbol interference can be disregarded or simply suppressed using a prefix of negligible duration. Without inter-symbol interference, the received signal can then be expressed as [50]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.12}$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the transmitted and the received symbol vectors at discrete time instants with period $1/B$, respectively, $\mathbf{n}$ is the i.i.d. circularly symmetric complex Gaussian noise vector, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_o\mathbf{I})$, and $\mathbf{H}$ represents the $N_r \times N_t$ channel matrix that varies following a fast block-fading model that remains constant for a small number of symbols and takes independent identically distributed values across blocks; this means that the rate of the system is the *ergodic rate* (average mutual information over the realizations of $\mathbf{H}$). The mmWave channel matrix is randomly distributed following a random geometry with a moderate number of propagation paths (order of tens) grouped in very few clusters of similar paths (average 1.9) [14, 48]

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{\rho N_c N_p}} \sum_{k=1}^{N_c} \sum_{\ell=1}^{N_p} g_{k,\ell}\mathbf{a}_r(\phi_k + \Delta\phi_{k,\ell})\mathbf{a}_t^H(\theta_k + \Delta\theta_{k,\ell}) \tag{2.13}$$

where ρ is the distance dependent path-loss, N_c is the number of independent clusters, N_p represents the number of paths per cluster, $g_{k,\ell} \sim \mathcal{CN}(0, 1)$ is the small scale fading associated with the ℓ^{th} path of the k^{th} cluster, ϕ_k and $\theta_k \in [0, 2\pi)$ represent the mean angle of arrival (AoA) and angle of departure (AoD) of the k^{th} cluster at the receiver and at the transmitter, respectively. The AoA and AoD of each path within each cluster vary around the mean direction of that cluster, with a standard deviation θ_{RMS}. We represent by $\Delta\phi_{k,\ell}$ and $\Delta\theta_{k,\ell}$ $\sim \mathcal{N}(0, \theta_{RMS}^2)$ the differential AoA and AoD of the ℓ^{th} path of the k^{th} cluster.

Here, we model the antenna arrays at both the transmitter and the receiver as uniform linear arrays (ULA) with adjacent antenna spacing of half the wavelength of the transmitted signal ($\lambda/2$). Under this model, a spatial signature vector $\mathbf{a}_t$ for the transmit array can be expressed as a function of the AoD as follows

$$\mathbf{a}_t = \frac{1}{\sqrt{N_t}}[1, e^{j\pi\sin(\theta)}, ..., e^{j(N_t-1)\pi\sin(\theta)}]^T \tag{2.14}$$

where T denotes transpose, and an analogous expression characterizes the spatial signature vector for the receiver, $\mathbf{a}_r$. Finally, for a 28 GHz channel, the path-loss from [51] is

$$\rho_{LOS}(dB) = 61.5 + 20\log_{10}(d) + \xi,\ \xi \sim \mathcal{N}(0, 5.8),$$

$$\rho_{NLOS}(dB) = 72 + 29.2\log_{10}(d) + \xi,\ \xi \sim \mathcal{N}(0, 8.7),$$

where d represents the distance between the transmitter and the receiver on a straight line while the variation of distances traversed by different paths is captured in $g_{k,\ell}$.

The parameters suggested to model 28 GHz channels in the literature [51] are $N_p = 20$, $N_c \sim \max\{\text{Poisson}(1.8), 1\}$ and $\theta_{RMS} \sim 10^o$. In our evaluations, in addition to $N_c \sim \max\{\text{Poisson}(1.8), 1\}$ and $N_p = 20$, we also consider a rank 1 channel with $N_c = N_p = 1$ to study the effect of the low rank channel matrix $\mathbf{H}$ on the choice of an appropriate combining scheme.

2.3.1.1 Quantized Received signal

The rate of the quantized MIMO channel with a 1 bit ADC under an exact non linear quantization model is shown in [32]. However, such exact non-linear models are difficult to derive for a higher number of bits. A common approximation for the performance of quantized systems is the lower bound for the achievable rate, obtained modeling the quantization as an additive Gaussian noise with power inversely proportional to the resolution of the quantizer, that is, 2^{-b} times the receiver input power where b is the number of ADC bits. In recent studies [34], [38], this Additive Quantization Noise Model (AQNM) has been applied to the study of mmWave quantized signal modeling with an arbitrary number of ADC bits.

2.3.1.2 Received Signal Model with AQNM

We consider that the received signals at each antenna may be subject to some analog processing prior to quantization. This RF-processed received signal is converted to the digital domain by multiple ADCs (one ADC for each inphase and quadrature component for each vector dimension). The AQNM represents the quantized version $\mathbf{y}_q$ of the received signal (3.1) as

$$\mathbf{y}_q = (1 - \eta)(\mathbf{Hx} + \mathbf{n}) + \mathbf{n}_q \tag{2.15}$$

where $\mathbf{n}_q$ is the additive quantization noise and η is the inverse of the signal-to-quantization noise ratio, which is inversely proportional to the square of the resolution of an ADC (i.e., $\eta \propto 2^{-2b}$). For a Gaussian input distribution, the values of η for $b \leqslant 5$ are listed in Table 2.1, and for $b > 3$ can be approximated by $\eta = \frac{\pi\sqrt{3}}{2}2^{-2b}$ [38]. We denote by $\boldsymbol{\gamma}_q$ the signal-to-noise ratio SNR of $\mathbf{y}_q$, given by

$$\gamma_q = (1-\eta)^2(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H)((1-\eta)^2 N_o\mathbf{I} + \mathbf{R}_{\mathbf{n}_q\mathbf{n}_q})^{-1} \tag{2.16}$$

where superscript H denotes conjugate transpose, $\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H$ is the received signal at the output of the quantizer, $\mathbf{R}_{\mathbf{xx}}$ is the input covariance, N_o is the noise power, and $\mathbf{R}_{\mathbf{n}_q\mathbf{n}_q} = \eta(1-\eta)(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H) + N_o\mathbf{I}$ [52] is the covariance of the quantization noise. Substituting $\mathbf{R}_{\mathbf{n}_q\mathbf{n}_q}$ in (2.16) yields

$$\gamma_q = (1-\eta)(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H)(N_o\mathbf{I} + \eta(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H))^{-1}, \tag{2.17}$$

and finally, in terms of the SNR of the unquantized signal (γ), γ_q can be written as [46]

$$\gamma_q = \frac{(1-\eta)\gamma}{1+\eta\gamma} \tag{2.18}$$

At low SNR, γ_q can be approximated as $(1-\eta)\gamma$, while at high SNR and for finite bits b, the quantized SNR γ_q is tightly upper bounded by $\min(\frac{1-\eta}{\eta}, \gamma)$. Note that, for a very high resolution, $\eta \to 0$, and γ_q in (2.17) will be equal to the SNR of the unquantized signal γ. Finally, the achievable rate of the MIMO link with an AQNM signal in (2.15) is given as

$$C_q = \mathrm{E}_{\mathbf{H}}\left[\max_{\mathbf{R}_{\mathbf{xx}}} B\log_2\left|\mathbf{I} + (1-\eta)(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H)(N_o\mathbf{I} + \eta(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H))^{-1}\right|\right] \tag{2.19}$$

2.3.2 Receiver Architectures

This section characterizes the achievable rates in mmWave links with three types of receiver architectures featuring quantization. For all three cases, the transmitter architecture is considered to be always fully digital with ideal MIMO processing. We also assume the availability of channel state information (CSI) both at the transmitter and at the receiver and we design the MIMO processing accordingly.

The three architectures are termed Analog, Digital, and Hybrid, and represented in Figures 2.1, 2.2 and 2.3, respectively. The difference between the three architectures consists in their different analog processing hardware prior to the ADC, which modifies the total number of RF and ADC units, and the number of digital signal dimensions that may be further processed by digital stages.

2.3.2.1 Analog Combining

The first architecture is Analog Combining, which is motivated by the fact that typically ADCs are considered as the most power-hungry receiver components. Therefore, in this architecture all multiple-antenna processing is performed in the analog domain to minimize power consumption. The architecture in Figure 2.1 features one phase shifter per receive antenna and an analog signal adder; together, these devices implement analog beamforming and deliver a scalar combined signal to a one-dimensional RF and ADC chain.

The quantized received signal y_q with analog combining at the receiver and digital beamforming at the transmitter is given by

$$y_q = (1-\eta)(\mathbf{w}_r^H \mathbf{H} \mathbf{w}_t x + \mathbf{w}_r^H \mathbf{n}) + \mathbf{n}_q \tag{2.20}$$

where $\mathbf{w}_t$ represents the digital beamforming vector at the transmitter such that $||\mathbf{w}_t||^2 = 1$ and $\mathbf{w}_r$ is the analog combining vector at the receiver with a constant amplitude per coefficient $|w_{r,i}| = 1/\sqrt{N_r}$ due to its implementation using phase shifters.

The additive quantization noise variance is given by $\eta(1-\eta)(|\mathbf{w}_r^H \mathbf{H} \mathbf{w}_t|^2 P + N_o)$, where P is the average transmitter power. Finally, the ergodic rate maximization problem[6] with analog combining is given as

$$\begin{aligned} C_{AC} = \mathrm{E}_{\mathbf{H}}\Bigg[&\max_{\mathbf{w}_r,\mathbf{w}_t} B\log_2\left(1+\frac{(1-\eta)|\mathbf{w}_r^H \mathbf{H} \mathbf{w}_t|^2 P}{N_o + \eta|\mathbf{w}_r^H \mathbf{H} \mathbf{w}_t|^2 P}\right)\Bigg] \\ &s.t.\ |w_{r,i}| = \frac{1}{\sqrt{N_r}}, \\ &||\mathbf{w}_t||^2 = 1, \end{aligned} \tag{2.21}$$

Due to the fact that CSI is available for each channel realization, and the fact that (2.21) is a monotonic function of $|\mathbf{w}_r^H \mathbf{H} \mathbf{w}_t|^2$, the maximization of the ergodic rate is achieved by maximizing this beamforming gain independently for each channel realization[7].

---

[6] Notice that the achievable rates are evaluated based on a fast fading channel model (as used in [35]) which may have practical limitations for mmWave channels due to the very short coherence time. However, capacity evaluation based on long term channel statistics (which requires a separate system model) is left for future work.

[7] Note that at a particular time instant AC can only look in a single combining direction and therefore it requires a longer duration than DC or HC to acquire the complete CSI in all angular directions. This delay may degrade the performance of the AC scheme. However, in this work we compute the achievable rate of the AC scheme considering that the CSI of all angular direction is already available and therefore we are not incorporating the loss related to the identification of the right beamforming vector(or direction). However, note that the change in the directional beamforming vector is related to the long term channel statistics

The transmitter beamforming vector, which has fewer constraints, can be simply assigned the value that maximizes the gain for a given value of $\mathbf{w}_r$. This consists in implementing a matched filter at the transmitter with value $\mathbf{w}_t = \frac{\mathbf{H}^H \mathbf{w}_r}{||\mathbf{H}^H \mathbf{w}_r||^2}$, where the normalization is required to satisfy the transmit power constraint.

By using the matched filter at the transmitter, we can rewrite the problem as finding the receive beamforming vector that maximizes $|\mathbf{w}_r^H \mathbf{H}|^2$. If $\mathbf{w}_r$ had no per-coefficient amplitude constraints, the optimal receive beamforming would be the eigenvector $\mathbf{u}_{\max}$ associated with the largest singular value $\sigma_{\max}$ of matrix $\mathbf{H}$, i.e.,

$$\mathbf{u}_{\max} = \arg \max_{\mathbf{w}_r \in \mathbb{C}^{N_r}} |\mathbf{w}_r^H \mathbf{H}|^2, \tag{2.22}$$

However, since the analog scheme can only alter the phase of a constant-amplitude beamforming vector, the exact optimal analog beamforming vector is, in turn, expressed as

$$\mathbf{w}_r^* = \arg \max_{\mathbf{w}_r : |w_r^i| = 1/\sqrt{N_r}} |\mathbf{w}_r^H \mathbf{H}|^2. \tag{2.23}$$

Finally, due to the fact that the problem in (2.23) is more difficult, in our model we consider that the receiver instead settles for an approximate solution consisting in the projection of the unconstrained optimal beamforming vector $\mathbf{u}$ from (2.22) to the nearest point over the space of constant-amplitude vectors, i.e.,

$$\tilde{\mathbf{w}}_r^H = \frac{1}{\sqrt{N_r}} (e^{\measuredangle u_{\max}^1}, e^{\measuredangle u_{\max}^2}, \dots e^{\measuredangle u_{\max}^{N_r}})^T \tag{2.24}$$

2.3.2.2 Digital Combining

The second architecture is Digital Combining, which is motivated by the fact that digital MIMO processing has in general fewer constraints and can achieve higher gains. Therefore, in this architecture all multiple-antenna processing is performed in the digital domain to maximize the rate. The architecture in Figure 2.2 features no analog processing; each antenna directly delivers its received signal to a dedicated RF and ADC chain. A quantized signal with N_r dimensions is processed by a Digital MIMO processor that allows the spatial multiplexing of up to $N_s \leqslant \min(N_t, N_r)$ symbol streams. The quantized received signal with digital

---

whereas we evaluate the achievable rates considering a fast fading channel. Therefore, the loss associated with the selection of the right beamforming vectors can be negligible.

combining is given as

$$\mathbf{y}_q = (1-\eta)(\mathbf{W}_r^H\mathbf{H}\mathbf{W}_t\mathbf{x} + \mathbf{W}_r^H\mathbf{n}) + \mathbf{W}_r^H\mathbf{n}_q \tag{2.25}$$

where $\mathbf{W}_r$ and $\mathbf{W}_t$ are the digital combining and beamforming matrices, respectively. Note that, as combining is performed after quantization, the DC matrix also multiplies the quantization noise. To calculate the supremum achievable rate with DC, we design the beamforming and combining matrices corresponding to the singular value decomposition of the channel matrix i.e., $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular matrices, respectively, and Σ is a diagonal matrix with the singular values. Now, by applying a transmit beamforming matrix $\mathbf{W}_t = \mathbf{V}$ and receive combining matrix $\mathbf{W}_r = \mathbf{U}^H$, Eq. (2.25) can be written as

$$\begin{aligned}\mathbf{y}_q &= (1-\eta)(\mathbf{U}^H\mathbf{U}\Sigma\mathbf{V}^H\mathbf{V}\mathbf{x} + \mathbf{U}^H\mathbf{n}) + \mathbf{U}^H\mathbf{n}_q \\ &= (1-\eta)(\Sigma\mathbf{x} + \mathbf{U}^H\mathbf{n}) + \mathbf{U}^H\mathbf{n}_q\end{aligned} \tag{2.26}$$

Finally, we allocate the transmit power across the singular values of Σ using the water filling algorithm, and the rate with DC results in

$$C_{DC} = \mathrm{E}_{\mathbf{H}}\left[\max_{\mathbf{R_{xx}}} B\log_2\det\left|\mathbf{I} + (1-\eta)\Sigma\mathbf{R_{xx}}\Sigma^H(N_o\mathbf{I} + \eta\mathbf{U}^H\mathrm{diag}(\mathbf{U}\Sigma\mathbf{R_{xx}}\Sigma^H\mathbf{U}^H)\mathbf{U})^{-1}\right|\right] \tag{2.27}$$

where the input covariance matrix that maximizes the rate for each channel realization, $\mathbf{R_{xx}}$, appears also in the noise term. This means that with b bits the optimal $(\mathbf{R_{xx}})_b^*$ is not exactly the same as the optimal for the unquantized channel, $(\mathbf{R_{xx}})_\infty^*$, calculated using the water-filling algorithm. However, as the number of quantization bits grows we have $\lim_{b\to\infty}(\mathbf{R_{xx}})_b^* = (\mathbf{R_{xx}})_\infty^*$. Therefore, in our DC model we use a near-optimal water-filling input distribution that lower bounds the achievable rates for DC and approaches the optimal as b increases.

At low SNR, the waterfilling algorithm allocates all the power to the maximum singular value, and therefore the beamforming and combining vectors are just the right and left singular vectors ($\mathbf{v}_{\max}$ and $\mathbf{u}_{\max}$) corresponding to the maximum singular value, respectively. This shows how the optimal number of spatial streams, N_s, may be smaller than $\min(N_t, N_r)$; compared with AC, which only provides beamforming gain, DC can provide the advantages of both spatial multiplexing gains at high SNR and beamforming power gains at low SNR.

Although DC achieves a rate similar to AC at low SNR, the former is always better than the latter, due to the fact that the DC architecture does not impose constant-amplitude constraints in the receive combining vector. The rate with DC as $SNR \to 0$ is given as

$$C_{DC} \overset{SNR\to 0}{=} \mathrm{E}_{\mathbf{H}}\left[B\log_2\left(1+\frac{(1-\eta)|\mathbf{u}_1^H\mathbf{H}\mathbf{v}_1|^2P}{N_o+\eta|\mathbf{u}_1^H\mathbf{H}\mathbf{v}_1|^2P}\right)\right] \tag{2.28}$$

2.3.2.3 Hybrid Combining

The third architecture is Hybrid Combining, which is motivated by the high rate and the low power consumption of Digital and Analog combining architectures, respectively, and tries to strike a balance between the two. The architecture in Fig. 2.3 features an analog processing stage with multiple banks of phase shifters, each with an independent analog adder, RF and ADC chain. The analog processing reduces the dimensions of the received signal to a number N_{RF} greater than one (analog case) but smaller than N_r (digital case). The analog processed signal with N_{RF} dimensions is quantized and digitally processed allowing the spatial multiplexing of up to $N_s \leqslant \min(N_t, N_{RF}) \leqslant N_r$ symbol streams. The quantized signal with HC is given by

$$\mathbf{y}_q = (1-\eta)(\mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{H}\mathbf{W}_t\mathbf{x} + \mathbf{W}_{BB}^H\mathbf{W}_{RF}^H\mathbf{n}) + \mathbf{W}_{BB}^H\mathbf{n}_q \tag{2.29}$$

where $\mathbf{W}_{RF}$ and $\mathbf{W}_{BB}$ are the RF and the baseband combining vectors, respectively. Let $\mathbf{n}_e = \mathbf{W}_{RF}^H\mathbf{n}$ represent equivalent receiver noise, $\mathbf{H}_e = \mathbf{W}_{RF}^H\mathbf{H}$ represent an $N_{RF} \times N_t$ equivalent channel matrix, and $\mathbf{H}_e = \mathbf{U}\Sigma\mathbf{V}^H$ represent its singular value decomposition. The digital baseband combiner may be set to $\mathbf{W}_{BB} = \mathbf{U}^H$, and its corresponding transmit beamformer to $\mathbf{W}_t = \mathbf{V}$, which leaves the quantized signal in (2.29) written as

$$\mathbf{y}_q = (1-\eta)(\Sigma\mathbf{x} + \mathbf{U}^H\mathbf{n}_e) + \mathbf{U}^H\mathbf{n}_q \tag{2.30}$$

Similar to the analog case, the design of $\mathbf{W}_{RF}$ is affected by the phase shifter with constant amplitude constraints. The unconstrained optimization should maximize the result of a water-filling over the N_{RF} singular values of the equivalent channel $\mathbf{H}_e = \mathbf{W}_{RF}\mathbf{H}$, and results in the following optimization problem

$$\mathbf{W}_{RF}^{ideal} = \arg \max_{\mathbf{W}_{RF} \in \mathbb{C}^{N_r \times N_{RF}}} \max_{\sum p_i = P} \sum_{\mathcal{S}(\mathbf{W}_{RF}\mathbf{H})} \sigma_i^2 p_i, \tag{2.31}$$

where $\mathcal{S}(\mathbf{W}_{RF}\mathbf{H})$ denotes the spectrum of the matrix and σ_i^2 is the i^{th} eigenvalue. However, the analog processors can only alter the phase of constant-amplitude values of $\mathbf{W}_{RF}$. This, in turn, is expressed as

$$\mathbf{W}_r^* = \arg \max_{\mathbf{W}_{RF}:|W_r^{ij}|=1/\sqrt{N_r}} \max_{\sum p_i = P} \sum_{\mathcal{S}(\mathbf{W}_{RF}\mathbf{H})} \sigma_i^2 p_i. \tag{2.32}$$

However, the latter optimization poses a number of issues. First, the selection of the best beamforming matrix in (2.32) is not a linear problem because the target function contains a power allocation. Second, the formulation of (2.32) satisfies the norm constraints but does not satisfy the semi unitary constraints of the hardware. Therefore, to satisfy both norm and semi unitary constraints we select $\mathbf{W}_{RF}$ using a practical approximate solution by using an alternate projection method [53]. The procedure is summarized in Algorithm 1 and consists in an alternate-projection iterative method. The algorithm starts by generating a semi unitary matrix $\mathbf{W}_{SU}$, which is equal to the first N_{RF} vectors of the left singular matrix of $\mathbf{H}$. In the next step, the RF combining matrix $\tilde{\mathbf{W}}_{RF}$ is generated such that the corresponding elements have similar phase as that of $\mathbf{W}_{SU}$ but with constant normalized amplitude. In the next step, $\tilde{\mathbf{W}}_{RF}$ is again projected back to the semi unitary matrix $\mathbf{W}_{SU}$. The process continues until the algorithm converges (i.e., the matrix coefficients change less than a small step threshold value). The convergence properties of the algorithm are studied in [53]. This results in a similar $\mathbf{y}_q$ as obtained with DC in (2.26). However, $\mathbf{U}$ has a dimension of N_{RF} instead of

Algorithm 1 Alternate projection method for design RF combining matrix for HC

```
Initialize $\mathbf{W}_{SU} = [\mathbf{u}_1 \ldots \mathbf{u}_{N_{RF}}] \in \mathbf{U}$,where $\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^H$
while not converging do
    $[\tilde{\mathbf{W}}_{RF}]_{ij} = \frac{1}{\sqrt{N_r}} \exp(j\measuredangle[\mathbf{W}_{SU}]_{ij}), \forall i, j$
    $\mathbf{W}_{SU} = (\tilde{\mathbf{W}}_{RF}\tilde{\mathbf{W}}_{RF}^*)^{-\frac{1}{2}}\tilde{\mathbf{W}}_{RF}$
end while
```

N_r, due to the limited number of RF chains. The achievable rate with HC is given as

$$C_{HC} = \mathrm{E}_{\mathbf{H}}\left[\max_{\mathbf{R_{xx}}} B \log_2 \det \left|\mathbf{I} + (1-\eta)\Sigma\mathbf{R_{xx}}\Sigma^H(N_o\mathbf{I} + \eta\mathbf{U}^H \mathrm{diag}(\mathbf{U}\Sigma\mathbf{R_{xx}}\Sigma^H\mathbf{U}^H)\mathbf{U})^{-1}\right|\right] \tag{2.33}$$

Although the two rate expressions for DC and HC are similar, the rate of HC is upper bounded by the rate of DC due to the fact that the selection of $\mathbf{W}_{RF}$ is subject to a constant amplitude constraint associated with the analog combiners. We also point out that in our formulation we use infinite-resolution phase shifters for AC or HC, whereas in practice only finite-resolution quantized phase steps are employed. Therefore, the rates achieved following (2.21) and (2.33) are in fact greater than what would be obtained with a quantized phase shifter constraint.

Lastly, note that as was the case with DC, the input covariance matrix that maximizes the HC rate with b bits, $(\mathbf{R_{xx}})^*_b$, is not that which is calculated with the water-filling algorithm, $(\mathbf{R_{xx}})^*_\infty$. This means that our implementation using water-filling is slightly lower than the maximal rates achievable with a HC scheme in (2.33). Nevertheless, our comparison between DC and HC remains valid due to the fact that we apply a simplification of the same type to both models and, since the simplification is worse for low b and we are primarily interested in when DC with a few bits outperforms HC with many bits, in fact this simplification is slightly skewed in favor of HC.

2.3.2.4 Spectral Efficiency Calculation

We compare the average achievable rates for AC, DC and HC mmWave links as a function of the number of ADC bits. Figures 2.11 and 2.12 show the rate vs ADC bits when the number of propagation clusters and paths per cluster (N_c, N_p) are $(1, 1)$ for an ideal rank 1 channel, and when they are $(Poisson(1.8), 20)$ for a more realistic channel model as in [51], respectively. We evaluate the theoretical integrals for the averages of Eqs. (2.21), (2.27), and (2.33) using the Monte-Carlo numerical integration method with 1000 realizations of the channel distribution per point. We consider $N_t = 64$, $N_r = 16$, $N_{RF} = 4$, bandwidth $B = 1$ GHz, and a total transmit power of 30 dBm. We show results for -20 and 0 dB SNR not including antenna gain, which corresponds to an approximate communication range of 100 m for NLOS and LOS, respectively.

In both figures we can see that, for all architectures, the rate grows up to a certain number

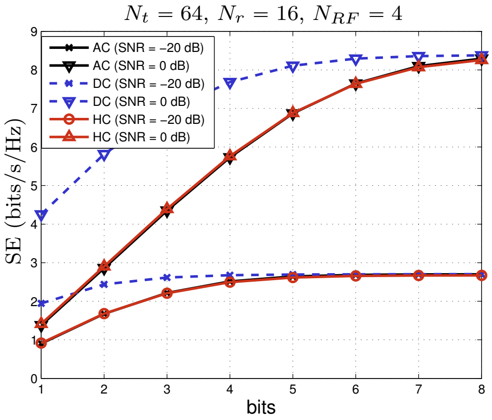

Fig. 2.11. *Achievable rate vs ADC bits comparison for AC, DC and HC schemes for* $(N_c, N_p) = (1, 1)$.

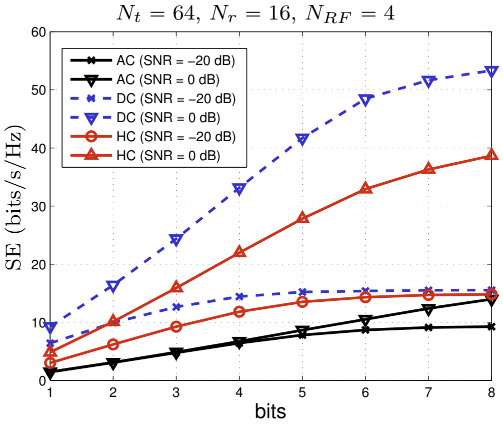

Fig. 2.12. *Achievable rate vs ADC bits comparison for AC, DC and HC schemes for* $(N_c, N_p) = (Poisson(1.8), 20)$.

of ADC bits and saturates afterwards, so that a further increase in b does not improve the SNR of the quantized signal. The threshold appears later when the unquantized signal SNR (γ) is higher, and more bits are necessary to reach saturation.

The results show that the DC architecture outperforms the other schemes for both cases of propagation paths. This is due to the fact that DC does not have constant amplitude constraints in the beamforming coefficients. Moreover, in the case of $(N_c, N_p) = (1, 1)$, i.e., if the channel has rank 1, spatial multiplexing is not possible, and AC can perform similarly to HC or DC, but only for a high number of ADC bits. In the opposite case, for $(N_c, N_p) = (Poisson(1.8), 20)$, AC achieves significantly lower rates at high SNR because it cannot exploit spatial multiplexing, unlike HC and DC. At low SNR, AC is closer to HC and DC because the water-filling algorithm concentrates all the power in one eigenvalue and there is no multiplexing gain anyways. This shows that the appeal of AC schemes is strongest for single-path sparse propagation environments and low-SNR links. Finally, even though HC exploits spatial multiplexing, its rate is always slightly lower than DC due to the constant amplitude constraint of the RF combining matrix ($\mathbf{W}_{RF}$) and also to the fact that DC's spatial multiplexing gain is not upper bounded by N_{RF}.

2.3.3 Energy Efficiency Analysis

Looking only at rate, the straightforward choice for a mmWave receiver design would be a fully digital architecture (i.e., DC), which can exploit the maximum advantages of both beamforming and spatial multiplexing techniques, outperforming AC and HC. However,

Table 2.2. *Power consumption of each device*

Device	Notation	Value
Low Noise Amplifier (LNA) [41]	P_{LNA}	39 mW
Splitter	P_{SP}	19.5 mW
Combiner [41]	P_C	19.5 mW
Phase shifter [54, 55]	P_{PS}	2 mW or 0
Mixer [45]	P_M	16.8 mW
Local oscillator [37]	P_{LO}	5 mW
Low pass filter [37]	P_{LPF}	14 mW
Base-band amplifier [37]	$P_{BB_{amp}}$	5 mW
ADC	P_{ADC}	$cB2^b$

generally, these advantages of DC are tied to a higher power consumption at the receiver. Thus, although DC results in the maximum achievable rates, it may not be an energy efficient receiver option.

In particular, in the large bandwidth operation expected at mmW, the ADC is usually considered to be the most power hungry block and thus the power consumption of DC is penalized by its high number of ADCs (N_{ADC}), equal to twice the number of receive antennas. In comparison, AC, only requiring 2 ADCs, would be the the least power consuming scheme, and HC, requiring $N_{ADC} = 2 \times N_{RF}$, is generally assumed to have a power consumption in-between AC and DC.

Nonetheless, looking only at the ADC, and disregarding the power consumption of other analog components that are not necessary for DC, may be misleading. Particularly, if the power consumption of phase shifters and analog combiners is non-negligible, HC may be penalized due to the fact that it requires a large number (up to N_{RF}) of both analog blocks and ADCs at the same time.

In this section we develop a detailed study of the EE of the three architectures, defined as

$$EE = \frac{C_q}{P_{Tot}} \tag{2.34}$$

where C_q is the achievable rate of the quantized signal corresponding to different combining schemes and P_{Tot} is the total power consumption of the mmWave receiver design corresponding to analog, digital and hybrid combining architectures.

Table 2.3. *ADC power per sample and per level, c*

Scenario	Value	Generation
LPADC	5 fJ/step/Hz	Ideal future value
IPADC	65 fJ/step/Hz	Intermediate Power (Recently proposed)
HPADC	494 fJ/step/Hz	State of the art

2.3.3.1 Power Consumption Model

The devices required to implement each mmWave receiver architecture are displayed in Figures 2.1, 2.2 and 2.3, respectively. The total power consumption P_{Tot} of AC, HC and DC schemes is evaluated by the expressions shown in Equations (2.1), (2.2), and (2.3), respectively. The component power consumptions are detailed in Table 2.2. The power consumption of all components except the ADC is independent of the bandwidth B and the number of bits b, whereas P_{ADC} increases exponentially with b and linearly with B and with the ADC Walden's figure of merit c [43] (the energy consumption per conversion step per Hz).

It must be noted that, depending on the choice of power values for each component, our model may give a different outcome about whether HC or DC performs better. For example, if we used a very high ADC Walden's figure of merit, e.g., 12.5 pJ/step/Hz as in [42], we would heavily penalize DC, giving an unfair advantage to HC. Likewise, if we considered a very high Phase Shifter power consumption, such as 19.5 mW in [41], we would be similarly giving an unfair advantage to DC.

In this work we have selected two reasonable component power consumption models that we approximately identify with two generations of technology, current and upcoming. The considered ADCs Walden's figure of merit is detailed in Table 2.3.

- The High Power ADC (HPADC) model is based on an existing device that supports sampling at Gs/s and has been referenced in related literature such as [34]. In order to give Phase Shifters appropriate power consumption values, we pair the existing ADC model with an existing PS model with $P_{PS} = 2$ mW, referenced in [54].

- The Low Power ADC (LPADC) model considers a likely future best-case scenario deduced from the hardware survey in [56]. Likewise, we pair this "best case future scenario" with a best case phase shifter model, with negligible power consumption ~ 0

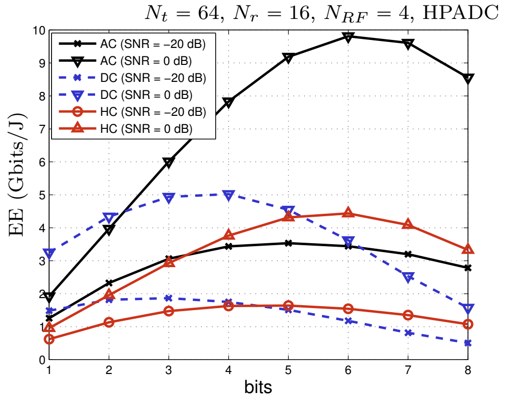

Fig. 2.13. *EE vs.ADC bits comparison for AC, DC and HC schemes for* $(N_c, N_p) = (1, 1)$mW, as in [55][8].

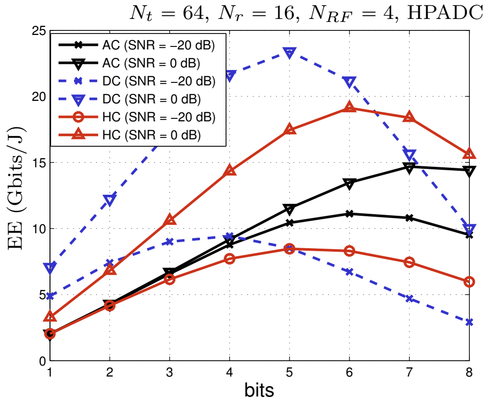

Fig. 2.14. *EE vs.ADC bits comparison for AC, DC and HC schemes for* $(N_c, N_p) = (Poisson(1.8), 20)$.

2.3.3.2 Energy Efficiency Calculation

We analyze the EE of each receiver architecture using the power values per component defined in Tables 2.2 and 2.3. We show the EE vs number of ADC bits in the HPADC scenario for two SNR values, −20 dB and 0 dB, when the number of propagation clusters and paths per cluster (N_c, N_p) are $(1, 1)$ and $(Poisson(1.8), 20)$, in Figures 2.13 and 2.14, respectively. In this analysis we have considered a mmWave antenna array with $N_t = 64$, and $N_r = 16$, a hybrid scheme with $N_{RF} = 4$, and a channel with bandwidth $B = 1$ GHz.

The plots show that all combining schemes have an optimal number of ADC bits which results in a maximum EE, with EE increasing for b lower than the optimal, and decreasing for any further increase of b above the optimal. The optimal point is influenced by both the flat saturation of rate as a function of b, which depends on SNR, and by the exponential increase in ADC power consumption with b.

Comparing the three receiver architectures, we note that surprisingly DC offers the highest EE of the three schemes with a HPADC model in a dense multi-path environment with high SNR (Figure 2.14 for 0 dB) whereas AC achieves highest EE in case of a single cluster/single path scenario or low SNR (Figure 2.13 and Figure 2.14 for −20 dB). Note that DC EE decays rapidly when b is beyond the optimal, and thus HC outperforms DC when both have a high number of bits. However, it is incorrect to extrapolate from this that HC always outperforms

---

[8] ADC power values are rapidly changing, and we consider the extreme cases of well-established HPADC [34] and future LPADC [56]. Very recently a new Intermediate Power ADC (IPADC) has been proposed in [57], given this dynamic situation, we complement our study with a method and a web visualization tool so that researchers can reproduce the charts and easily bring in the advances in ADC technology to our model.

DC. Also note that in Fig. 2.14 at high SNR, DC has higher EE than other schemes up to 6 bits, and this number of bits is high enough to avoid any noticeable degradation in SE (see Fig. 2.12). We also mention that the number of ADC bits also affects the choice of the modulation scheme, for instance ADCs with only 1 or 2 bits cannot support higher order modulation schemes such as 64-QAM, which may also result in a reduction of the achievable rate compared to a high resolution ADC.

Comparing the SNR across the same architecture, we observe that the rate saturation occurs at a higher number of bits when SNR is larger, and therefore the optimal number of bits increases with SNR. Moreover, when the SNR is low, AC has better performance than HC and DC, due to the fact that at low SNR the water-filling algorithm focuses all the power in a single singular value of the channel and HC and DC do not exploit spatial multiplexing gains.

It must be noted that the DC receiver with highest EE and the receiver with the highest SE do not have the same number of bits. For this reason, in the next section we develop a framework to compare receivers in the two dimensions.

2.3.4 Trade-off Analysis and Effect of Parameters

In this section, we develop the full two-dimensional SE and EE comparison of DC, HC, and AC receiver architectures. To this end, we create comparison charts that represent the EE and the SE of each receiver design, making it possible to study the choice of appropriate receiver schemes for different needs, and to observe the trade-off between the two metrics. We write EE as a function of SE as

$$\begin{aligned} EE = \frac{B \times SE}{P_o + P_a N_r + 2N_r P_{ADC}} &\overset{(a)}{=} \frac{SE}{(P_o + P_a N_r)/B + 2N_r c 2^b} \\ &\overset{(b)}{\simeq} \frac{SE}{(P_o + P_a N_r)/B + 2N_r c\sqrt{\frac{2}{\pi\sqrt{3}}\frac{\gamma}{(1+\gamma)2^{-SE}-1}}} \end{aligned} \tag{2.35}$$

where we denote by P_o the power consumption of all fixed receiver components and by P_a the power consumption of all per-antenna analog receiver components except ADCs, (a) comes from $P_{ADC} = cB2^b$ in Table 2.2 and (b) comes from combining (2.18) and $\eta \simeq \frac{\pi\sqrt{3}}{2}2^{-2b}$.

This expression displays three behavior regimes:

1. At $(P_o + P_a N_r)/B \gg 2N_r c\sqrt{\frac{2}{\pi\sqrt{3}}\frac{\gamma}{(1+\gamma)2^{-SE}-1}}$ EE grows linearly with SE.

2. At $(P_o + P_a N_r)/B \simeq 2N_r c\sqrt{\frac{2}{\pi\sqrt{3}}\frac{\gamma}{(1+\gamma)2^{-SE}-1}}$ a transition occurs.

3. At $(P_o + P_a N_r)/B \ll 2N_r c\sqrt{\frac{2}{\pi\sqrt{3}}\frac{\gamma}{(1+\gamma)2^{-SE}-1}}$ EE decreases exponentially with SE.

In order to know whether a receiver architecture is better or worse than another, we must look at the SE and EE around the transition point in regime 2. For the sake of a simple discussion let us consider $P_o \ll N_r P_a$ and focus on the term $N_r P_a$. We can calculate a point in the transition regime $\tilde{SE}$ by solving the equation $P_a/B = 2c\sqrt{\frac{2}{\pi\sqrt{3}}\frac{\gamma}{(1+\gamma)2^{-\tilde{SE}}-1}}$, which may be rewritten as $\tilde{SE} = \log_2 \frac{\gamma+1}{\gamma+\frac{2}{\pi\sqrt{3}}\left[\frac{P_a}{2Bc}\right]^2}$. Ideally we would be interested in finding some optimal SE^* that maximizes EE in (2.35). However, note that if SE is much higher or lower than $\tilde{SE}$ the system would behave in regimes 1 and 3, respectively, whereas SE^* belongs to regime 2 by definition. Therefore, we can obtain some general insight about SE^* through its similarity to $\tilde{SE}$.

Let us compare two receivers with SNRs $\gamma^{(1)}$ and $\gamma^{(2)}$ and analog RF chain power consumptions $P_a^{(1)}$ and $P_a^{(2)}$. Let us define a difference in parameters $\Delta P_a^2 = (P_a^{(2)})^2 - (P_a^{(1)})^2$, and define the required SNR difference $\Delta\gamma^* = \gamma^{(2)} - \gamma^{(1)}$ such that receiver #2 achieves a greater or equal value in $\tilde{SE}$ than receiver #1.

$$\Delta\gamma^* = \frac{(\gamma^{(1)}+1)((P_a^{(2)})^2 - (P_a^{(1)})^2)}{(P_a^{(1)})^2 - \frac{\pi\sqrt{3}}{2}(Bc)^2} \tag{2.36}$$

where we observe that if Walden's figure of merit c is low, a higher SNR is required to achieve the same SE in a receiver with the optimal EE. However, if c is high, $\Delta\gamma^*$ changes sign and is negative. This means that the value $\tilde{SE} \simeq SE^*$ to maximize EE decreases, and that fewer bits are used in the optimal ADC configuration. But since the use of fewer bits increases the quantization noise, such a system can operate with a lower unquantized SNR without losing performance. In summary, switching from DC to HC can produce a significant gain or a significant loss depending critically on the component parameters and preference of EE over SE. This calls for an analytic framework that optimizes both EE and SE at the same time.

For each architecture, we plot a curve in a chart representing the evolution of its SE versus EE performance as the number of ADC bits b increases from 1 to 8, at increments of 1. The highest points in the chart correspond to highest SE and the rightmost points in the chart correspond to highest EE. Generally speaking, the closer a point to the top-right corner, the better. However, this guideline is not precise enough to fully describe the needs of receiver designers. Thus, we construct a multi-objective utility optimization interpretation of the charts that allows to describe all the points of interest in the graph quantitatively. For this we consider a free parameter $\alpha \in [0, 1]$ that represents the receiver designer's preference between higher EE and higher SE. The "receiver utility" according to the designer's preference can be expressed and maximized as

$$U = \max_{\{HC,DC,AC\}} \max_{b \in \{1 \ldots 8\}} \alpha \mathrm{EE} + (1 - \alpha)\mathrm{SE} \tag{2.37}$$

Clearly, with $\alpha = 0$ the above problem maximizes SE, and with $\alpha = 1$ it maximizes EE. Moreover, the above problem is easy to solve due to the small size of the exploration set, so we can obtain the solution for all values in the range $\alpha \in [0, 1]$. The set of solutions obtained for all values in the range of α correspond to all the receiver designs that designers with different preferences would choose. Hence, receivers not in the set of feasible solutions would never be of use. In the following SE-vs-EE charts, we have highlighted the points that are possible utility maximization solutions with a circle.

In this analysis, we observe the SE vs EE trade-off between AC, HC and DC for both the Downlink and the Uplink scenarios. We first compare the performance of different beamforming schemes both in the high and the low SNR regimes. We then investigate the dependence of the SE vs EE trade-off results on the component parameters (ADC and the phase shifter power consumption). Finally, we study how the behavior of HC varies with a change in N_{RF}.

2.3.4.1 Uplink-Downlink differences at high SNR

We now investigate the SE vs EE trade-off for AC, DC and HC schemes considering both the Uplink and the Downlink scenarios. For both scenarios, we set $(N_c, N_p) = (Poisson(1.8), 20)$, and $N_{RF} = 4$. For the Downlink, we set $N_t = 64$ and $N_r = 16$, whereas for the Uplink, $N_t = 16$ and $N_r = 64$.

Figures 2.15 and 2.16 show the SE vs EE trade-off in the high SNR regime for the Downlink and the Uplink scenarios, respectively. In the charts we observe that as b is increased, the curves first reach upward and right, and then wrap around and EE returns to the left corner while SE continues increasing. This is consistent with the analysis of (2.35).

Each curve can be interpreted as an "achievable region" where receivers of each type can operate. The larger the area covered by a curve, the more versatile the corresponding receiver

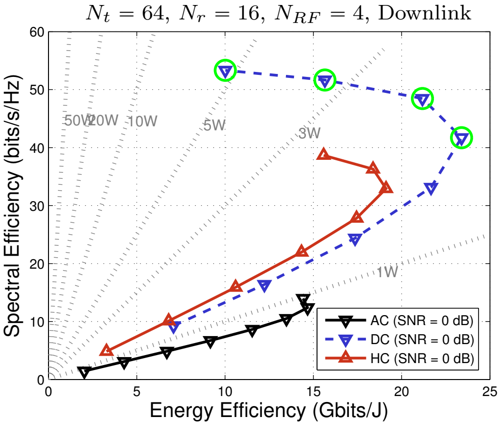

Fig. 2.15. *SE vs. EE comparison in high SNR regime for AC, DC and HC schemes for a Downlink scenario with a HPADC model, and with* $(N_c, N_p) = (Poisson(1.8), 20)$.

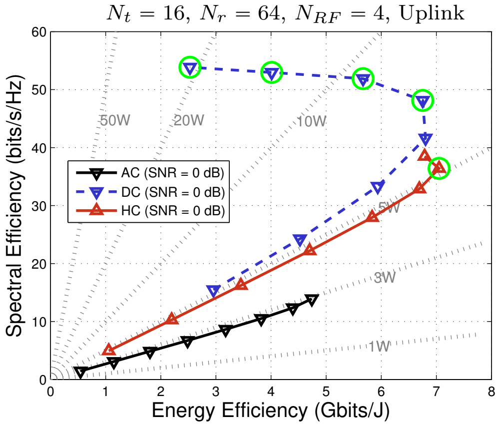

Fig. 2.16. *SE vs. EE comparison in high SNR regime for AC, DC and HC schemes for an Uplink scenario with a HPADC model,and with* $(N_c, N_p) = (Poisson(1.8), 20)$.

design. Moreover, the points maximizing (2.37) for different values of α are highlighted.

Note that the number of bits that maximize SE may not maximize EE. For instance, in Figure 2.15 the circled point in the left top corner maximizes SE ($\alpha = 0$) corresponds to $b = 8$, whereas the circled point in the right top corner maximizes EE ($\alpha = 1$) corresponds to $b = 5$. Notice also that the number of bits which maximizes for different combining schemes may not be the same, i.e., AC, HC, and DC achieve maximum EE with $b = 7, 6$ and 5, respectively.

In the charts we observe that, for Downlink, DC dominates other schemes. Note that DC with 4 or 5 bits consumes lower power than the HC scheme (for any number of ADC bits) while achieving better EE than AC or HC and also achieving higher SE.

Note that in Downlink the receiver is equipped with only $N_r = 16$ antennas. The power consumption increases in Uplink where the receiver has $N_r = 64$ antennas. This increase affects all schemes but is more severe for DC. Uplink HC with 6-8 bits achieves better (or equivalent) EE than DC with any number of bits, but only HC with 7 bits is actually selected for its utility, whereas DC is used not only for the highest SE ($b = 8$), but also at many points with a good EE ($b = 5, 6$) that meet the utility-maximization criteria depending on the designer preference parameter $\alpha \in [0, 1]$.

Finally, a constant device power consumption reference grid has been added to the chart (diagonal dotted lines). For a given power constraint, only points below the corresponding rule in the grid may be selected. In Downlink, only AC is a viable receiver below 1 W,

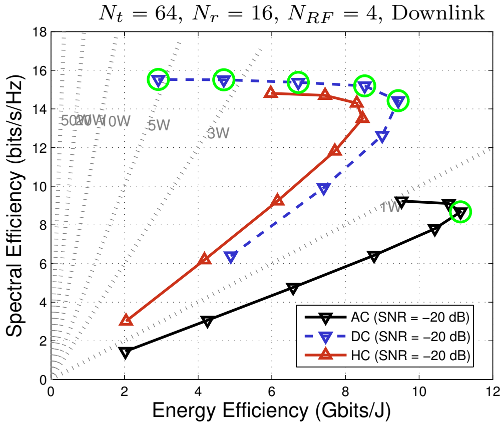

Fig. 2.17. *SE vs. EE comparison in low SNR regime for AC, DC and HC schemes for a Downlink scenario with a HPADC model, and with* $(N_c, N_p) = (Poisson(1.8), 20)$.

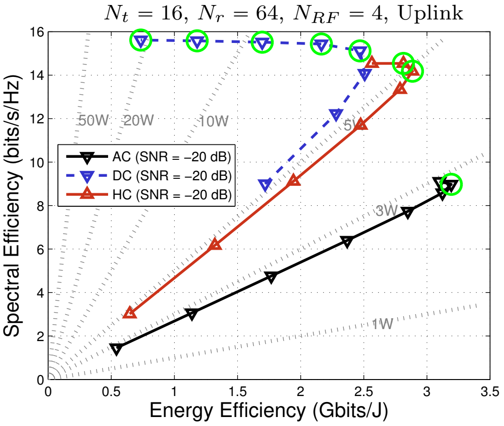

Fig. 2.18. *SE vs. EE comparison in low SNR regime for AC, DC and HC schemes for an Uplink scenario with a HPADC model, and with* $(N_c, N_p) = (Poisson(1.8), 20)$.

whereas above 1 W DC outperforms the other schemes. In Uplink, only AC is viable below 3 W. In the range $3 - 5$ W, HC presents a better choice. Finally above 5 W, DC outperforms HC again. Thus, surprisingly, DC is a better receiver for both smaller devices (such as a UE with 1-3 W) and larger devices (such as a macro cell BS with 6-50W), while HC is better for mid-range power devices (such as a pico cell BS with $3 - 5$W). In devices below 1 W AC should be used. Note that these observations apply only to receivers fabricated with the component consumption values listed in Tables 2.2 and 2.3. Engineers should generate charts for their own components following the example in this paragraph and using the tool available at [49].

Note that we consider high SNR where HC and DC schemes exploit spatial multiplexing gains and are significantly better than AC. Moreover, we consider the HPADC scenario with existing ADC and PS device models. Finally, we only consider $N_{RF} = 4$ in the HC scheme, where an improved EE can be achieved with fewer RF chains, at the cost of SE, and since the channel can support higher spatial dimensions, up to $\min(N_t, N_r) = 16$, the SE can also be augmented by increasing N_{RF}. We investigate the impact of all these parameters in the following sections.

2.3.4.2 Uplink-Downlink differences at Low SNR

The performance trade-off of AC, HC, and DC schemes in low SNR regime for the Downlink and the Uplink is shown in Figures 2.17 and 2.18, respectively.

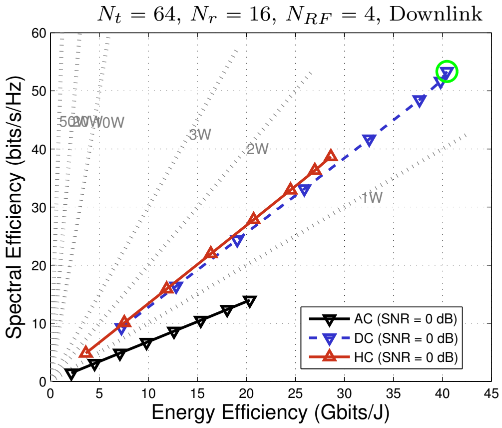

Fig. 2.19. *SE vs. EE comparison for AC, DC and HC schemes for a Downlink scenario with a LPADC model with* $(N_c, N_p) = (Poisson(1.8), 20)$ *and* $N_{RF} = 4$ *for HC.*

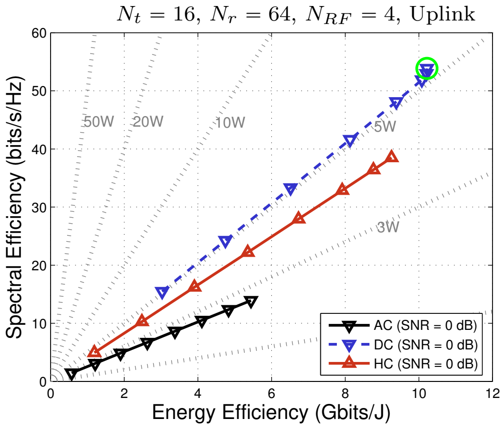

Fig. 2.20. *SE vs. EE comparison for AC, DC and HC schemes for an Uplink scenario with a LPADC model with* $(N_c, N_p) = (Poisson(1.8), 20)$ *and* $N_{RF} = 4$ *for HC.*

We note that at low SNR (and both in the Downlink and Uplink scenario) water-filling concentrates all the transmitted power in one single strongest dimension (singular value) of the channel. Thus, low-SNR links with water-filling become effectively rank-1 channels, which were shown in Subsection 2.3.2.4 to achieve very similar SE with AC, HC and DC for a sufficient number of bits. HC and DC cannot exploit any multiplexing gain and have a fairly small difference in SE compared to AC, so AC with its simpler hardware displays much better energy efficiency and covers a wider area towards the right.

The relation between HC and DC follows a similar trend as what was observed in the high SNR case, with the addition of a point of the AC architecture being added to the set of solutions to the multi-objective utility maximization for the maximal values of EE.

2.3.4.3 Improvement of ADC and Phase Shifter Power Model

Now we study the influence of the component characteristics, and particularly ADC and Phase Shifter power consumption, on the EE vs SE trade-off. We improve the HPADC model employed in Subsection 2.3.4.1, where we considered state of the art devices with parameters $c = 494$ fJ and $P_{PS} = 2$ mW. Now, we introduce a “reasonable best-case future evolution” set of values considering that the Walden’s figure of merit of the ADC can improve up to 5 fJ [56] and the Phase Shifter power consumption can be reduced almost to zero, $P_{PS} \sim 0$ [55].

Figures 2.19 and 2.20 show the SE vs EE plot with the improved component characteristics for the Downlink and the Uplink scenarios, respectively. The charts show that, in general, EE is improved after switching to better ADC and PS parameters. In all schemes, the ADCs power consumption is so low that, in the considered range, the lines do not twist at some optimal number of bits, after which EE starts decreasing. Instead, for $b \leqslant 8$, EE always increases with b and the lines are nearly straight. This improvement of EE is better exploited by DC, which dominates HC completely in Downlink, whereas in Uplink achieves significantly greater SE with very slightly worse EE than HC. Still, HC is selected if maximal EE is desired.

If we compare schemes point-to-point at the same number of bits, DC combining always performs better than HC for the same value of b. However, if we make a comparison based on a fixed SE value, for instance SE = 40 bps/Hz, then HC with 8 bits would be a better choice than DC with 4 bits in Uplink scenarios.

Note that in the utility maximization combining SE and EE, again DC is superior in the Downlink while both HC and DC are possible choices in the Uplink scenario on similar terms as in previous sections. However, all points of interest are so close to one another that the SE-vs-EE trade-off is rendered nearly irrelevant by the ADC technology improvement. Thus, for the rest of our analysis below we only focus on the HPADC component model.

2.3.4.4 RF chains comparison

We now observe the effect on the SE vs EE trade-off for HC as the number of RF chains is changed. Increasing N_{RF} increases SE but also the power consumption, potentially leading to a drop in EE. We compare the performance for both the Downlink and the Uplink while considering both the high and the low SNR regime.

Figures 2.21 and 2.22 show the SE vs EE trade-off for the Downlink in the high and in the low SNR regimes, respectively. At high SNR, an improvement both in SE and in EE is observed by increasing N_{RF} from 2 to 4, i.e., EE is improved from 15 Gbits/J to 19 Gbits/J whereas SE is improved from 22 bits/s/Hz to 32 bits/s/Hz. This is due to the fact that $N_{RF} = 2$ is not sufficient to fully exploit all spatial multiplexing gains available in the channel. Thus, SE increases much more than the power consumption, and as a result EE improves. However, a further increase in N_{RF} only improves SE slightly and EE starts decreasing because a spatial multiplexing gain greater than 4 is much less likely in the random channel with $(N_c, N_p) = (Poisson(1.8), 20)$. Also note that DC is indifferent to the number of RF chains and provides all the best SE and EE multi-objective utility maximization solutions

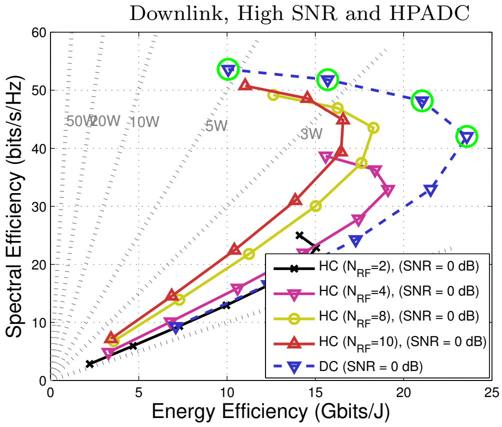

Fig. 2.21. *SE vs. EE comparison in high SNR regime for AC, DC and HC schemes for a Downlink scenario with a HPADC model with* $(N_c, N_p) = (Poisson(1.8), 20)$ *and* $N_{RF} = 2, 4, 8, 10$ *for HC.*

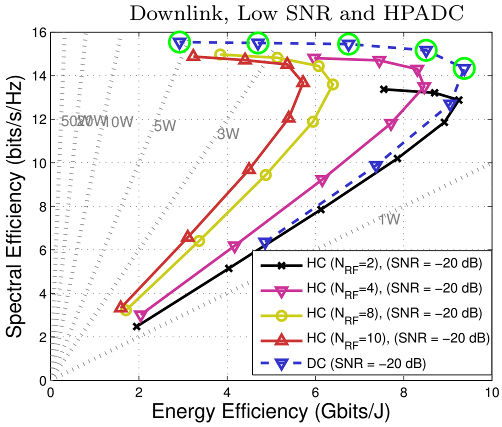

Fig. 2.22. *SE vs. EE comparison in low SNR regime for AC, DC and HC schemes for a Downlink scenario with a HPADC model with* $(N_c, N_p) = (Poisson(1.8), 20)$ *and* $N_{RF} = 2, 4, 8, 10$ *for HC.*

ranging from 5 to 8 bits.

At low SNR, there is no significant spatial multiplexing advantage, and therefore HC seems to perform better with $N_{RF} = 2$. Thus, the optimal number of RF chains is SNR dependent, a troublesome handicap for HC since usually the same hardware is built in all devices and has to be able to operate at different distances in a network. The only scenario where HC achieves EE similar to DC is with $N_{RF} = 2$ and under low SNR, whereas at high SNR the same scheme has a 35% lower EE with HC than with DC. Therefore, another advantage of DC is the ability to adapt digital signal processing to varying channel conditions, which suggests that DC may be a preferable option in the Downlink.

Figures 2.23 and 2.24 show the SE vs EE trade-off for the Uplink in the high and in the low SNR regimes, respectively. Here we considered $N_{RF} = 2, 4, 8$ and 12. The curves for the Uplink follow a similar trend as what was observed for the Downlink scenario, i.e., there is an optimal number of RF chains which maximizes the SE vs EE trade-off and a further increase in N_{RF} decreases EE while SE remains nearly the same. At high SNR, $N_{RF} = 8$ achieves the best SE vs EE trade-off whereas at low SNR $N_{RF} = 4$ performs best. Note that in the Uplink with either $N_{RF} = 4$ or 8, HC with the optimal N_{RF} achieves better EE than DC in both low and high SNR regimes and the multi-objective utility maximization does switch between DC (when greater SE is preferred) and HC (when greater EE is preferred). Again, the problem of SNR dependence in the selection of the number of RF chains negatively affects

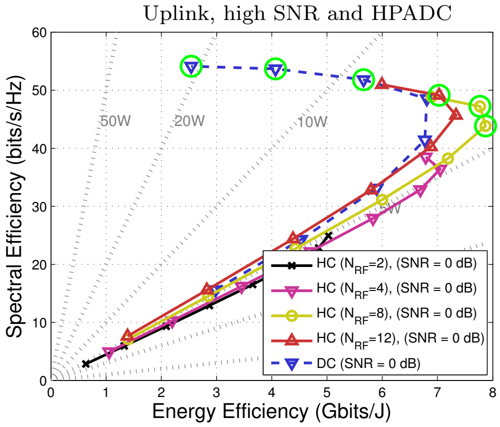

Fig. 2.23. *SE vs. EE comparison in high SNR regime for AC, DC and HC schemes for an Uplink scenario with a HPADC model with* $(N_c, N_p) = (Poisson(1.8), 20)$ *and* $N_{RF} = 2, 4, 8, 12$ *for HC.*

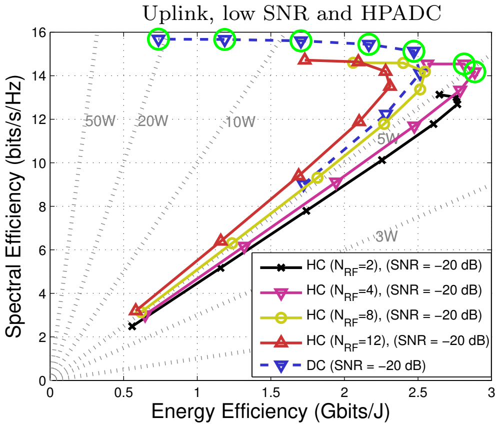

Fig. 2.24. *SE vs. EE comparison in low SNR regime for AC, DC and HC schemes for an Uplink scenario with a HPADC model with* $(N_c, N_p) = (Poisson(1.8), 20)$ *and* $N_{RF} = 2, 4, 8, 12$ *for HC.*

the versatility of HC.

2.4 Conclusions

In this chapter, we studied and analyzed the performance of popular millimeter wave receiver architectures based on analog, digital and hybrid combining schemes. We compared these combining architectures focusing on low resolution ADCs and taking into account the power consumption of all receiver components. Results showed that the common perception regarding high power and high energy consumption of DC is not always justified. Rather, there are many use-cases where DC results in a better power, spectral and energy efficiency than other combining schemes. Particularly, DC with fewer ADC bits than HC shows a better spectral efficiency vs energy efficiency trade-off. However, a comparison with the same high number of ADC bits results in favor of HC. Moreover, AC is only favorable for single dominant path channels or devices with stringent power constraints. We also showed that there is an optimal number of bits that maximizes the energy efficiency.

3

Bit Allocation in Variable Resolution ADCs

In future high-capacity wireless systems based on mmWave or massive multiple input multiple output (MIMO), the power consumption of receiver Analog to Digital Converters (ADCs) is a concern. Although hybrid or analog systems with fewer ADCs have been proposed, fully digital receivers with many lower resolution ADCs (and lower power) may be a more versatile solution. In this chapter, focusing on an uplink scenario, we propose to take the optimization of ADC resolution one step further by enabling variable resolutions in the ADCs that sample the signal received at each antenna. This allows to give more bits to the antennas that capture the strongest incoming signal and fewer bits to the antennas that capture little signal energy and mostly noise. Simulation results show that, depending on the unquantized link SNR, a power saving in the order of 20-80% can be obtained by our variable resolution proposal in comparison with a reference fully digital receiver with a fixed low number of bits in all its ADCs.

3.1 Introduction

Future wireless communications are expected to leverage large antenna arrays at the base station to achieve higher data rates, both in new mmWave bands and at standard frequencies with massive MIMO [14, 15]. Fully digital receiver architectures, where each antenna is connected to an independent Analog to Digital Converter, can provide maximum flexibility but could display too high component power consumption due to the exponential increase of ADC power with the number of bits [17]. Note that the concept of green communication and the deployment of ultra-dense small cells motivate the reduction of power consumption at the base station.

There are two strategies to mitigate the power consumption of receivers with many antennas:

1. Use Analog or Hybrid Combining (AC or HC) to perform all or a part of the MIMO operations in analog circuitry and sample only one or a few signals with ADCs [18,19].
2. Use fully Digital Combining (DC) with reduced ADC resolution (for example, 1 or a few bits), which can offer even better power efficiency if the power of the radio-frequency (RF) components is taken into account [20,39,58].

In this work, focusing on an uplink scenario, we propose a further improvement to the fully-digital low-resolution strategy by studying the possibility of enabling a variable number of bits in each ADC of the DC system. Compared to a conventional approach to low-resolution DC, where each RF chain has equal ADCs with the same fixed number of bits b_{ref}, we propose to assign some ADCs a slighly higher number of bits $b_{\text{high}} > b_{\text{ref}}$, while the rest of the RF chains have an even lower number of bits ($b_{\text{low}} < b_{\text{ref}}$). Our results show that the same capacity of the fixed-bit system can be achieved using two variable-bit values with a power saving between 20 and 80%, depending on the link pre-quantization Signal to Noise Ratio (SNR).

3.1.1 Related Work

Recent works such as [29–31] study the capacity and energy efficiency (EE) of large antenna array receiver designs depending on the ADC resolution. The effect of the number of ADC bits b and sampling rate B on capacity and power consumption is analyzed in [34] for both AC and DC.

DC systems using low-resolution ADCs to reduce power consumption are further analyzed in [38,39], showing that a few bits are enough to achieve almost the spectral efficiency (SE) of an unquantized system of the same characteristics.

It is possible to use analog switches instead of analog mixers to create a hybrid scheme that samples only the best subset of the antennas of the array to reduce power consumption, as proposed in [37]. In addition to switching the best antennas to high resolution ADCs, it is possible to add 1-bit ADCs to sample the rest of the antennas as in [59], achieving a large fraction of the capacity of a full-high-resolution architecture.

In this work we show that these configurations with antenna selection (equivalent to 0 bit ADCs) or only 1 bit, combined with a few very high resolution ADCs, are not necessarily optimal, and a milder variation in the number of bits such as $b_{\text{low}} = 4$ vs $b_{\text{high}} = 6$ works better.

3.1.2 Our Contribution

In this work, we focus on low resolution ADCs and discuss how the availability of variable resolution ADCs can reduce the receiver power consumption compared to a receiver with fixed resolution on each ADC. We primarily focus on a simple case where the ADCs offer only two operation modes, with low and high resolutions. We believe such simplified model is a good starting point to open this topic and can be more easily imagined as a practically feasible hardware.

We propose two different algorithms depending on whether ADCs can only take two resolution values, or can also be completely shut off to also incorporate the benefits of antenna selection. We perform an analysis of power consumption with a constraint that the varibable-resolution scheme achieves the same SNR after quantization and capacity compared to the reference fixed-resolution model. We show that

- A receiver with two-level ADC resolution can achieve similar capacity to a fixed resolution scheme such as [38, 39] in practical systems with high SNR. However, instead of 2 bits like in [38] our reference is the fixed resolution that achieves the best trade-off between SE and EE under our channel model, which is $b_{\text{ref}} = 5$ according to results in Chapter 2[1].

- The power saving is related to the two levels of resolution selected, b_{low} and b_{high}. Robust performance at low SNR is obtained for not-too-low b_{low} and not-too-high b_{high}, whereas at high SNR more power is saved with more extreme differences between b_{low} and b_{high}.

- The power saving increases slightly with a larger number of antennas, and so enabling variable resolution is even more interesting in “massive MIMO” systems.

---

[1] This may be visualized using the tool available at `http://enigma.det.uvigo.es/~fgomez/mmWaveADCwebviewer/`, by selecting the HPADC (High-Power ADC) option and the values of the parameters discussed later in this work.

3.2 System Model

3.2.1 mmWave Channel

We study mmWave point-to-point uplink MIMO links with an N_t antenna transmitter, an N_r antenna receiver and bandwidth B. We assume that there is no inter-symbol interference, as in previous models such as [60]. The received signal in each symbol period $1/B$ is

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3.1}$$

where $\mathbf{x}$ represents the transmitted symbol vector, $\mathbf{n}$ is the independent and identically distributed (i.i.d) circularly symmetric complex Gaussian noise vector, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, N_o\mathbf{I})$, where N_o represents the noise power, and $\mathbf{H}$ represents the $N_r \times N_t$ channel matrix. The mmWave channel matrix $\mathbf{H}$ is randomly distributed following a random geometry with a small number of propagation paths (order of tens) grouped in very few clusters of similar paths [60], and is obtained as

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{\rho N_c N_p}} \sum_{k=1}^{N_c} \sum_{\ell=1}^{N_p} g_{k,\ell} \mathbf{a}_r(\phi_k + \Delta\phi_{k,\ell}) \mathbf{a}_t^H(\theta_k + \Delta\theta_{k,\ell}) \tag{3.2}$$

where the terms in this expression are generated according to the mmWave channel model in Chapter 2. Here ρ is the pathloss, $g_{k,\ell}$ is the small scale fading coefficient associated with the ℓ^{th} path of the k^{th} cluster, $\mathbf{a}_t$ and $\mathbf{a}_r$ are spatial signatures of the transmit and receive arrays, and the θ's and $\Delta\theta$'s and ϕ's and $\Delta\phi$'s are the angles of departure and arrival for a small number of propagation paths N_p grouped in even fewer non-independent clusters N_c.

It must be noted that, due to this small number of paths, despite having large dimensions, the matrix $\mathbf{H}$ has a low rank and an even lower number of dominant eigenvalues are responsible for 95% of the energy transfer in the channel. $\mathbf{H}$ is generated in [60] using $N_c \sim$ Poisson(1.8) and $N_p = 20$. In this chapter, we generate $\mathbf{H}$ instead with $N_c = 2$, $N_p = 10$ for absolute compatibility with fixed-resolution power consumption values obtained in Chapter 2, where N_c is selected as a constant and varied to study its effect. Moreover, it is noted in [60] that for the median channel a single spatial dimension captures approximately 50% of the channel energy and two degrees of freedom capture 80% of the channel energy. We also performed our own Monte-Carlo verification with 10^4 channel realizations, and found

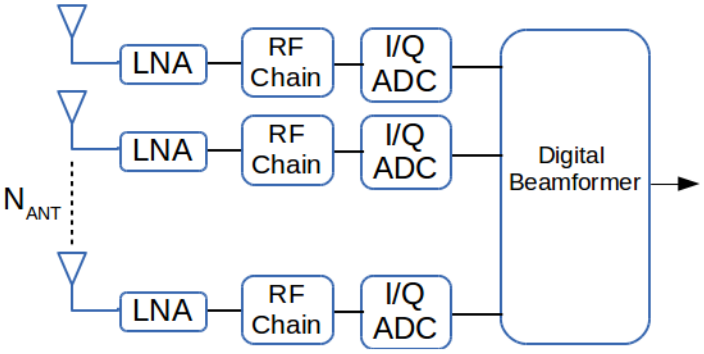

Fig. 3.1. *Digital Receiver with ADC with b_i bits on each antenna RF chain.*

that the first eigenvalue is responsible for over 50% energy transfer with probability 0.95 and for over 75% of the energy transfer with probability 0.6.

Therefore, for the sake of space and simplicity, in our analysis we assume that the transmitter has the channel state information (CSIT) and implements a beamforming scheme that concentrates all the signal in the single strongest eigenvalue of the channel matrix. That is, if $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ is the Singular Value Decomposition of the channel, the transmitter sends a scalar symbol, x, projected over the row $\mathbf{v}_m$ of $\mathbf{V}^H$ associated with the strongest eigenvalue on the diagonal $\mathbf{\Sigma}$. Thus the signal at the transmitter array is $\mathbf{x} = \mathbf{v}_m x$ and the received signal may be expressed as

$$\mathbf{y} = \mathbf{u}_m \sigma_m x + \mathbf{n} \tag{3.3}$$

where σ_m is the maximum singular value, and $\mathbf{u_m}$ is the corresponding left singular vector.

3.2.2 Variable-bits ADC Receiver

The DC receiver is illustrated in Fig. 3.1. After the signal (3.1) is received, the signal at each antenna i is quantized by an ADC with b_i bits. Due to the fact that we are only concerned about power consumption in this work, but not the maximum number of components, we illustrate the variable resolution ADC with the simplified architecture in Fig. 3.2, consisting in a pair of fixed-resolution ADCs that can be alternatively switched in and out of the circuit. This simplified architecture serves to illustrate some interesting gains and open the discussion on variable resolution ADCs, while we leave the design of resource-efficient variable-resolution ADC architectures for future research. Moreover, the use of switching hardware guarantees that ADCs can be commuted with the same time resolutions used in

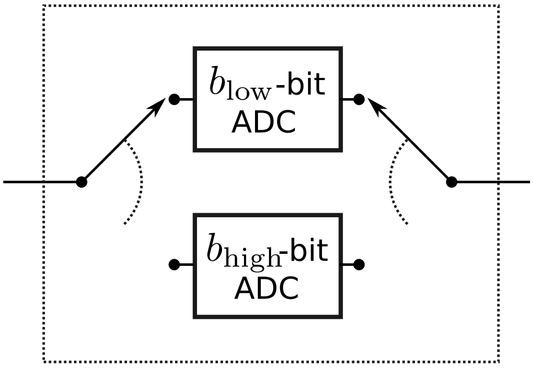

Fig. 3.2. *A simplified 2-level variable-resolution ADC.*

antenna selection schemes, allowing us to disregard potential issues with power-up time of ADCs. Note that there are very few works regarding the design of variable-bit ADC, for instance [61], however, we believe that the improvement shown with variable resolution ADCs for large antenna arrays will motivates ADC researcher to develop power efficient variable resolution ADCs.

We represent the signal after quantization using the Additive Quantization Noise Model (AQNM) [34] approximation by adding an additive white noise $\mathbf{n}_q$ that models the quantization distortion of each coefficient y_i of the signal (3.1), producing a quantized output in each ADC that satisfies

$$y_i^q = (1 - \eta_i) y_i + n_i^q \tag{3.4}$$

where η_i is the inverse of the signal-to-quantization noise ratio at antenna i, and is inversely proportional to the square of the resolution of the i-th ADC (i.e., $\eta_i \propto 2^{-2b_i}$). The quantization noise in each antenna n_i^q is AWGN distributed with variance $\eta_i(1 - \eta_i)\mathrm{E}\left[|y_i|^2\right]$.

We can write the quantized signal as a vector by denoting the quantization terms in a diagonal matrix, producing

$$\begin{aligned} \mathbf{y}^q &= \mathbf{D}(\mathbf{H}\mathbf{x} + \mathbf{n}) + \mathbf{n}^q, \\ \mathbf{D} &= \begin{pmatrix} (1-\eta_1) & 0\dots & 0 & \\ 0 & (1-\eta_2)\dots & 0 & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (1-\eta_{N_r}) \end{pmatrix}, \end{aligned} \tag{3.5}$$

where $\mathbf{D} = (1 - \eta_{\text{ref}})\mathbf{I}$ when the number of bits is the same in all ADCs.

For a Gaussian input distribution, the values of η for $b \leqslant 5$ are listed in Table 2.1, and for $b > 5$ can be approximated by $\eta = \frac{\pi\sqrt{3}}{2}2^{-2b}$ [38]. We denote by γ_q the SNR of $\mathbf{y}_q$, given by

$$\gamma_q = (\mathbf{D}\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H\mathbf{D}^H)(\mathbf{D}^2 N_o + \mathbf{R}_{\mathbf{n}^q\mathbf{n}^q})^{-1} \tag{3.6}$$

$$= (\mathbf{D}\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H\mathbf{D}^H)(\mathbf{D}^2 N_o + ((\mathbf{I} - \mathbf{D})(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H + N_0\mathbf{I})\mathbf{D}^H))^{-1} \tag{3.7}$$

where $\mathbf{R}_{\mathbf{n_q n_q}}$ and $\mathbf{R}_{\mathbf{xx}}$ represent the covariance matrix of the quantization noise and of the transmitted symbol, respectively. Note that (3.7) is valid for any type of transmission, and in this work we can simplify it for our particular choice of single dominant eigenvalue beamforming. We replace $\mathbf{R}_{\mathbf{xx}} = |x|^2\mathbf{v}_m\mathbf{v}_m^H$, and the optimal receiver for this transmission is Maximum Ratio Combining so the SNR becomes

$$\gamma_q = \sum_i \frac{\sigma_m^2|u_m^i|^2(1-\eta_i)^2}{N_0(1-\eta_i)^2 + (N_0 + \sigma_m^2|u_m^i|^2)\eta_i(1-\eta_i)} \tag{3.8}$$

where u_m^i is the ith element of the dominant eigen vector. By imposing the constraint of using single dominant eigenvalue beamforming, we are able to express the SNR as a sum of per-RF-chain partial SNR's as $\gamma_q = \sum_i \gamma_q^i$, where the quantized signal from each antenna has partial SNR $\gamma_q^i = \frac{\sigma_m^2|u_m^i|^2(1-\eta_i)^2}{N_0(1-\eta_i)^2+(N_0+\sigma_m^2|u_m^i|^2)\eta_i(1-\eta_i)}$. Also we can define the unquantized SNR per antenna $\gamma_i = \frac{|u_m^i|^2\sigma_m^2}{N_0}$ such that $\gamma_q^i = \frac{(1-\eta_i)\gamma_i}{1+\eta_i\gamma_i}$.

Finally, the capacity of the MIMO link, which in general is

$$C_q = \mathrm{E}_{\mathbf{H}}\left[\max_{\mathbf{R}_{\mathbf{xx}}} B\log_2\left|\mathbf{I} + (1-\eta)(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H)(N_o\mathbf{I} + \eta(\mathbf{H}\mathbf{R}_{\mathbf{xx}}\mathbf{H}^H))^{-1}\right|\right], \tag{3.9}$$

can be particularized with our transmitter constraints to

$$C_q = \mathrm{E}_{\mathbf{H}}\left[\log(1+\gamma_q)\right] \tag{3.10}$$

where E [.] represents the expectation.

3.2.3 Digital Receiver Power Consumption

The devices required to implement the mmW receiver architecture are displayed in Fig. 3.1. All receiver schemes considered in this work have the same RF components and we are only interested in the variation of ADC power consumption as a function of their number of bits.

The power consumption of the i-th ADC, denoted as $P^i_{ADC} = cB2^{b_i}$, increases exponentially with the number of bits b_i and linearly with the bandwidth B and with the ADC Walden's figure of merit c [43] (the energy consumption per conversion step per Hz). The aggregate power consumption across all the ADCs in the system is

$$P^{Tot}_{ADC} = \sum_i P^i_{ADC} = cB\left(\sum_i 2^{b_i}\right) \tag{3.11}$$

where it must be noted that we consider that all ADCs have the same Walden's figure of merit despite the variation of bits.

3.3 Analysis of the Variable-Bit ADC System

We compare a reference receiver with a fixed number of bits in all ADCs, b_{ref}, and a receiver where the number of bits on each ADC can be selected at any time instant between two possible values, b_{low} and b_{high}, depending on the instantaneous unquantized SNR's in each of the antennas γ_i.

With regard to the capacity of the channel, we must note that as b_i grows, η_i decreases exponentially to 0. Each term in (3.8) of the form $\gamma^i_q = \frac{\sigma^2_m |u^i_m|^2 (1-\eta_i)^2}{N_0 + \sigma^2_m |u^i_m|^2 \eta_i (1-\eta_i)}$ increases monotonically as $\eta_i \to 0$. Thus, the per-antenna SNRs always increase with b_i. Moreover, the higher $|u^i_m|^2$, the more the increase in SNR derived from assigning more bits to the i-th antenna.

With regard to the power consumption, denoting the number of antennas with b_{high} bits by N_{high}, we write the normalized power consumption (ratio between the power consumption with variable and fixed resolution) as

$$\begin{aligned} \xi &= \frac{cB(N_{\text{high}} 2^{b_{\text{high}}} + (N_r - N_{\text{high}}) 2^{b_{\text{low}}})}{cB N_r 2^{b_{\text{ref}}}} \\ &= \frac{N_{\text{high}}}{N_r} 2^{b_{\text{high}} - b_{\text{ref}}} + \left(1 - \frac{N_{\text{high}}}{N_r}\right) 2^{b_{\text{low}} - b_{\text{ref}}} \end{aligned} \tag{3.12}$$

The design of our variable-bit ADC system must satisfy that

- In order to be able to replicate the capacity and/or power consumption of the reference system, we select bit values

$$b_{\text{low}} \leqslant b_{\text{ref}} \leqslant b_{\text{high}}.$$

- If we wish for the variable-bit system to consume less, or equal power as the reference system, we must have
$$N_{\text{high}} \leqslant \frac{2^{b_{\text{ref}}-b_{\text{low}}}-1}{2^{b_{\text{high}}-b_{\text{low}}}-1} N_r \tag{3.13}$$
- For any antenna i, the increase in SNR obtained by increasing its number of bits to $b'_i > b_i$ is independent of the state of the other antennas.
- If any pair of antennas i, j satisfies $|u_m^i| > |u_m^j|$ but $b_i < b_j$, then the system achieves higher effective SNR and the same power consumption if we swap b_i and b_j.

The above ideas inspire the Greedy Bit Allocation (GBA) algorithm that allocates bits to ADCs in descending order of the values of $|u_m^i|$, implicitly obtained from the order of γ_i. The GBA algorithm only allows antennas to use b_{low} or b_{high} bits, starts with all ADCs in the low assignment, and swaps to a higher number of bits one antenna at a time until the system has the same effective SNR of the reference. Two outcomes are possible: if at the end of the algorithm the number of high-resolution RF chains is $N_{\text{high}} < \frac{2^{b_{\text{ref}}-b_{\text{low}}}-1}{2^{b_{\text{high}}-b_{\text{low}}}-1} N_r$, power has been saved by GBA. Otherwise, GBA wastes more power than a reference scheme with fixed resolution b_{ref}.

Algorithm 2 Greedy Bit Allocation

```
  Reference effective SNR $\gamma^{\text{ref}}$
  Measure unquantized SNR on each antenna $\gamma^i$
  Order the RF chains as $\gamma^1 \geqslant \gamma^2 \geqslant \dots \gamma^{N_r}$
  Start assuming $N_{\text{high}} = 0$, $b_i = b_{\text{low}} \forall i$
while $\sum_{i=1}^{N_r} \gamma_q^i < \gamma_q^{\text{ref}}$ do
    $N_{\text{high}} = N_{\text{high}} + 1$
    $b_i = b_{\text{high}}$
end while
```

Two weaknesses of GBA are that it does not work for $b_{\text{low}} = b_{\text{ref}}$ and that it always requires to have all antennas active. We add these functionalities to the improved Greedy Antenna Selection and Bit Allocation (GASBA) algorithm.

In the GASBA algorithm, three values are allowed per ADC, 0, b_{low} or b_{high} bits. The algorithm combines the mechanics of antenna selection and bit allocation, enabling the receiver to exploit a combination of high-resolution RF chains, lower-resolution RF chains, and disabled RF chains. This algorithm can replicate all allocations possible with GBA and also perform some new allocations in the style of antenna selection.

Algorithm 3 Greedy Antenna Selection and Bit Allocation

```
  Reference effective SNR $\gamma^{\text{ref}}$
  Measure unquantized SNR on each antenna $\gamma^i$
  Order the RF chains as $\gamma^1 \geqslant \gamma^2 \geqslant \ldots \gamma^{N_r}$
  Start assuming $N_{\text{on}} = 0$, $N_{\text{high}} = 0$, $b_i = 0 \forall i$
while $\sum_{i=1}^{N_{\text{on}}} \gamma_q^i < \gamma_{\text{ref}}$ do
  $N_{\text{on}} = N_{\text{on}} + 1$
  if $N_{\text{high}} < \frac{2^{b_{\text{ref}} - b_{\text{low}}} - 1}{2^{b_{\text{high}} - b_{\text{low}}} - 1} N_r$ then
    $N_{\text{high}} = N_{\text{high}} + 1$
    $b_i = b_{\text{high}}$
  else
    $b_i = b_{\text{low}}$
  end if
end while
```

3.3.1 Initial SNR Measurement

We assume that channel state information of each receive antenna is perfectly known. In practice, an estimation of the SNR will be needed. The impact of imperfect initial SNR estimation is left as part of our future work. For example in [61], an ADC design is proposed where inputs with lower voltage use 6 bits and inputs with higher voltage use 4 bits. Such a design could be easily modified to operate in the opposite way, giving 6 bits to the signals with the higher voltage. Since the average thermal noise per antenna is the same, the SNR is directly proportional to the squared voltage in the antenna circuits in receivers without Automatic Gain Control (AGC), and is proportional to the ratio between the squared voltage and the amplifier gain in receivers with independent AGC for each antenna.

3.4 Numerical Simulations

In this section we present the power saving achieved by GBA and GASBA algorithms, studied by Monte Carlo simulation with results averaged over 1000 independent realizations. We discuss the performance for different combinations of $b_{\text{low}}, b_{\text{high}}$ for both algorithms. In the simulations, the transmitter is always equipped with 4 antennas whereas the number of receive antennas can be either 64 or 256. We consider a mmWave link with $B = 1$ GHz and vary the unquantized link SNR from -20 to 20 dB (except for Figures 3.3 and 3.4, where it

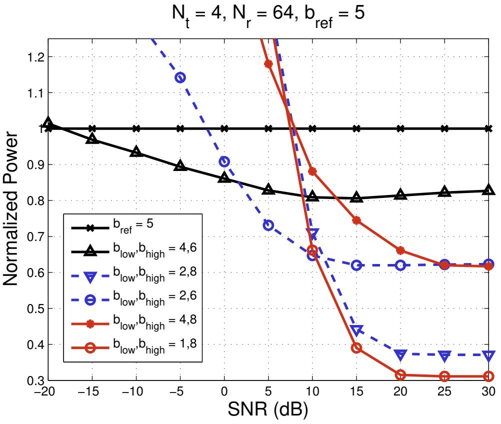

Fig. 3.3. *Greedy Bit Allocation with* $b_{\text{ref}} = 5$ *and* $N_r = 64$

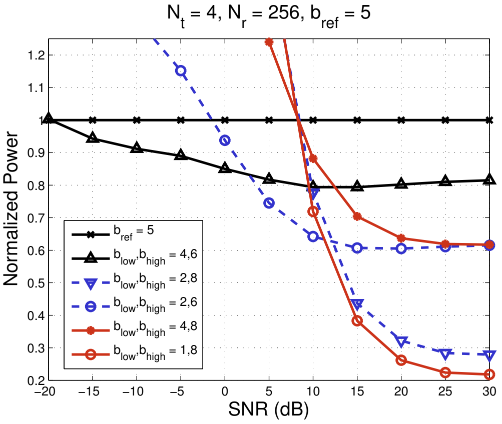

Fig. 3.4. *Greedy Bit Allocation with* $b_{\text{ref}} = 5$ *and* $N_r = 256$

is varied from -20 to 30 dB) with a step of 5 dB. The unquantized SNR is the product of transmitter power and pathloss, divided by the noise power N_0.

We display the normalized power consumption ξ (3.12) vs the unquantized SNR for systems that achieve the same quantized SNR (and thus, capacity). We use references that have 5 or 4 bits resolution, which achieve the best EE of a fixed-resolution system according to results shown in Chapter 2.

3.4.1 GBA

We begin with the results for the GBA algorithm. Figs. 3.3 and 3.4 show the normalized power consumed by the receivers with GBA for $N_r = 64$ and $N_r = 256$ receive antennas, respectively, and $b_{\text{ref}} = 5$. The result shows that above certain SNR values the variable resolution architecture displays lower power consumption than the fixed resolution architecture.

Note that the configurations where $b_{\text{low}}, b_{\text{high}}$ is relatively close to b_{ref} (e.g., $b_{\text{low}}, b_{\text{high}} = 4, 6$) result in a lower power consumption for almost the complete range of an unquantized SNR. This is due to the fact that, in circumstances where the quantization noise of the reference is smaller or comparable to the unquantized signal noise, using a high b_{low} (i.e., close to b_{ref}) already achieves a capacity close to the reference, and thus it only takes very few RF chains with b_{high} bits to close the gap.

Secondly, note that in the settings with very low b_{low} (i.e., with $b_{\text{low}} = 1, 2$ bits) a large power is saved at very high SNR, but the variable resolution system consumes even more power than the reference at low SNRs. This is due to the fact that if the number of bits is sufficient, a quantized system operates very close to the capacity of an unquantized system, but the threshold that marks this "sufficient" number of bits grows with the unquantized SNR (Section 2.2). Therefore, at high operating SNR the contribution of RF chains with b_{low} is not significant, and the use of smaller b_{low} saves power without harming the capacity much, while on the other hand the few ADCs with very high resolution (for instance $b_{\text{high}} = 8$) improve the SNR much more than many RF chains with moderate resolution. Therefore smaller b_{low} and higher b_{high} works better at high SNR.

Conversely, at low SNR, to achieve the capacity of the reference resolution system, a dramatic increase in the number of high resolution ADCs is required, which increases the power consumption of variables bits scheme even more than the reference fixed-bit architecture. For instance, with $(b_{\text{low}}, b_{\text{high}})$ set to $(1, 8)$, $(2, 8)$ and $(4, 8)$, the consumed power is lower than the reference only when the SNR is above 10 dB. This is because b_{high} produces no significant improvement in the SNR (i.e., "zooming in the noise") and choosing b_{high} closer to b_{ref} results in a lower power consumption.

The impact of the number of receive antennas can be observed by comparing Figs. 3.3 and 3.4. For instance, with $b_{\text{low}}, b_{\text{high}} = 4, 6$, the power of the variable resolution system compared to the fixed resolution reference is in the range 95-80% for both antenna configurations $N_r = 64$ and $N_r = 256$. On the other hand, at the same high SNR of 20 dB, the variable architecture with $b_{\text{low}}, b_{\text{high}} = 1, 8$ can achieve a power saving of 69% with 64 antennas, whereas the system with 256 antennas can increase the power saving to 75%. It seems that the number of antennas affects the gains more for more extremely separate resolution values that work well at high SNR, whereas it has a negligible impact for narrowly separated resolution values that work well at all SNRs.

Finally, note that there is a certain operating SNR after which the normalized power saturates. This is more obvious in Figure 3.3 where the plot is extended up to 30 dB. This is due to the fact that few fixed number of b_{high} ADCs are enough to get a capacity of a reference. Moreover, with $b_{\text{low}}, b_{\text{high}} = 4, 6$, a minimum power is achieved at 10 dB and 15 dB SNR for $N_r = 64$ and $N_r = 256$, respectively. This is due to the phenomena explained above that at very low (high) unquantized SNR the contribution of b_{high} (b_{low}) to the quantized SNR is not very significant.

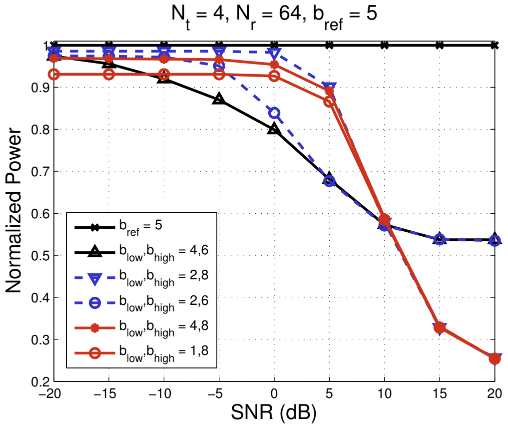

Fig. 3.5. *Greedy Antenna Selection and Bit Allocation with* $b_{\text{ref}} = 5$*,* $N_r = 64$

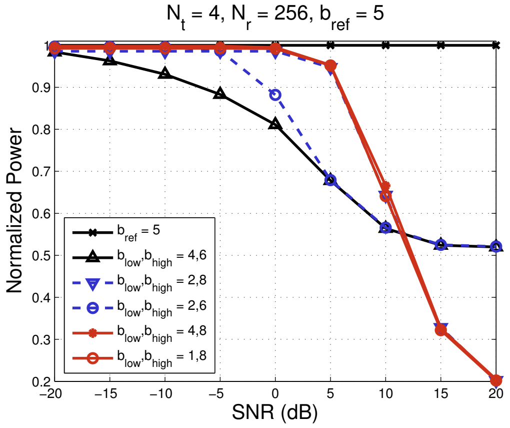

Fig. 3.6. *Greedy Antenna Selection and Bit Allocation with* $b_{\text{ref}} = 5$*,* $N_r = 256$

3.4.2 GASBA

To address the increased power consumption issues of GBA, we now discuss the performance of GASBA which also puts constraints on the total power consumed by the high-resolution antennas. Note that GBA always tries to achieve the rate of the reference, and sometimes this causes power consumption to be greater than the reference. However, in GASBA, the maximum N_{high} (3.13) ensures that the power consumption of the variable architecture is bounded by the reference. Therefore, after reaching the maximum power limit, GASBA algorithm stops assigning more higher resolution antennas to match the capacity of the reference fixed-bit architecture. GASBA algorithm keeps the power under control while trying to achieve a capacity as close to the reference as possible. Particularly, depending on the configuration of variable-bit, GASBA may result in a slight degradation and improvement in system capacity (compared to the fixed resolution architecture) at low and high SNR, respectively.

Figs. 3.5, 3.6 show the results for GASBA with $b_{\text{ref}} = 5$, $N_r = 64$ and $N_r = 256$, respectively. Note that the normalized power of any variable resolution configuration at any operating SNR is less than or equal to the power consumed by the reference architecture, thanks to the design of the algorithm to stop activating high-resolution RF chains when a certain limit is reached. At the lowest SNR values, the normalized power consumption of GASBA algorithms becomes flat as all antennas are assigned with ADCs with maximum N_{high} (3.13), and in these scenarios the GASBA algorithm gives up trying to make the variable-resolution system capacity match the reference. Fig. 3.8 illustrates this. Also note that, for

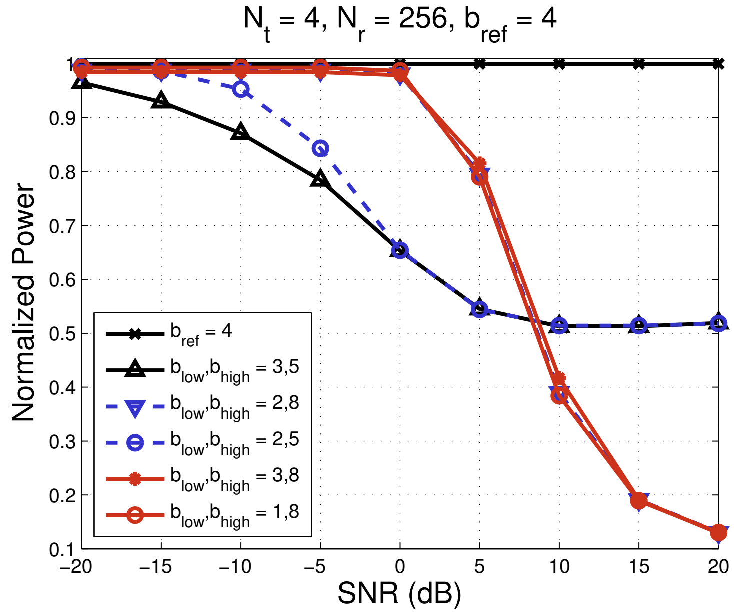

Fig. 3.7. *Greedy Antenna Selection and Bit Allocation with* $b_{\text{ref}} = 4$, $N_r = 256$

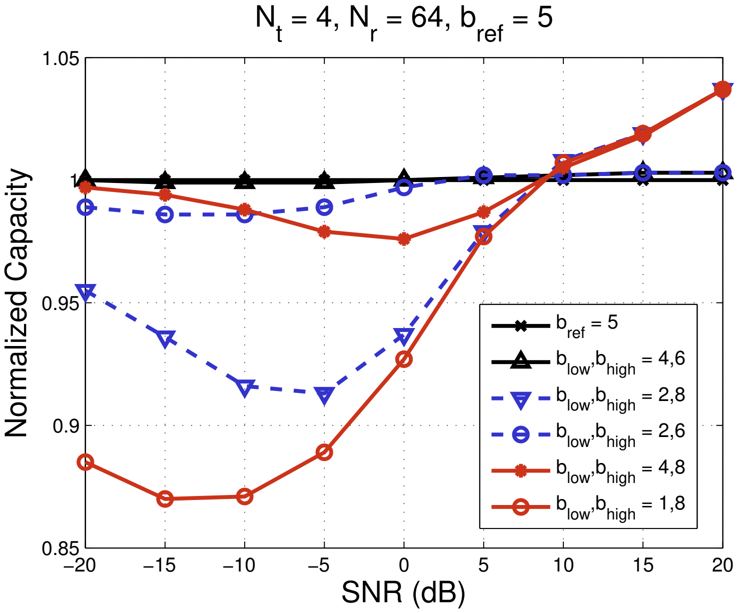

Fig. 3.8. *GASBA does not always match the reference capacity at low SNR.*

some configurations of $b_{\text{low}}, b_{\text{high}}$, the maximum value of N_{high} may not be an integer (Eq. 3.13), and therefore in those cases floor of N_{high} is selected due to which the curves stay below reference even when all antennas are utilized (Figure 3.5-3.7).

The power savings improve at high SNR. This is due both to the effects described for GBA above, and to the fact that a system with higher captured energy per antenna does not need so much beamforming gain and can make do with activating fewer RF chains. In fact, at higher SNRs using only the RF chains with b_{high} bit ADCs is enough to provide equal capacity as the reference, and the GASBA algorithm greatly reduces the subset of selected antennas at high SNR, keeping power consumption minimal.

The results show that at low SNRs the configurations with b_{low} and b_{high} closer to b_{ref} (i.e., $b_{\text{low}}, b_{\text{high}} = 4, 6$ or $2, 6$) are the only ones that can always achieve the same capacity as the reference. Moreover, these configurations also results in some power consumption savings at low SNR (e.g., at SNR = 0 dB). This follows from the fact that at low SNR too many bits do not provide significant improvement (i.e., "zooming in the noise"), so many antennas are activated regardless. A smaller b_{high} results in a higher upper limit to N_{high}, and the use of many moderately-high resolution antennas is preferable over few very high resolution ones.

Conversely, as SNR increases it is preferable to use few antennas with many bits. At higher SNRs, the systems with very high b_{high} start to display the best power savings. However, these values must not be too extreme for a given SNR. For example, at 10 dB the configuration $2, 6$ actually performs better than $2, 8$. Nonetheless, the curves for $b_{\text{high}} = 8$ display a more abrupt decrease. The value of b_{low} is irrelevant at high SNR because those RF chains are never turned on, differently from GBA where it had some small impact.

Note that in the entire range -20 to 20 dB SNR, which is quite reasonable for wireless communication, the variable resolution scheme with $b_{\text{low}}, b_{\text{high}} = 4, 6$ always results in a reasonable power reduction without any degradation in the achievable capacity for both GASBA and GBA algorithms. Thus, the variable-resolution model in [59] with $b_{\text{low}} = 1$, despite showing impressive gains at very high SNR, may not be the best configurations for all systems.

Note that although the GASBA algorithm always results in a lower power consumption than the reference fixed resolution architecture, there are regimes where GBA consumes lower power than GASBA for the same number of bits configuration in their ADCs. For instance, with 256 antennas the configuration $b_{\text{low}}, b_{\text{high}} = 4, 6$ at -10 dB SNR displays a normalized power of 0.90 (Fig. 3.4) and 0.95 (Fig. 3.6) for GBA and GASBA, respectively. This is because the GASBA algorithm starts by trying to approximate the capacity of the reference using only b_{high} ADCs and therefore in certain regimes GASBA may assign high-resolution ADCs to antennas where low-resolution ADCs could have done the same job (low-SNR antennas where the quantization noise is small or negligible compared to the thermal noise).

Finally, in Fig. 3.7 we observe the effect of changing the reference for a variable resolution GASBA system compared versus $b_{\text{ref}} = 4$. The results show that with a reduction in b_{ref} the difference in power consumption between the fixed and variable resolution techniques increases (compare Figs. 3.6 and 3.7). This is because a reduction in b_{ref} also reduces the target reference capacity and therefore the required number of ADCs with b_{high} bits decreases.

In summary, the use of variable resolution ADCs can provide a significant decrease in the power consumption in comparison to a fixed resolution ADC architecture, ranging from 20% at 0 dB SNR up to 80% at 20 dB.

3.5 Conclusions

In this chapter, we proposed and analyzed a variable-bit DC receiver architecture. The results showed that the proposed solution can provide 20-80% power reduction (depending on the SNR regime) compared to a fixed-bit receiver architecture. The performance is analyzed based on two greedy algorithm, 1) greedy bit allocation which saves power only at high SNR but always tries to achieve the capacity of the fixed-bit reference architecture, and 2) greedy antenna selection and bit allocation which always consumes a power less than or equal to that of the reference fixed-bit design but may result in a slight capacity degradation at low SNR.

4

Directional Cell Discovery

In this chapter, we study receiver efficiency from a control plane perspective. Particularly, we investigate the choice of an appropriate beamforming scheme based on latency, power consumption, energy consumption and probability of access associated with directional cell search in mmWave communication.

In mmWave communication, beamforming is considered as a key technology to combat the high path-loss, and unlike in conventional microwave communication, beamforming may be necessary even during initial access/cell search. Beamforming solutions proposed for LTE are quite different from what is required for mmWave communication. In LTE, initial access is performed using omnidirectional communication, and beamforming is used only when directional information is available after initial physical layer access. However, in mmWave communication beamforming may be required even during the initial access/cell search process to overcome the coverage problems of omnidirectional communication with mmWaves. In this chapter, we study and compare analog, digital, hybrid beamforming and also propose a low power phase shifters network based architecture. Focusing on the directional initial cell search problem, we compare these schemes in terms of their consumed power and energy and of the associated delay and probability of access error.

Particularly, we introduce the topic and discuss the related work in Section 4.1. In Section 4.2 we discuss that analog beamforming can still be a viable choice when context information about mmWave base stations (BS) is available at the mobile station (MS), and also propose a low power phase shifters network based architecture. In Section 4.3 we further extend our analysis and compare the energy consumption associated with different beamforming architectures in the context of directional cell search. Finally, we conclude the chapter in

Section 4.4.

4.1 Introduction and Related Work

In the past few years, there has been an increased interest in millimeter wave technology to fulfill the data rate requirements foreseen for fifth generation cellular communication (5G) [62]. However, frequencies in mmWave bands experience high path-loss, which in comparison to microwave bands may result in a significant coverage reduction when considering omnidirectional communication. To overcome these coverage issues, beamforming at mmWave is an effective solution, and is necessary even during initial cell discovery. Moreover, the small wavelengths at mmWave frequencies allow a large number of antennas to be packed in a small space, and this allows to generate high gains and highly directional beams.

In addition, low latency is also an essential requirement for 5G [12]. Low latency can be achieved by incorporating a high bandwidth sub-carrier. However, this increase in the bandwidth requires higher sampling frequency for an analog-to-digital converter which corresponds to a higher power consumption. This power consumption and latency product is even more critical during initial cell discovery (due to large search delays) and directly dictates the choice of an appropriate BF scheme (especially for the MS's receiver design).

Analog, Digital and Hybrid beamforming (ABF, DBF and HBF) are the commonly discussed BF schemes for mmWave communication [16]. The choice of an appropriate BF scheme is critically related to its power consumption, and it is even of more importance for a mmWave receiver design especially at the MS which has limited available battery. In general, as discussed in earlier chapters, DBF requires a separate RF chain for each antenna element and therefore consumes the highest power, whereas ABF consumes the least power. However, DBF allows to form multiple simultaneous beams to cover the whole angular space, whereas ABF only converges to a single beam at a particular time instant. Therefore, ABF results in a larger time delay than DBF or HBF due to the additional time required to form beams in different directions to cover the whole angular space. This shows a trade-off between the power consumption and the directional search delay, which is critical during initial cell discovery (ICD). In this chapter, we study the performance of different beamforming schemes in terms of probability of access error and also study the power consumption vs time delay trade-off along with the energy consumption, and identify the appropriate BF scheme at the

MS for ICD.

4.1.1 Related Work

Recently, two different approaches have been considered for directional initial access/cell discovery. Firstly, in [47], considering a HetNet scenario, context information (CI) regarding mobile station (MS) positioning is provided to the mmWave base station (BS) by the microwave BS. Based on this, the mmWave BS points its beam (using analog beamforming) in the desired direction. The authors also address the issue of erroneous CI, proposing that the BS, in addition to searching in the CI based direction, also searches the rest of the angular space by forming beams in different directions and also with different beamwidth (to increase the coverage). Results showed that the enhanced cell discovery, where in case of positioning error the BS searches the adjacent angular directions, outperforms the greedy search approach where the BS searches the angular space sequentially. In addition, at the MS, omnidirectional reception is considered, which in comparison to a directional reception results in a reduced gain. Recently, the authors of [47] extended their work by considering a more complex channel model with multiple rays and obstacles [63].

Secondly, a non context information based approach is used in [46, 64, 65]. In [46], omnidirectional and random directional transmissions are considered from the mmWave BS, and the MS can perform either analog, digital or hybrid beamforming. It is observed that using an omnidirectional transmission at the BS and digital beamforming at the MS outperforms the other schemes. However, the use of omnidirectional transmission at the BS during initial synchronization/access will reduce coverage, while random directional transmission will result in large delays.

In [64], exhaustive and hierarchical searches are compared while considering analog, digital and hybrid beamforming at the BS and the MS. In exhaustive search, the whole angular space is covered by sequentially transmitting beams in a time division multiplexing fashion, and initial beamforming is done by selecting the best combination of Tx-Rx beams. The hierarchical search, instead, is a multiple step process. In the first step, a MS initially utilizes fewer antennas to form a relatively small number of wide beams. The received signal is combined with all the beams and the best combiner beam is selected as a reference for the next step, where several narrower beamwidth combiners are formed, within the initially selected beam. Considering scenarios with limited mobility, the process finishes when the combiner beam is within the range of 5° to 10°. However, selecting an incorrect combiner in the initial stage can result in an initial access error in the following stage.

Recently, in [65], iterative and exhaustive search schemes using analog beamforming have been studied and compared, and the authors showed that the optimal scheme depends on the target SNR regime.

The authors in [66] suggested to scan the complete angular space sequentially to identify the best BF direction both at the MS and the base station (BS). A delay based comparison for initial access in mmWave cellular networks is studied in [67], where it is shown that DBF has a lower delay than ABF without any performance degradation.

In [44], the total power consumption based comparison of an ABF and a DBF for the mmWave receiver at the MS is studied. Results showed that with low power ADC and a fixed power budget, DBF may result in a higher capacity than ABF when the channel state information is available at the transmitter. In contrast to previous works, in this work we consider the availability of CI regarding the mmWave BS positioning at the MS, and consider beamforming both at the MS and at the BS. Moreover, focusing on initial cell discovery (ICD) for the CI and the non-CI based scenario, we also study power consumption vs time delay trade-off and identify the choice of an appropriate BF scheme based on the total energy consumption.

4.2 Context Information based Directional Cell Discovery

In this section, we study and compare analog, digital, hybrid and phase shifters network based beamforming schemes and evaluate their performance considering the availability at the MS of context information regarding positioning of the mmWave BS.

Among the proposed beamforming schemes for initial cell search, analog beamforming is a power efficient approach but suffers from its inherent delay during the initial cell search phase. In this work, we argue that the availability of context information can reduce the search delay associated with the analog beamforming architecture. However, the performance of analog beamforming degrades in case of angular errors in the available context information. To address this, we present an analog beamforming receiver architecture that uses multiple arrays of Phase Shifters and a single RF chain to combat the effect of angular errors, showing that it can achieve the same performance as hybrid beamforming.

4.2.1 Our Approach

Analog beamforming (ABF) results in a lower power consumption compared to digital and hybrid beamforming (DBF and HBF) but has been shown to have comparable performance in terms of initial access probability when combined with an exhaustive search [64]. Although the exhaustive approach is not preferred in general due to its inherent search delay, the availability of CI can greatly improve its delay performance. Particularly, CI reduces the uncertainty in AoA, i.e., based on angular error statistics, it either provides the correct beamforming direction or reduces the directional search to some short angular interval $[-\phi_{max}, \phi_{max}]$. To address this, in contrast to [47], we consider the availability of CI at the MS, so that the MS, rather than searching the whole angular space, will form its combining beam only in the direction provided by the CI[1]. This will result in a reduced initial cell search delay and, in comparison to omnidirectional reception (as in [47]), the ABF at the MS will provide a higher gain. Moreover, the availability of CI at the BS does not necessarily minimize the delay, as if there are multiple MSs belonging to different beams the BS has to scan all of them, losing at least part of the delay savings. On the contrary, in initial cell search the MS typically listens to a single (or a small number of) BS, and therefore the availability of CI allows it to form a single beam in the right direction, thereby avoiding the beam scanning delay.

We consider a HetNet scenario for transferring the CI to the MS, where a microwave BS during the exchange of the initial control signals also transfers the location information of the mmWave BS (e.g., global positioning system (GPS) coordinates) to the MS. Although, GPS based CI provides accurate position information, it also has some limitations, e.g., 1) in indoor scenarios it is difficult to receive GPS signals, and 2) to form a beam in the desired direction, in addition to GPS coordinates, information regarding mobile orientation is also necessary. To address this, possible options could be, 1) to consider CI as the beamforming vector associated with the last communicating direction or 2) CI regarding the desired beamforming direction can be predicted based on the statistics of recent communicating directions.

---

[1] The initial cell search delay associated with the directional communication (for $N_{BS} = 64$ and $N_{MS} = 16$ with ABF) is roughly around 10 ms (Section 4.3.3), and is small enough to ingnore the rotational motion of the MS, which makes the directional signal reception at the MS a feasible option.

These options (non-GPS based) also have their pros and cons. For instace, CI can be computed locally, however, the accuracy of CI depends on how the current beamforming vector is associated with the last communicating direction or with the statistics of recent communicating directions. In this work, we focued on GPS based CI and a detailed performance comparison of different schemes for obtaining CI is left as part of the future work.

Due to the recent increase in location based applications and GPS positioning accuracy, GPS based CI may be considered as a viable option. We assume that the GPS coordinates of the MS are already available. The MS, after acquiring the CI, figures out the expected angle of arrival (AoA) and aligns its beam in the required direction to receive the initial synchronization signals from the mmWave BS. In this work, our results give a quantitative evaluation of the benefit of having CI, by measuring how much better a mmWave system can perform with the availability of CI. A detailed assessment of the cost of obtaining the CI with assisted GPS will be provided in the next section.

Specifically, in this section we address the following issues:

- how ABF with CI performs in comparison to ABF with non CI based approaches (Random [46] and Exhaustive [64] search);
- how the angular error in the provided CI will affect the performance of the initial access process;
- how the optimal number of MS antennas that results in the best access performance varies with the angular error in the available CI.

Finally, we propose an analog beamforming based phase shifters network (PSN) architecture (see Figure 4.1) to mitigate the effect of the angular error in the available CI, and also compare its performance with hybrid and digital beamforming in terms of initial access error and power consumption.

The rest of this section is organized as follows. In Section 4.2.2, we define the MIMO based system with beamforming and extend the idea to CI and also present our PSN architecture. Next, we discuss the simulation results for the different mentioned beamforming schemes in Section 4.2.3.

4.2.2 System Model

Consider a downlink mmWave MIMO communication system, with N_{BS} BS antennas and N_{MS} MS antennas. A typical MIMO received signal model, without considering beamforming, is given as [50]

$$\mathbf{y} = \sqrt{P}\mathbf{H}\mathbf{s} \; + \; \mathbf{n} \tag{4.1}$$

where $\mathbf{H}$ is $N_{MS} \times N_{BS}$ matrix which represents the channel between the BS and the MS, $\mathbf{s}$ and $\mathbf{y}$ are the transmitted and received symbol vectors, respectively, P is the transmitted power, and $\mathbf{n}$ is the complex white Gaussian noise, $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2)$. The mmWave MIMO channel can be modeled with a few scatterers (L) [62] and is well represented by the following geometric model [48]

$$\mathbf{H} = \sqrt{\frac{N_{BS}N_{MS}}{\rho L}} \sum_{l=1}^{L} \eta_l \mathbf{a}_{MS}(\phi_l)\mathbf{a}_{BS}^{H}(\theta_l) \tag{4.2}$$

where H represents the conjugate transpose, ρ is the path-loss, η_l is the complex gain associated with the l^{th} path, $\mathbf{a}_{MS}$ and $\mathbf{a}_{BS}$ are the spatial signatures of the MS and the BS, respectively, and ϕ_l and $\theta_l \in [0, 2\pi)$ represent the AoA and angle of departure (AoD) of the l^{th} path at the MS and the BS, respectively. In this work, for simplicity we restrict our analysis to single path line of sight (LOS) scenarios and therefore will not consider the subscript l in further analysis, whereas the study of a multi-path scenario is left as future work. Moreover, both the MS and the BS are equipped with a uniform linear array (ULA). Now, considering the ULA at the BS, $\mathbf{a}_{BS}$ is defined as

$$\mathbf{a}_{BS} = \frac{1}{\sqrt{N_{BS}}}[1, e^{j(2\pi/\lambda)d\sin(\theta)}, ..., e^{j(N_{BS}-1)(2\pi/\lambda)d\sin(\theta)}]^T \tag{4.3}$$

where T represents the transpose, d is the spacing between antenna elements, and λ is the wavelength of the transmitted signal. For $d = \lambda/2$, $\mathbf{a}_{BS}$ becomes

$$\mathbf{a}_{BS} = \frac{1}{\sqrt{N_{BS}}}[1, e^{j\pi\sin(\theta)}, ..., e^{j(N_{BS}-1)\pi\sin(\theta)}]^T \tag{4.4}$$

and the spatial signature $\mathbf{a}_{MS}$ can be defined similarly.

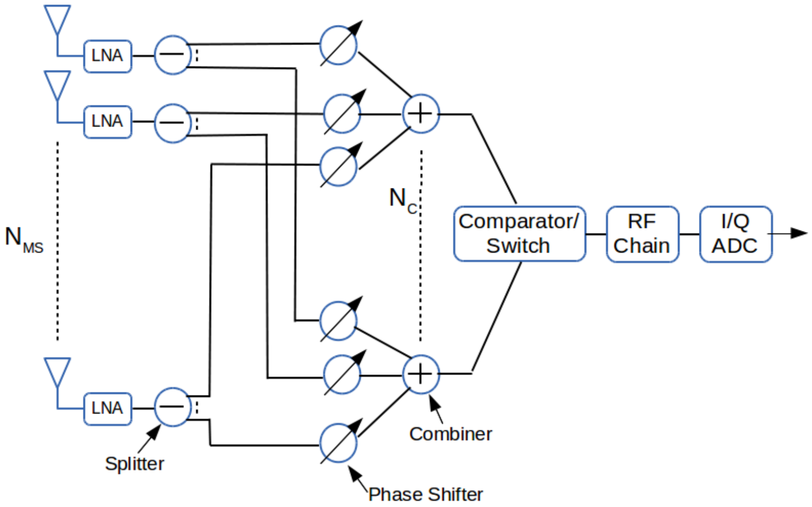

Fig. 4.1. *A Phase Shifters Network based receiver architecture.*

4.2.2.1 Signal Model with Analog Beamforming

The received signal, after applying beamforming and combining at the transmitter (the BS) and the receiver (the MS), can be written as

$$y = \sqrt{P}\mathbf{w}_{MS}^{H}\mathbf{H}\mathbf{w}_{BS}s \ + \ \mathbf{w}_{MS}^{H}\mathbf{n} \tag{4.5}$$

where $\mathbf{w}_{BS}$ and $\mathbf{w}_{MS}$ are the transmit beamforming and receiver combining vectors, respectively. For simplicity in the notation, the subscripts for the BS and the MS are removed in our general description of beamforming vectors, however, we explicitly use them whenever necessary.

Considering ABF, the beamforming vector $\mathbf{w}$ models the analog phase shifters which apply successively progressive phases to the antenna elements. These phase shifters are digitally controlled and thus result in a finite number of possible phases. Considering q quantization bits, 2^q different phases can be applied to each element of an antenna array. The set of these quantized phases is represented as

$$t = \left\{0,\ 2\pi(\frac{1}{2^q}),\ 2\pi(\frac{2}{2^q}),\ ...,\ 2\pi(\frac{2^q-1}{2^q})\right\} \tag{4.6}$$

and with N antennas, a codebook of 2^{qN} beamforming vectors can be generated. However, in order to reduce the complexity, we only consider a subset of this codebook, and let $\mathbf{W}$ =

$[\mathbf{w}^1 \ \mathbf{w}^2 \ \ldots \ \mathbf{w}^{N_q}]$ represent an $N \times N_q$ reduced size codebook, where each column of $\mathbf{W}$ represents a unique beamforming vector. Now for a quantized phase $\vartheta_i \in t$, we generate the receiver combining vector $\mathbf{w}^i$ as

$$\mathbf{w}^i = \frac{1}{\sqrt{N}}[1, \ e^{j\vartheta_i}, \ e^{j2\vartheta_i}, \ ..., \ e^{j(N-1)\vartheta_i}]^T \tag{4.7}$$

where $i = 1, \ 2, \ldots \ N_q$. Moreover, we set the cardinality of $\mathbf{W}$ as $card(\mathbf{W}) = N_q = 2N$, and therefore by definition of $\mathbf{w}^i$ in (4.7), the required number of quantization bits is $q = \log_2(2N)$. This gives an acceptable performance, as shown in [68]. The transmitter and the receiver beamforming vectors are assumed to be defined according to the above approach, and $\mathbf{W}_{MS}$ and $\mathbf{W}_{BS}$ represent the beamforming codebooks at the MS and the BS, respectively.

4.2.2.2 MS Combining Vector Selection with the Available CI

We consider a scenario where the BS transmits sequentially in different angular directions using ABF, and the MS combines the received signal with a combining vector selected based on the available CI. Note that the quantized phase ϑ_i is related to the physical AoA (ϕ) as $\phi = \sin^{-1}(\vartheta_i/\pi)$. The MS first estimates the CI based AoA (ϕ_{CI}) and then identifies ϑ_i that minimizes the angular distance with ϕ_{CI}. Finally, the combining vector that corresponds to the best ϑ_i is selected. This combining vector maximizes the norm $|{\mathbf{w}^i_{MS}}^H \mathbf{a}_{MS}(\phi_{CI})|$, where $\mathbf{a}_{MS}(\phi_{CI})$ represents the CI based estimated spatial signature at the MS. This maximization problem can be expressed as

$$\begin{aligned} \mathbf{w}^{\star}_{MS} = \ & \underset{\mathbf{w}^i_{MS}}{\arg\max} \, |{\mathbf{w}^i_{MS}}^H \mathbf{a}_{MS}(\phi_{CI})| \\ & s.t. \ \mathbf{w}^i_{MS} \in \mathbf{W}_{MS} \end{aligned} \tag{4.8}$$

the formulation of (4.8) also maximizes the signal to noise ratio (SNR) of the received signal (Equation (4.5)), expressed as

$$SNR = \frac{|\mathbf{w}^H_{MS}\mathbf{H}\mathbf{w}_{BS}|^2 P}{||\mathbf{w}^H_{MS}||^2 \sigma^2} \tag{4.9}$$

where P is the power of the transmitted signal, and σ^2 is the noise power.

We now extend this approach by considering an error in the available CI. This will add an angular error ϕ_e to the estimated AoA, which may result in a wrong selection of ϑ_i and $\mathbf{w}_{MS}$, which finally decreases the SNR of the received signal. We model ϕ_e as uniformly distributed in $[-\phi_e^{max}, \phi_e^{max}]$, where ϕ_e^{max} represents the maximum angular error[2]. The amount by which ϕ_e affects the selection of ϑ_i is related to the 3 dB beamwidth θ_{BW} of the antenna array, which for a ULA with $d = \lambda/2$ and N antennas can be computed as $\theta_{BW} = 2\sin^{-1}(0.891/N)$ [69]. Therefore, for a reliable performance it is desirable to have $\phi_e^{max} \leqslant \theta_{BW}$. However, to address larger ϕ_e^{max}, i.e., $\phi_e^{max} > \theta_{BW}$, one possible solution is to increase θ_{BW} by reducing the number of antennas at the cost of a reduced gain. The other option is to form multiple beams simultaneously, as discussed in the next section.

4.2.2.3 Phase Shifters Network to Combat ϕ_e

The performance of ABF (which only results in a single beam) starts degrading as ϕ_e^{max} becomes comparable to θ_{BW}. A solution to mitigate the effect of large ϕ_e^{max} is to form multiple simultaneous beams and then identify among them the best combining vector. HBF or DBF are attractive options to form multiple simultaneous beams[3], but at the cost of higher power consumption. Compared to a fully Digital architecture (DBF), HBF is preferred for mmWave communication due to its lower power consumption because of its reduced number of RF chains. However, the required number of ADCs in HBF, which are considered as the main power hungry blocks in mmWave receiver design, increases proportionally to the number of RF chains. Therefore, in comparison to ABF, HBF and DBF may result in significantly higher power consumption.

To jointly address the issue of higher power consumption and large ϕ_e^{max}, we propose a phase shifters network (PSN) architecture. The architecture of PSN is similar to ABF but instead of forming a single beam with one combining vector, PSN allows the formation of multiple beams simultaneously by using multiple combining vectors. The idea is to form multiple beams and identify the best combining vector which corresponds to the desired AoA all in the analog domain. Therefore, PSN provides a power efficient design with only two ADCs (for the inphase and the quadrature phase signals), whereas in HBF the number of ADCs is directly proportional to the number of RF chains. The PSN receiver architecture is

---

[2] The general trends followed by different beamforming schemes (as shown in Section 4.2.3) will be maintained if different angular error distributions are used.

[3] For details regarding HBF and DBF, see [48] and [70], respectively.

shown in Figure 4.1[4], where N_C is the number of combiners. The number of simultaneous beams is equal to N_C, where phase shifters connected to a particular combiner represent a unique receiver combining vector. The received signal is combined with N_C combining vectors and the output of each combiner is compared using a comparator and with the help of a switch only the output of the best combiner (the one with the strongest signal) is forwarded to the ADC for further digital processing.

The architecture of PSN is also similar to HBF, with the exception that it consists of a single RF chain and allows beamforming only in the analog domain. Therefore, in comparison to HBF and DBF, PSN does not provide the advantages of digital beamforming. However, during cell search the inherent advantages of digital beamforming like multiplexing, interference cancellation or multiuser communication are not required at the MS[5]. Therefore, PSN, thanks to its lower power consumption receiver design, is a viable option for mmWave initial cell search.

To form multiple beams, the selection of the main combining vector for PSN follows a similar formulation as in Equation (4.8). However, in contrast to ABF where only a single combining vector is required, to form multiple beams in PSN N_C different combining vectors from codebook $\mathbf{W}_{MS}$ must be used. Among the N_C different combining vector, the main combining vector is selected according to (4.8), while the other $N_C - 1$ combining vectors are selected from $\mathbf{W}_{MS}$ such that they have minimum angular distance from the main combining vector. According to the formation of the combining matrix $\mathbf{W}_{MS}$, this results in a selection of $N_C - 1$ combining vectors that are adjacent to the main combining vector. This is a suitable choice as $\mathbf{W}_{MS}$ is generated in a way to ensure that the gain fluctuation among any two adjacent combining vectors is within 1 dB [68].

4.2.3 Simulation Setup and Results

In our performance evaluation, we assume that the CI of the mmWave BS is already available at the MS and with this location information the MS can point its beam in the desired direction. We also assume that the mmWave BS always transmits the signal in differ-

---

[4] Note that PSN compares and selects the best beam in analog. Therefore, the advantage of an SNR improvement using matched filer is not available which may result in a wrong beam selection at low SNRs.

[5] Note that, in order to get the highest energy efficiency in the first phase of initial access while enjoying the advantages of digital beamforming in subsequent phases and during data communication, the PSN architecture can be easily converted to HBF by switching the output of each combiner to a separate RF chain rather than to the comparator. A detailed study of such hybrid architecture is part of our future work.

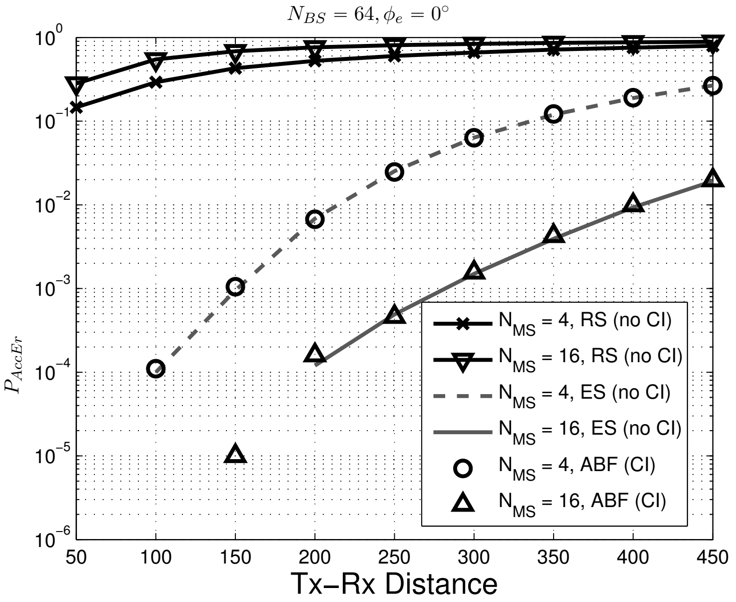

Fig. 4.2. *A comparison of Random Search, Exhaustive Search, and CI based schemes.*

ent directions sequentially using ABF, while for a better comparison different beamforming schemes are considered at the MS. A more comprehensive comparison considering different beamforming options also at the BS is left for future study.

In our simulations, we consider a single path LOS scenario ($L = 1$), carrier frequency of 28 GHz, transmit power of 30 dBm, noise figure of 5 dB, thermal noise power of -174 dBm/Hz, path-loss exponent of 2.2, and 64 transmit antennas at the BS. The simulation results are averaged over 10^5 different channel realizations for each Tx-Rx distance. We evaluate the performance in terms of access error probability P_{AccEr}, which is defined as the probability that the SNR of the received signal at the MS (Eq. (4.9)) is below a certain threshold Θ.

$$P_{AccEr} = \Pr(SNR < \Theta) \tag{4.10}$$

In our simulations this threshold Θ is set to -4 dB [64].

We start with a comparison of our proposed CI (with ABF) based initial cell search scheme with two non CI (with ABF) approaches, namely Random search (RS) [46] and Exhaustive search (ES) [64]. Figure 4.2 shows how P_{AccEr} for CI and non CI based approaches varies for different values of Tx-Rx Distance and N_{MS}. Results show that RS performs worst in comparison to the other two schemes. This is because in RS a MS forms its beam in a randomly selected direction, and therefore it is less likely to form a beam in the required direction, which results in a higher initial access error. Also in RS, as N_{MS} increases, the beamwidth gets narrower and the MS has to form more beams to cover the complete angular space, and therefore the probability of finding the right combining beam decreases, which results in an increase in P_{AccEr}. On the other hand ES, where a MS searches the whole angular space sequentially, and our proposed CI based initial cell search approach show similar performance. However, considering ABF both at the BS and the MS, the exhaustive approach has a large search delay because $card(\mathbf{W}_{BS}) \times card(\mathbf{W}_{MS})$ joint angular directions need to be considered, whereas for our CI based approach this search delay is reduced to $card(\mathbf{W}_{BS}) \times 1$, at the cost of obtaining CI. For the rest of this section, the performance evaluation is only focused on the CI based approach. We now discuss how the CI based approach performs in the presence of the angular error in the available CI, and also how this angular error is related to N_{MS}.

In Figure 4.3, P_{AccEr} is plotted for different values of ϕ_e^{max}, while considering 4 and 16 antennas at the MS, respectively. As expected, for $\phi_e^{max} = 0$, with an increase in N_{MS} the beamforming gain increases, and therefore P_{AccEr} decreases. For both $N_{MS} = 4$ and $N_{MS} = 16$, P_{AccEr} increases with an increase in the Tx-Rx distance or the angular error ϕ_e. The increase in P_{AccEr} with an increase in ϕ_e^{max} for $N_{MS} = 4$ is not very significant due to its wider beamwidth ($\theta_{BW}^4 = 25.7^\circ$). However, in the case of $N_{MS} = 16$, the beamwidth decreases ($\theta_{BW}^{16} = 6.38^\circ$), and therefore with an increase in the angular error it is more likely that the incoming signal will not fall within θ_{BW}, and this results in a significant increase in P_{AccEr}.

Figure 4.3 also highlights how the angular error is related to the optimal number of receive antennas. It is observed that when the angular error starts falling outside of θ_{BW} of the estimated beam, P_{AccEr} also increases. Therefore, an increase in N_{MS} makes it more susceptible to angular errors and hence results in even worse performance. It is obvious from Figure 4.3 that the optimal number of antennas varies with ϕ_e^{max}. For instance, with $\phi_e^{max} = 0^\circ$, $N_{MS} = 16$ is the better option, whereas for $\phi_e^{max} = 10^\circ$, $N_{MS} = 4$ performs better than $N_{MS} = 16$. Moreover, for $\phi_e^{max} = 5^\circ$, $N_{MS} = 16$ has a lower P_{AccEr} for Tx-Rx distance 50 m and above, whereas below 50 m $N_{MS} = 4$ is a preferable choice. This shows a tradeoff between distance dependent path-loss and ϕ_e. At short Tx-Rx distance,

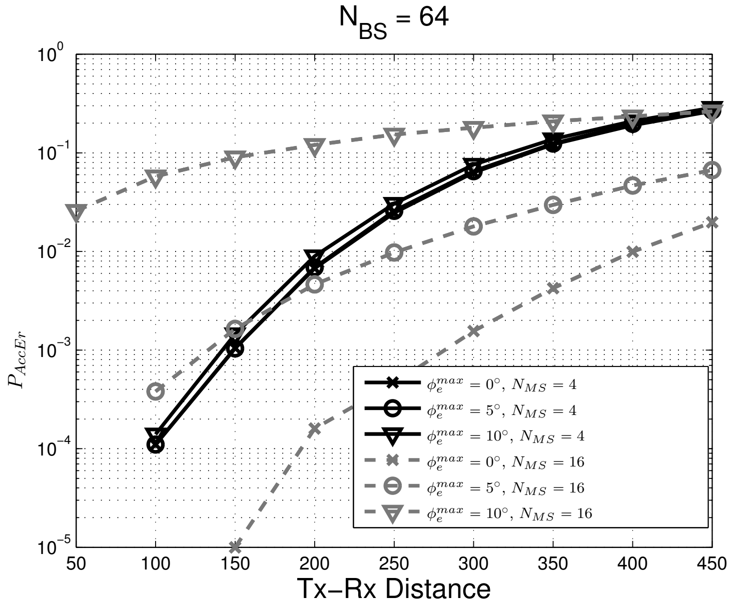

Fig. 4.3. *Access error probability vs Tx-Rx distance for ABF with CI for different values of* N_{MS} *and with* ϕ_e^{max}.

ϕ_e has more significance on P_{AccEr} and therefore $N_{MS} = 4$ performs better, however as the Tx-Rx distance increases the path-loss also increases and therefore $N_{MS} = 16$ with higher beamforming gain results in a lower P_{AccEr}.

We now compare the performance of PSN with ABF and show how PSN can mitigate the effect of ϕ_e. For PSN, we consider $N_C = 3$. Figure 4.4 presents a comparison of ABF and PSN for different angular errors. For both schemes, P_{AccEr} increases with an increase in ϕ_e^{max} or Tx-Rx distance. However, PSN performs better than ABF, and is less prone to a small increase in ϕ_e^{max}. For instance, the performance of PSN is unaffected for $\phi_e^{max} = 5°$, whereas there is an increase in P_{AccEr} for $\phi_e^{max} = 10°$. This is because the PSN architecture allows to form multiple simultaneous beams and is able to capture the signal energy even in the presence of angular errors. The performance of PSN in case of angular error improves with N_C, as a larger N_C will result in more combiner beams covering more angular space and thus is less susceptible to angular errors.

As PSN outperforms ABF in case of erroneous CI, we now further evaluate its performance by comparing it with HBF and DBF. In simulations, for PSN and HBF $N_C = N_{RF} = 3$.

Figure 4.5 presents a comparison of DBF and HBF with PSN. It is depicted that DBF shows a stable performance with an increase in the angular error. This is because DBF allows

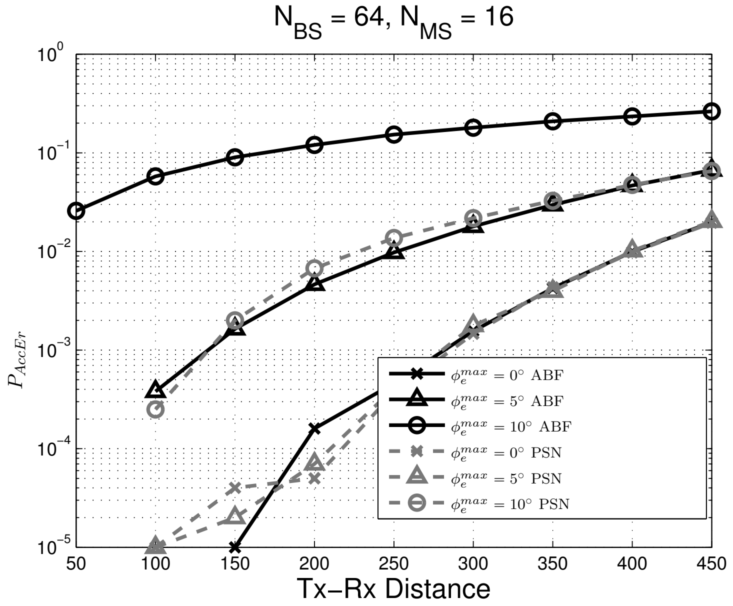

Fig. 4.4. *Access error probability vs Tx-Rx distance for ABF and PSN with* $N_C = 3$.

to form multiple beams which allows the MS to look in all angular directions simultaneously, and therefore the performance of DBF is not effected by angular errors (but at the cost of higher power consumption). Note that the performance of PSN and HBF for small angular errors is similar to that of DBF.

Moreover, for $\phi_e^{max} = 0°$, HBF results in a slightly lower P_{AccEr} than PSN due its additional digital baseband combiners (which further minimizes the distance between ϕ_{CI} and ϑ_i), but the performance difference between PSN and HBF diminishes with an increase in ϕ_e^{max}. Also, note that the digital baseband combiner in HBF requires additional RF chains, which results in higher power consumption. Therefore, PSN is a viable solution as it provides significant power savings at the cost of only a small performance degradation at lower ϕ_e^{max}.

In order to make the above considerations about power consumption of the various beamforming schemes more precise, we study how the total receiver power consumption P_{Tot}, varies with an increase in the number of ADC bits for ABF, DBF, HBF and PSN. In our evaluation, we consider $N_C = N_{RF} = 3$ and $N_{MS} = 16$. P_{Tot} for ABF, HBF and DBF can be calculated similar to Equations (2.1), (2.2) and (2.3), respectively, whereas for PSN can be calculated as

$$\begin{aligned} P_{Tot}^{PSN} = {} & N_{MS}(P_{LNA} + P_{SP} + N_C P_{PS}) \\ & + N_C P_C + P_{RF} + P_{Comp} + P_{Sw} + 2P_{ADC} \end{aligned} \tag{4.11}$$

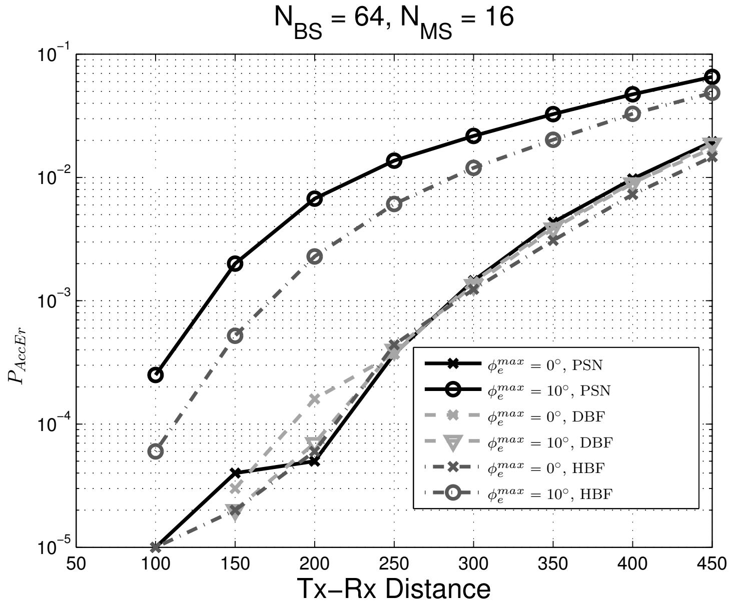

Fig. 4.5. *Access error probability vs Tx-Rx distance for PSN, HBF and DBF.*

where $P_{RF} = P_M + P_{LO} + P_{LPF} + P_{BB_{amp}}$ represents the power consumption of the RF chain, and where P_{LNA}, P_{PS}, P_C, P_M, P_{LO}, P_{LPF}, $P_{BB_{amp}}$, P_{ADC}, P_{SP}, P_{Sw} and P_{Comp} represent the power consumption of low noise amplifier (LNA), phase shifter, combiner, mixer, local oscillator, low pass filter, baseband amplifier, ADC, splitter, switch and comparator, respectively. The power consumption of an ADC scales exponentially with the number of bits and linearly with the sampling rate [43] and can be calculated similar to Eq. 2.5. For $B = 500$ MHz and $b = 5$, the power consumption of an ADC in [42] is considered as 200 mW, and this correponds to $c = 12.5$ pJ. For details regarding the power consumption of other components is according to the values mentioned in Chapter 2.

Figure 4.6 shows that with an increase of b there is an exponential increase in power consumption. As expected, P_{Tot} for ABF is always lower than for the other beamforming schemes, and hence ABF with CI is a suitable option for initial access considering small angular errors. On the other hand, DBF always has higher power consumption than other beamforming schemes except for $b = 1$. Finally, PSN consumes lower power than HBF irrespective of the number of ADC bits. Moreover, the P_{Tot} difference for PSN and HBF increases with an increase in b. Therefore, PSN with P_{Tot} less than DBF and HBF is a viable option even for high resolution (i.e., higher number of bits) ADC based mmWave receiver design.

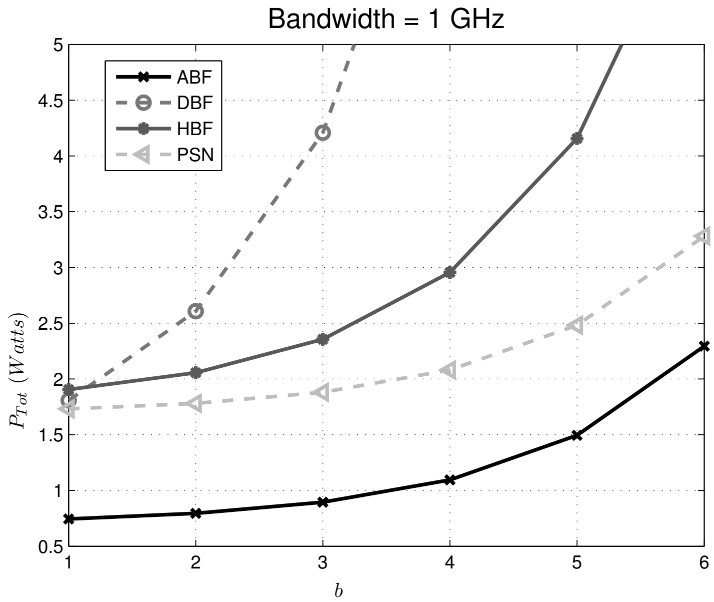

Fig. 4.6. *Power consumption comparison of ABF, DBF, HBF and PSN.*

In next section we will take our analysis one step further and study how delay and energy consumption associated with the initial cell discovery process vary with a change in the sub-carrier bandwidth.

4.3 Directional Cell Discovery: Energy Consumption

In this section, we provide an energy consumption based comparison of different beamforming schemes while considering both a low power and a high power analog-to-digital converter for a millimeter wave based receiver at the MS. We analyze both context information (GPS positioning based) and non context information based schemes, and show that analog beamforming with CI (where the mmWave base station's positioning information is already available at the MS) can result in a lower energy consumption, while in all other scenarios digital beamforming has a lower energy consumption than analog and hybrid beamforming. We also show that, under certain scenarios, the recently proposed phase shifters network architecture can result in a lower energy consumption compared to other beamforming schemes. Moreover, we show that the energy consumption trend among different beamforming schemes is valid irrespective of the number of ADC bits. Finally, we propose a new signaling structure which utilizes a relatively higher frequency sub-carrier for primary synchronization signals compared to other signaling, which allows a further reduction in initial cell search delay and energy consumption at the MS.

4.3.1 Our Work

In this work, we compare the receiver energy consumption (EC) associated with different BF schemes during initial cell discovery. Primarily, we focus on an EC comparison among the ABF, DBF, HBF and PSN receiver architectures at the MS. EC is computed as the product of power and time delay. Power consumption for different beamforming schemes is computed considering similar mmWave receiver architectures as in Chapter 2, while delay is calculated following a synchronization signaling structure as used in LTE. For power consumption, we consider both a low power ADC (LPADC) and a high power ADC (HPADC) model. We also consider both context information and non CI based scenarios for initial cell discovery. In CI based ICD, we consider that the GPS coordinates of the mmWave base stations (BS) position are provided to the MS by the microwave BS. We divide the CI based scenario into 1) where the GPS coordinates of the MSs positioning are already available, and 2) where the MS has to acquire its GPS coordinates by using assisted GPS (AGPS) which introduces an additional delay[6]. Moreover, we only consider CI based scenario for AFB, HBF and PSN, as DBF already allows to form multiple beams simultaneously to cover the complete angular space, and therefore CI for DBF will not result in any notable advantage. In the rest of the section, we refer to non CI based scenario as nCI, CI with GPS coordinates already available, i.e., CI with no delay as CInD, and CI with the delay of acquiring the GPS coordinates as CID. Considering the above mentioned scenarios, we show that

- All BF schemes have a similar time delay for CInD scenario, whereas in other scenarios DBF has a minimum delay.
- EC of all beamforming schemes decreases with an increase in the sub-carrier bandwidth.
- ABF (for both LPADC and HPADC) only results in a lower EC in the CInD scenario.
- In all scenarios except CInD, DBF results in a lower EC than ABF and HBF.
- PSN has a lower EC than other BF schemes for higher sub-carrier bandwidth.

---

[6] As acquiring GPS coordinates directly from the MS results in a larger delay and energy consumption, we restrict our analysis to the AGPS case.

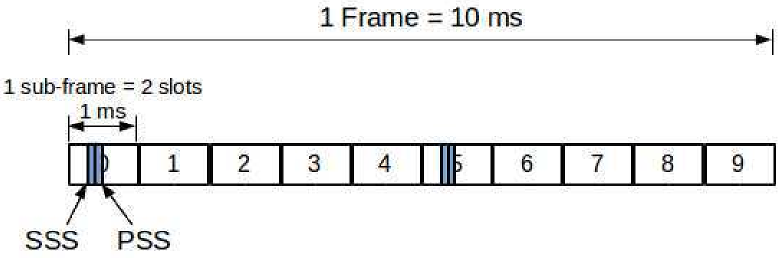

Fig. 4.7. *LTE FDD frame structure.*

- The EC trends among different BF schemes are independent of the number of ADC bits.

Finally, we propose a new signaling structure for 5G, where the primary synchronization signals are generated with a higher sub-carrier bandwidth while all other signaling is performed at a lower bandwidth. This allows the BS to cover the angular space with a relatively short delay than with lower sub-carrier bandwidth, and therefore also reduces the energy consumption.

4.3.2 Initial Cell Discovery

Initial cell discovery procedures proposed for conventional cellular networks are relatively different from what is required for mmWave cellular communications. Conventionally, the transmission and the reception of the primary synchronization signals (PSS) and secondary synchronization signals (SSS) (essential for ICD and synchronization) at the BS and the MS, respectively, is done omnidirectionally. However, in mmWave cellular networks, directional communication is essential even during the ICD process to exploit the BF gain. In addition to beamforming, reduced latency is another important requirement for 5G cellular networks and this can be achieved by increasing the bandwidth of the sub-carrier B_{SC} which also increase the total system bandwidth B_{Tot}. An increase in B_{Tot} reduces the latency, but results in a higher power consumption at the mmWave receiver, mainly because the ADC needs to sample the incoming signal at a higher rate. We next discuss the LTE control signaling model to evaluate the synchronization signaling time delay.

Table 4.1. *Synchronization Signal Time Period and Total System Bandwidth (B_{Tot}) Associated to Different B_{SC}*

B_{SC} (MHz)	0.015 (LTE)	0.25	0.5	1	10
T_{PSS}	5 ms	0.3 ms	0.15 ms	75 μs	7.5 μs
B_{Tot} (MHz)	1.4	23.3	46.7	93.3	933.3

4.3.2.1 Initial Cell Discovery in LTE

To initialize the ICD procedure, the MS starts looking for the reception of the synchronization signals (i.e, PSS and SSS). These synchronization signals provide the initial synchronization and also identify the physical layer cell identity. There are 504 different physical layer cell identities, where the SSS has 168 sequences and provides the cell identity group and the PSS has 3 different sequences and provides identity within the cell identity group. In FDD, the PSS and the SSS are transmitted once every 5 ms for a duration of 71.4 μs (Figure 4.7). Both synchronization signals constitute 6 resource blocks (corresponds to 72 sub-carriers) with 62 sub-carriers centered around DC (irrespective of the operating bandwidth of LTE) while the rest 10 sub-carriers are unused. After the successful reception of the PSS and the SSS, the MS finishes the initial cell discovery and starts the initial access process.

4.3.2.2 Extension to mmWave

Unlike in LTE, in mmWave 5G cellular networks an additional requirement for initial cell discovery is to identify the required beamforming direction to improve the SNR at the receiver. In comparison to LTE, this identification of the beamforming direction results in an additional ICD delay. This delay depends on the size of the antenna array and on the beamforming scheme incorporated both at the MS and at the BS. In this work, we consider 64 antennas at the BS and 16 antennas at the MS, which with beamspace beamforming requires $64 \times 16 = 1024$ different directional scans (N_D) to cover the complete angular space. The total ICD delay is a combination of the delay associated with 1) the control signaling (PSS and SSS) and 2) the identification of the beamforming direction.

4.3.3 Energy Consumption in mmWave ICD

We measure the energy consumption for different BF schemes considering the three different scenarios discussed in Section 4.3.1. As EC is computed as the product of power

Table 4.2. *t_{Del} for Different BF Schemes Corresponding to Different ICD Scenarios, with ABF at the BS*

	t_{Del} (nCI)	t_{Del} (CInD)	t_{Del} (CID)
ABF	$1024 \times T_{PSS}$	$64 \times T_{PSS}$	$64 \times T_{PSS} + t_{CI}$
DBF	$64 \times T_{PSS}$	$64 \times T_{PSS}$	$64 \times T_{PSS}$
HBF	$256 \times T_{PSS}$	$64 \times T_{PSS}$	$64 \times T_{PSS} + t_{CI}$
PSN	$256 \times T_{PSS}$	$64 \times T_{PSS}$	$64 \times T_{PSS} + t_{CI}$

consumption and time delay, we first compute the time delay associated with nCI, CInD and CID, whereas the power consumption of the different BF schemes will be studied in the following subsections.

4.3.3.1 Time Delay

We follow the LTE signaling structure to calculate the delay associated with the reception of synchronization signals, and identify the delay corresponding to different values of the sub-carrier bandwidth B_{SC}. This delay corresponds to the time period associated to the transmission of the PSS signal, T_{PSS}, for different ICD scenarios, as shown in Table 4.1. To receive the initial synchronization signals, the MS has to look in each BF direction for a duration of T_{PSS}. Therefore, the total time delay t_{Del} can be computed as the product of T_{PSS} and N_D. The total time delay corresponding to different BF schemes and different ICD scenarios is shown in Table 4.2. We consider only ABF at the BS while the MS can form its combining beam using either ABF, DBF, HBF or PSN. This shows that the BS at a particular instant can transmit the synchronization signal only in one angular direction, therefore the BS requires 64 (equal to the number of antennas) different transmissions at separate time instants to cover the whole angular space. Moreover, the angular search delay at the MS is related to the available beamforming scheme.

With ABF, the MS has to look in 16 different angular directions, which results in $N_D = 64 \times 16 = 1024$, the maximum scan search delay. Unlike ABF, N_D for HBF (PSN) is directly related to the number of available RF chains (combiners). In this work, the number of RF chains and combiners for HBF and PSN is set to 4. This allows both architectures to form 4 simultaneous beams in different angular directions, and therefore $N_D = 64 \times 16/4 = 256$ for both HBF and PSN. Finally, DBF only requires 64 directional scans to cover the complete angular space. The above mentioned N_D corresponds to the nCI scenario, and t_{Del} can be computed as the product of N_D and T_{PSS} (Table 4.2). This t_{Del} nCI showed that DBF has

Table 4.3. *Receiver Power Consumption for Different BF Schemes for HPADC*

B_{SC}	DBF	ABF	HBF	PSN
15 KHz	1.31	1	2.43	2.49
250 KHz	1.87	1.03	2.57	2.52
500 KHz	2.57	1.07	2.72	2.56
1 MHz	3.67	1.15	3.02	2.63
10 MHz	25.1616	2.4891	8.369	3.978

Table 4.4. *Receiver Power Consumption for Different BF Schemes for LPADC*

B_{SC}	DBF	ABF	HBF	PSN
15 KHz	1.28	0.996	2.43	2.3
250 KHz	1.3	0.998	2.43	2.3
500 KHz	1.32	0.999	2.44	2.3
1 MHz	1.37	1	2.45	2.31
10 MHz	2.22	1.06	2.66	2.36

the least delay whereas ABF results in maximum delay.

We consider the availability of CI only for ABF, HBF and PSN, as the fully digital architecture in DBF allows to form beams in all angular directions simultaneously, and therefore CI will not provide any significant advantage with DBF receivers. The availability of CI is advantageous for ABF, HBF and PSN as it reduces the number of directional scans. With the availability of CI with no delay (i.e., CInD), the receiver already has the knowledge of the appropriate combining direction and therefore the number of directional scans for ABF, HBF and PSN with CInD is similar to that for DBF (i.e., 64). This shows that t_{Del} for CInD for all beamforming schemes is similar.

Finally, a scenario where the GPS coordinates of the MS are not available (i.e., CID) results in an additional time delay. To acquire the GPS coordinates, we considered an AGPS protocol in which the time delay t_{CI} corresponding to the acquisition GPS coordinates is generally communication channel dependent. However, to simplify the analysis we consider the Time to first fix (TTFF) with AGPS to be approximately around 1 s (as mentioned in [71]), and assume $t_{CI} = 1.5$ s, slightly more than TTFF. This is a relatively short delay, however, we show that even this short t_{CI} results not only in a higher delay but also in a relatively higher energy consumption compared to the other design choices.

4.3.3.2 Power Consumption

Receiver power consumption is an important concern due to the utilization of huge bandwidths at mmWave communication. In this work, we consider the total power consumption of the receiver. The power consumption of all other components except ADC is considered independent of the total system bandwidth. We consider two different power consumption models for an ADC, i.e., a low power and a high power ADC[7]

The total system bandwidth (as shown in Table 4.1) is an important parameter, as an increase in B_{Tot} linearly increases the ADC power consumption (as B_{Tot} defines the required sampling rate of an ADC). B_{Tot} is directly proportional to the sub-carrier bandwidth due to which an increase in B_{SC} results in a higher B_{Tot} and therefore a higher ADC power consumption. B_{Tot} for different B_{SC} is computed following the LTE format used for the 1.4 MHz bandwidth. In LTE, with 1.4 MHz of bandwidth, 6 RBs are transmitted at each time instant. A RB consists of 12 sub-carriers with $B_{SC} = 15$ KHz. This corresponds to a bandwidth of 1.08 MHz for 6 RBs, and therefore with total bandwidth of 1.4 MHz, the utilized bandwidth is 77.14% (i.e., 1.08 MHz × 100/1.4 MHz = 77.14%). B_{Tot} for different B_{SC} can be computed similarly and is listed in Table 4.1.

The power consumption (in Watts) of different beamforming corresponding to different B_{SC} for a HPADC and a LPADC is listed in Table 4.3 and Table 4.4, respectively. The power consumption values are obtained for a 6 bits ADC. The listed power consumption values for different beamforming schemes are computed considering the same receiver architecture and per component power consumption values as mentioned in Chapter 2, Section 2.2 and Chapter 4, Section 4.2. The listed values show that the rate of increase in the power consumption is related not only to an increase in B_{Tot} but also to the choice of the ADC design. For instance, the rate at which power consumption increases for HPADC is much higher as compared to LPADC. Moreover, DBF with LPADC results in a lower power consumption than HBF and PSN even for $B_{SC} = 10$ MHz.

Finally, the energy consumption (EC) can be computed by evaluating the product of the total power consumption and the time delay.

---

[7] The power consumption of an ADC, denoted as $P_{ADC} = cB2^b$, increases exponentially with the number of bits b and linearly with the bandwidth B and with the ADC Waldens figure of merit c. For a HPADC $c = 12.5$ pJ whereas for a LPADC $c = 494$ fJ (Chapter 2, Section 2.2), to better quantify the energy consumption of different beamforming schemes at these two power extremes.

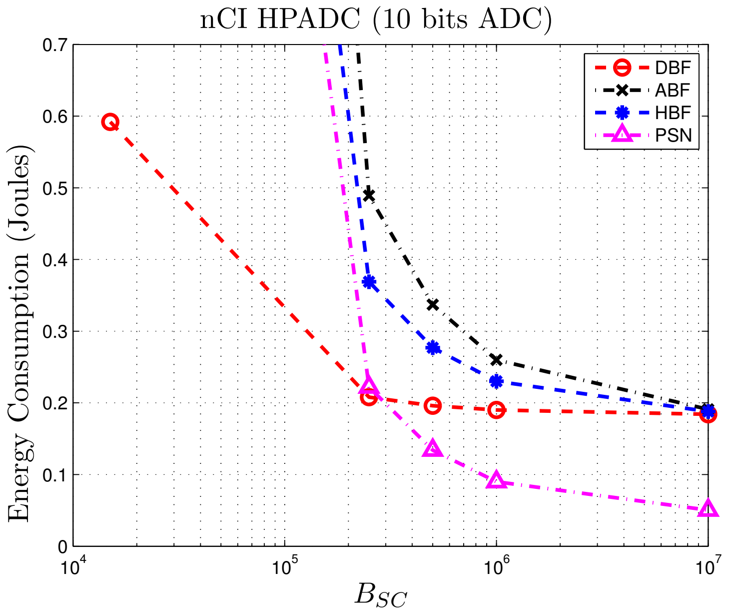

Fig. 4.8. *Energy consumption vs* B_{SC} *for different beamforming schemes in the nCI scenario and HPADC.*

4.3.4 Results

We now discuss the energy consumption trend followed by different BF receiver architectures during the ICD, and identify the choice of the appropriate BF scheme for nCI, CInD and CID scenarios while considering both a low power and a high power ADC model. The energy consumption of different BF schemes considering different B_{SC} is computed using the power consumption and time delay values listed in Tables 4.3, 4.4 and 4.2, respectively. All results except those in Figure 4.8 (where we considered 10 bits) are obtained for a 6 bits ADC.

Figures 4.8, 4.9 and 4.10 show the energy consumption vs B_{SC} for the nCI scenario. The plots show that the energy consumption of all BF schemes decreases (with a decreasing slope) with an increase in B_{SC}[8], and this decrease in EC is more rapid up to $B_{SC} = 0.3$ MHz. This follows from the fact that t_{Del} decreases with an increase in B_{SC}, therefore the EC of the receiver components other than ADC decreases. However, the EC of an ADC remains constant as the power consumption of the ADC increases with the same rate with which t_{Del} decreases. The plots also show that, even with a single RF chain, ABF results in a higher EC relative to the other beamforming for both the HPADC and the LPADC model.

---

[8] the plots of Figures 4.11-4.14 also show a similar trend i.e., decrease in the EC with an increase in the B_{SC}

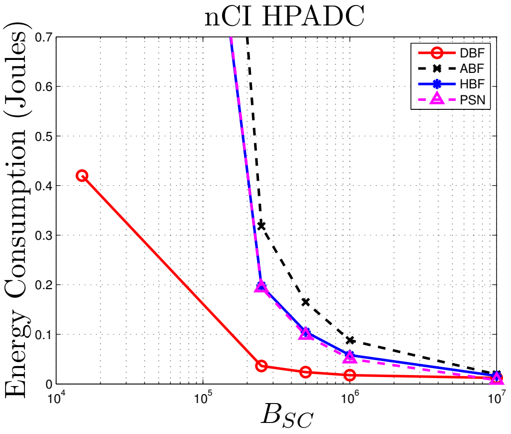

Fig. 4.9. *Energy consumption vs B_{SC} for different beamforming schemes in the nCI scenario and HPADC.*

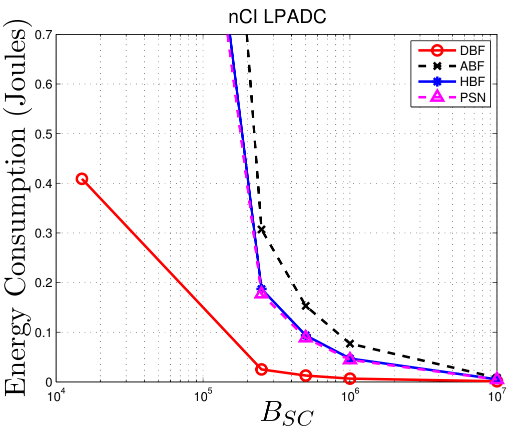

Fig. 4.10. *Energy consumption vs B_{SC} for different beamforming schemes in the nCI scenario and LPADC.*

This shows that t_{Del} for ABF is large enough (due to higher N_D) to dominate the lower power consumption of the ABF which results in the highest EC for ABF. On the other hand, interestingly, DBF always has a lower EC than ABF and HBF, irrespective of the number of ADC bits and of the ADC power model. However, the difference among the EC of DBF, ABF and HBF decreases with an increase in B_{SC} and starts converging to a similar EC. We will discuss this convergence later.

The EC trend followed by PSN is relatively different than what is observed for other BF schemes. Firstly, PSN has a higher energy consumption than DBF at lower B_{SC}, but results in a lower EC than DBF at higher B_{SC}. The converging value of the EC of PSN is also low compared to other BF schemes. This is because the PSN architecture requires ADCs similar to ABF but with much fewer N_D. Secondly, the intersection point between DBF and PSN shifts towards lower B_{SC} with an increase in the number of ADC bits. This is because an increase in bits results in a higher ADC power, due to which the ADC power starts dominating the power consumption of the other components, which results in a rapid convergence in the EC of PSN. Finally, from Figures 4.8 and 4.9 it is obvious that the EC trend among different BF schemes is valid irrespective of the number of ADC bits.

Figures 4.11 and 4.12 show an EC comparison among different BF schemes for a CInD scenario. In this scenario, it is considered that the GPS coordinates of the MS are already available and therefore all BF schemes have a similar t_{Del}. The plots show that ABF has the lowest EC in this scenario. This is due to the reduction in N_D thanks to the availability of CI regarding the direction of the mmWave BS. Also note that, despite the availability of

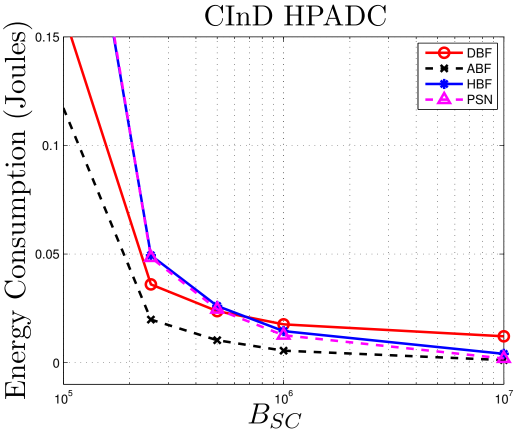

Fig. 4.11. *Energy consumption vs B_{SC} for different beamforming schemes in the CInD scenario and HPADC.*

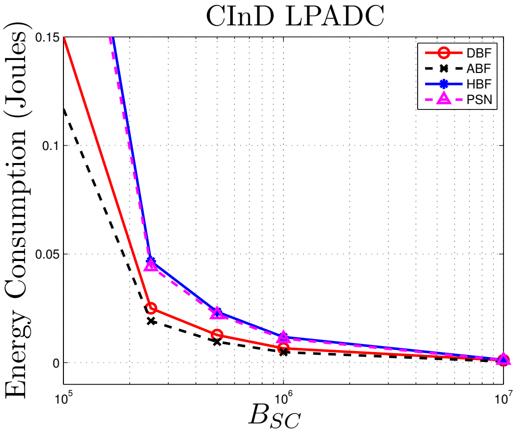

Fig. 4.12. *Energy consumption vs B_{SC} for different beam in the CInD scenario and LPADC.*

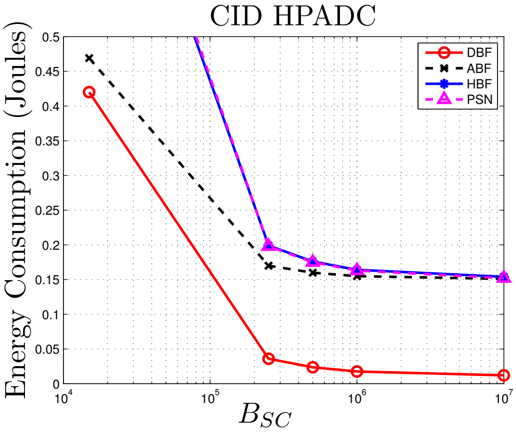

Fig. 4.13. *Energy consumption vs B_{SC} for different beam in the CID scenario and HPADC.*

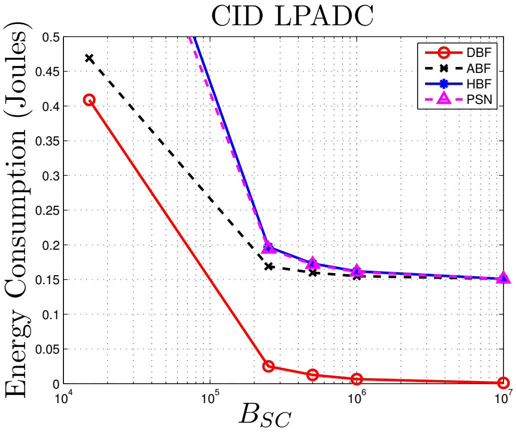

Fig. 4.14. *Energy consumption vs B_{SC} for different beam in the CID scenario and LPADC.*

CI, DBF has a lower or similar EC compared to PSN and HBF up to B_{SC} = 10 MHz and B_{SC} = 1 MHz for a LPADC model and a HPADC, respectively. The trend among HBF and PSN is similar to what was observed for the nCI scenario. Although these results depict the feasibility of ABF for ICD considering the CInD scenario, to better analyze the feasibility of ABF it is also important to note that the performance of ABF degrades with an increase in the error in the available CI, where PSN has a more robust performance (as shown in Section 4.2). Therefore the choice of the appropriate BF scheme for the CInD scenario not only depends on the minimum EC but is also important to quantify the performance of a particular BF scheme in the case of erroneous CI.

Figures 4.13 and 4.14 show the EC vs B_{SC} for a CID scenario. To compute the additional

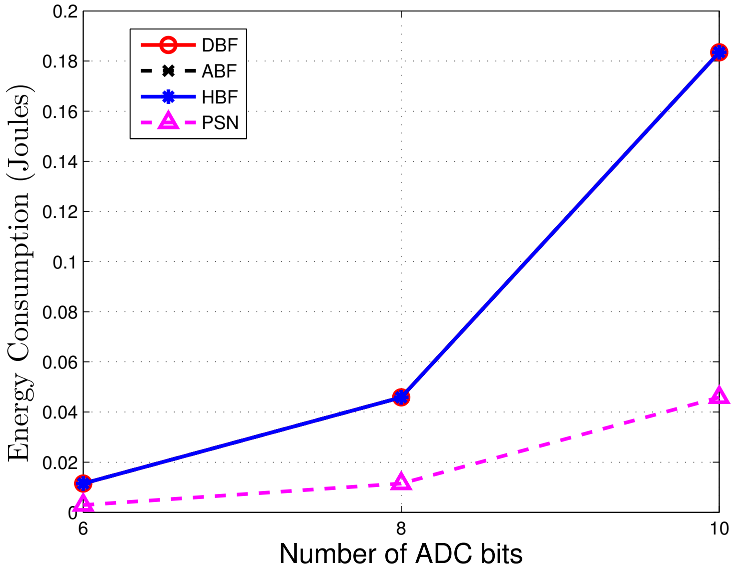

Fig. 4.15. *Energy consumption convergence for different BF schemes.*

EC corresponding to the acquisition of the GPS coordinates of the MS, the time delay is considered to be 1.5 s (as mentioned in Subsection 4.3.3.1) and the power consumption is assumed to be 100 mW. The results show that ABF has a lower EC than HBF and PSN for small values of B_{SC}, but the EC of ABF, HBF and PSN is higher compared to what observed for nCI and CInD scenarios. This is because the EC associated to the acquisition of GPS coordinates is high enough that all BF schemes result in a higher EC. Moreover, DBF (for both HPADC and LPADC model) has a lower EC than other schemes for all values of B_{SC}. Therefore, the CID scenario not only results in a higher EC for ABF, HBF and PSN but also introduces an extra time delay.

It is important to note that, for a very high value of B_{SC}, even though DBF may have a minimum search delay and results in a lower (in the CID scenario) or similar (in the nCI scenario) EC compared to other beamforming schemes, it can consume a very high power (e.g., with HPADC, $P_{DBF} = 25$ Watt for $B_{SC} = 10$ MHz, see Table 4.3). This suggests that DBF, for a very high values of B_{SC}, may not be an effective beamforming scheme for power constrained devices, such as a MS.

4.3.4.1 EC Convergence

A common trend that can be observed in Figures 4.8-4.14 is the convergence of the EC with an increase in B_{SC}. This is because the time delay is higher at lower B_{SC}, which makes the EC of all components other than the ADC more significant. However, with an increase in B_{SC} the EC of the ADC starts dominating. Therefore, the convergence value for the nCI scenario (convergence for CID and CInD can be computed similarly) can simply be calculated by evaluating the EC of the total number of ADCs present in a particular BF scheme. The EC of a single ADC can be computed as

$$E_{ADC} = P_{ADC} \times t_{Del} = cBR \times t_{Del} \tag{4.12}$$

where E_{ADC} and P_{ADC} are the energy and power consumption of an ADC, c is the energy consumption per conversion and R is the resolution of an ADC. Equation (4.12) shows that the EC of an ADC remains constant irrespective of B_{SC}, as an increase in B_{SC} will decrease t_{Del} and vice-versa. The convergence value of a particular BF scheme can be found by computing the product E_{ADC} and the number of ADC based on a particular BF scheme. The results for EC convergence are shown in Figure 4.15, where the convergence value is shown for different numbers of ADC bits while considering ABF, DBF, HBF and PSN schemes. The results show that PSN has the minimum value of the EC convergence for any number of ADC bits, whereas all other beamforming schemes converge to the same value. This is because the product of N_D and the number of ADC for ABF, DBF and HBF has a similar result. However, a PSN architecture which allows to form multiple simultaneous beams and compare them in the analog domain requires a single RF chain (i.e., similar number of ADCs as used for ABF) and can cover the complete angular space with fewer directional scans than ABF, and therefore PSN results in a lower EC.

4.3.5 Higher B_{SC} for PSS Signaling

Usually, control plane signaling requires less bandwidth compared to data plane signaling. For instance, in LTE the reserved bandwidth for synchronization signaling is limited to 1.4 MHz, whereas the total system bandwidth can be as large as 20 MHz based on the data rate requirement of the mobile user. A lower bandwidth definitely requires less power

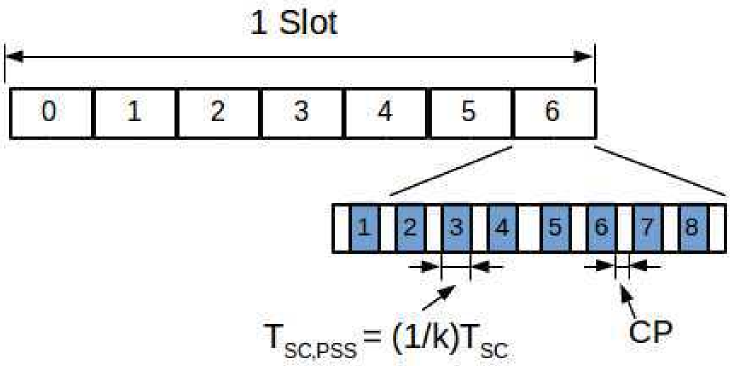

Fig. 4.16. *Proposed slot structure for PSS.*

consumption. However, as already shown, an increase in the sub-carrier frequency reduces not only the initial cell search delay but also the EC. By combining the above two points i.e., 1) the difference in the bandwidth of data plane and control plane signaling and 2) a higher sub-carrier bandwidth results in a lower latency and EC, we now propose a new signaling structure for 5G cellular networks, where PSS utilizes a relatively higher sub-carrier bandwidth compared to other signaling. The maximum increase in the PSS sub-carrier bandwidth ($B_{SC,PSS}$) can be set to the maximum available bandwidth (B_{Tot}) for data plane. We next discuss the proposed signaling structure and briefly discuss its performance, whereas a detailed analysis is left as future work.

The proposed slot structure for 5G containing PSS is shown in Figure 4.16, where $T_{SC,PSS}$ and T_{SC} represent the time period of the higher bandwidth PSS sub-carrier and reference sub-carrier bandwidth (i.e., B_{SC}^{ref}) used for other signaling, respectively, CP represents the cyclic prefix interval corresponding to $T_{SC,PSS}$ and k is a constant which represents the fractional increase or decrease in B_{SC}^{ref} or T_{SC} for PSS. In Figure 4.16, k is set to 8, and therefore in the time duration T_{SC}, 8 primary synchronization signals can be transmitted. This is because, according to the proposed structure, $B_{SC,PSS}$ is increased k times, which reduces the transmission time required for PSS by a factor of k compared to what is required for sub-carriers with bandwidth B_{SC}^{ref}. This allows to transmit multiple $B_{SC,PSS}$ signals in either a different or the same angular direction in a similar time interval as that of other signaling, i.e., T_{SC}. This has two main advantages, 1) k PSS signals can be transmitted in the same angular direction to capture higher energy in a reduced time, and 2) transmit k PSS signal into different angular direction to cover more angular space in a relatively shorter time interval to reduce the initial cell search delay. In this work we only discuss the latter.

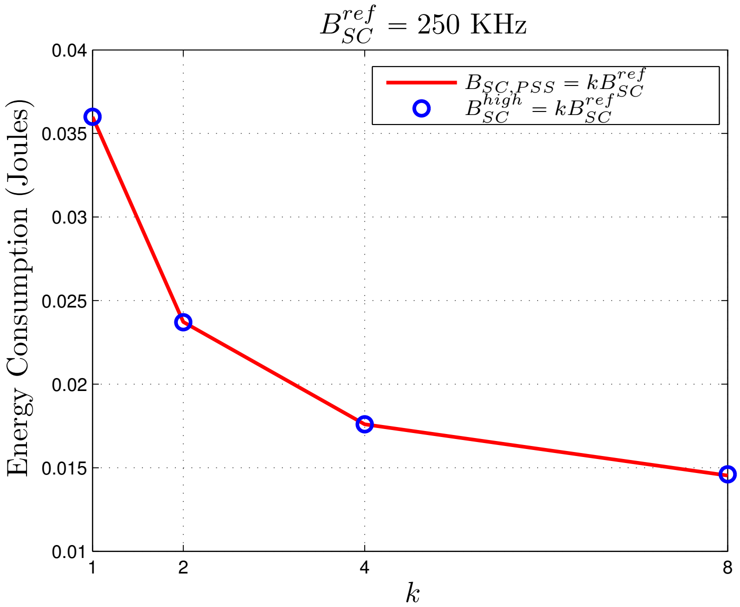

Fig. 4.17. *Energy consumption reduction with higher bandwidth sub-carrier for PSS.*

According to the proposed scheme, the reduction in the directional search delay with higher $B_{SC,PSS}$ compared to B_{SC}^{ref} can be given as

$$t_{Del,Red} = \frac{t_{Del}}{k} \tag{4.13}$$

where $t_{Del,Red}$ is a reduced delay. Note that this reduced delay is similar to what we would get with k times higher (fixed) sub-carrier bandwidth for all signaling. However, with the proposed structure (assuming ABF), the BS can transmit k PSS signals in consecutive time instants in k different directions whereas with same B_{SC} the BS can transmit PSS only in a single direction every T_{PSS}. We next discuss the EC vs k relationship.

Figure 4.17 shows the EC for different values of k. The plot shows that the EC decreases with an increase in k, and therefore the proposed signaling structure indeed reduces the EC. Moreover, the reduction in EC is similar to what can be achieved by using a fixed higher bandwidth sub-carrier B_{SC}^{high} for all signaling, as t_{Del} for both cases is similar. Therefore, the proposed structure with higher $B_{SC,PSS}$ can achieve a low EC (similar to what can be achieved with a higher B_{SC}^{high}), without increasing B_{SC} for other signaling (which allows ADCs to operate at lower sampling rate especially during data communication, which reduces the power consumption).

4.4 Conclusions

In this chapter we studied and analyzed different aspects of the directional initial cell discovery process in mmWave cellular systems. We compared the performance of ABF, HBF, DBF and PSN architectures in terms of probability of access error, latency, power consumption and energy consumption. The results showed that ABF can perform better than other schemes when context information regarding the mmWave BS is available without any significant delay. However, the performance of ABF degrades in case of erroneous CI. To address this, we proposed a low power phase shifters network based architecture which allows to form and compare multiple beams in analog simultaneously. We also showed that in all other scenarios (i.e., non CI based or CI with delay), DBF results in a lower delay and a similar or lower energy consumption then other beamforming schemes.

5

Energy Efficient Flooding of Fountain Codes

In this chapter, we investigate and analyze the transmission efficiency (which translates into energy efficiency) from a multi-hop network perspective. We consider a multihop network where a source node must reliably deliver a set of data packets to a given destination node. To do so, the source applies a fountain code and floods the encoded packets through the network, until they reach their destination or are lost in the process. We model the probability that the destination can recover the original transmissions from the received coded packets as a function of the network topology and of the code redundancy, and show that our analytical results predict the outcome of simulations very well. These results are employed to find a good working point on the tradeoff between the probability of success of end-to-end communications and the number of packet transmissions that take place in the network. Note that the reduction in the number of transmissions improves energy efficiency. We also propose practical policies to approximate the behavior of restricted and full flooding based only on local decisions at each node, and conclude by providing simple approximations for the probability that a given packet is lost on the route to the destination. Finally, we also analyze how the availability of a helper node which decodes and injects extra redundancy packets in the network can further reduce the required number of transmissions, and therefore improves the energy efficiency of the network.

5.1 Introduction

In the past decade, multi-hop flooding networks have been of great interest in the research community. Flooding based transmission improves the communication reliability as a destination can receive multiple copies of the same packet transmitted from different nodes. In addition, multi-hop routing allows to improve the communication range. Such networks are an attractive choice in scenarios where infrastructure is non-existent or unusable. Examples of such scenarios include emergency recovery, rescue operation, and underwater networks. For instance, the peculiar nature of the underwater channel limits the probability of successful communication. In addition, an underwater channel only supports acoustic waves over long range communication which substantially limits the data rate compared to terrestrial wireless networks. Note that communication at higher frequencies suffers from high attenuation which limits the communication range. In this scenario, communication reliability and data rate (while keeping a similar communication range) can be improved by using the advantages of flooding and multi-hop communication, respectively.

Although multi-hop flooding networks improve the reliability and range of communication, such wireless communication networks also result in a lower energy efficiency (due to redundant transmissions) which is of concern in future wireless communication. To address both reliability and energy efficiency, we study a trade-off between the required number of transmissions and the probability of success at the destination. We devise restricted flooding models to reduce the required number of transmissions while keeping the desired probability of success at the destination.

In addition to flooding, coding can be incorporated to further improve the communication reliability. In this work, we consider fountain codes which are well known for improving probability of successful communication. Fountain codes [25, 72] have been introduced to achieve reliable communications in generic networks, where the packet loss probability may be unknown a priori. The latter condition makes fountain codes particularly well suited to wireless networks [73, 74], where the ability to dynamically adjust the amount of redundancy offers a means to compensate for link breakages and communication failures [75].

In multihop wireless networks, and especially in those conditions where a high degree of reliability is sought [76], fountain-coded packets may be coupled with some form of redundant network-layer transmission pattern, e.g., flooding [77]. Flooded fountain codes introduce a double layer of redundancy: on one hand, each coded packet inherently carries information about other source packets, thereby providing utility not just to an intended relay, but rather to all upstream nodes receiving the packet; on the other hand, each coded source packet is repeated by a number of relays that participate in the flooding process. While this substantially increases the opportunities for each node between a source and its destination to correctly decode the packets (and thus be able to retransmit them), the amount of packet replication that results is potentially excessive, and may lead to a waste of resources and therefore is not energy efficient.

Still, fountain codes have been widely considered in the literature as a flexible error control technique in both wired and wireless networks experiencing erasure channels. For example, the SYNAPSE++ protocol [73] employs a fountain code to reliably disseminate firmware over a wireless sensor network. Fountain codes have also been shown to improve the performance of unicast communications, provided that early acknowledgment packets are sent in order to prevent the generation of useless redundancy [78]. In [76], the introduction of fountain codes in vehicular networks leads to a significant performance improvement compared to a simpler approach where data packets are periodically retransmitted. Delay-Tolerant Networks (DTNs) [79] and distributed storage systems [80] also benefit from the distribution of redundancy that comes from the use of fountain codes. Therefore, it is important to design a procedure to disseminate fountain-coded data through a network in an effective yet controlled way, that exploits the double level of redundancy of coding and flooding in order to deliver packets successfully, while at the same time avoiding the consumption of an excessive amount of resources (e.g., avoiding useless transmissions and the resulting energy wastage).

In our recent work [81], we analyzed the interplay between fountain codes and flooding, by considering both a full flooding and a restricted flooding policy. Given the amount of redundancy generated at the source, the number of hops, and the number of relays that interact at each hop, we derive the average number of transmissions required to achieve a prescribed probability of success at the destination. The latter is employed to analyze the interplay between the number of transmissions and the code redundancy. Note that in this chapter we focus on the reduction in the number of transmissions, which can directly be translated into a reduction in the energy consumption or an improvement in the energy efficiency.

In this work, we extend the work in [81] in several respects. First, while we maintain the analytical model simple and tractable, we extend simulations to include a more realistic physical layer model including path loss and fading. Simulations are then employed to validate our analytical results, which are extended to forecast the behavior of the network in different topologies and under different deployment assumptions. Our policies are evaluated based both on a simple error model and on a more realistic block fading model, which are compared to validate our analysis via simulation results. Second, we propose a simple family of policies that implement restricted flooding distributedly and based only on the local decisions of each node. We also formulate an adaptive policy that approaches the behavior of full flooding, but with a significantly lower retransmission overhead. Both policies combine forwarding ideas with practical medium access control-level issues that help reduce the complexity and cost of packet forwarding. Finally, we introduce low-complexity approximations to the analytical formulas in our model.

The remainder of this chapter is organized as follows. In the next section we survey the related literature. In Section 5.3 we introduce our network model and analytically characterize the probability of success and the number of transmissions taking place in the network under different packet forwarding schemes, and introduce practical and distributed forwarding policies; in Section 5.4, we evaluate the performance of these policies compared to the results from the analytical model; in Section 5.5 we discuss several possible extensions to our work and develop one such extension in detail. Section 5.7 provides the conclusions, and finally in Section 5.8 we provide low complexity approximations of our analysis.

5.2 Related Work

Several works in the literature have applied fountain codes to scenarios of interest, mostly considering end-to-end communications throughout a network. In [75], an analytical model is provided for the throughput performance of a fountain code transmitted through either a conventional or a cooperative multihop network, both in the delay tolerant and in the delay-constrained case, assuming a Nakagami fading model. The work in [82] derives an analytical expression for the average number of received packets in delay-constrained networks, whereas [78] measures the complexity of different fountain code relaying strategies. In [83], the authors consider a multihop scenario with a single node at each relay stage, and analyze the probability distribution of the transmission time in the presence of interference from randomly distributed nodes. However, in [75, 78, 82], multihop communications take place through a single, fixed path determined a priori, rather than considering a flooding scenario as we do in this work.

In [84], erasure codes have been applied to multicast transmissions of short duration in a single-hop cellular scenario. Fountain codes have also been shown to improve data dissemination [85], MAC layer communications [86] and reliable transport [74] in underwater networks. Random linear packet coding is applied to single-hop broadcast networks in [87]. Transmit power adjustment policies are proposed based either on the performance of the worst link or on the average performance of all links. In [88], in order to increase the reliability of packet delivery over multihop unicast connections, a node determines the lost packets thanks to an implicit acknowledgment procedure and retransmits them at a later time if needed. In [89], restricted flooding is proposed, where a node transmits with probability p after receiving the mth copy of a message, where p is selected based on percolation theory.

Several approaches have been proposed to mitigate the overhead coming from the excessive transmission redundancy. In [90], a network coding based approach is discussed to figure out the minimum forwarding probability required to achieve a given outreach probability over random graphs. The authors in [91, 92] discuss dynamic probabilistic broadcasting schemes in MANETs: the retransmission probability is adjusted based on the number of neighboring nodes (the higher the number of neighboring nodes, the lower the rebroadcasting probability). In [93], adaptive probabilistic flooding is applied to path discovery in multipath routing: the retransmission probability of advertisement messages is progressively reduced to limit the flooding process. A similar scheme is also considered in [94] to reduce transmissions in a vehicular ad hoc network, where the rebroadcasting probability is tied to the vehicle speed. Recently [95] proposed that a node tunes network coding based on estimates of the packet reception probability for the neighboring nodes, in order to maximize the aggregate number of source packets decoded by all neighbors while limiting the dissemination delay. These estimates are evaluated based on the link quality correlation among the neighboring nodes.

In [96], lower bounds for the required number of packet retransmissions at the MAC layer are derived in order to support broadcast with and without applying network coding techniques. In [97], a reliable broadcast transmission approach based on random linear network coding is considered, in which all source packets are initially transmitted, and once the ACK/NACK storm has finished, the source transmits network-coded packets to improve the performance of the worst receiver. In [98], the authors proposed a low complexity and energy efficient fractional transmission scheme (FTS) broadcasting approach that employs fountain codes over a multihop wireless network. Based on hop distance, a fraction of the encoded packets to be transmitted to a particular node is assigned among its neighboring nodes. If the sum of the received fractions exceeds one, then the assigned fractions can be adapted to reduce the number of redundant transmissions. After successful decoding, a node starts transmitting additional encoded packets.

Unlike in previous works, in this work we study the interaction of flooding and fountain codes via a fundamental approach. We start with a mathematically tractable model of the network performance that is proven to match realistic simulations quite accurately. We employ the analysis to identify desirable working points in terms of retransmission overhead and probability of success, and finally proceed to propose heuristic policies that can approximate the desired probabilistic forwarding behavior, in terms of both probability of success and overhead. Unlike in such works as [88, 89, 93], our practical policies rely on decisions that are made by each node in a fully distributed manner. Unlike [90, 95–97], our policies do not rely on further network coding, so that the source can be in control of the behavior of the flooding process.

5.3 Scenario Description, Analytical Model, and Practical Flooding Policies

We consider a multihop network scenario, where a source S has m packets to send to a given destination node D. To protect the communication from packet losses, the source encodes these m packets into $M > m$ packets using a random fountain code over the Galois field of size η, $\mathbb{F}_\eta$, $\eta \geqslant 2$. We define $E = M - m$ to be the number of redundancy packets injected in the network to favor the decoding process at the destination and to combat packet loss over the network links. All packets are forwarded towards the destination through a multihop network where we assume that there are r relay stages between S and D, where each stage i consists of N_i relays. This way, the shortest path joining S and D consists of $r+1$ hops.[1]

---

[1] It could be argued that allowing nodes at generic stage i to receive transmissions from farther stages rather than just $i-1$ would help improve the probability of correct reception at stage i and, as a consequence, at the destination as well. In our extensive set of simulations (not shown here due to lack of space) we confirmed that this is not the case, except in a very restricted set of scenarios with benign channels (i.e., very low p) and very few nodes in each stage. The latter is specifically typical of standard multi-hop routing scenarios, which are not the focus of this work. In all other cases, the gain is negligible because either a) the probability

In addition, we assume a Bernoulli error model, whereby the probability of packet error over each link is constant and equal to p. While this assumption is required to make the problem tractable, we will show in Section 5.4 that it is a very good approximation for more realistic link error models. In the following three subsections, we will model the forwarding process from S to D, by considering different types of flooding. We will derive the probability that D can decode the fountain code, as well as the number of transmissions that the network carries out to advance the fountain code packets towards the destination. Specifically, Section 5.3.1 focuses on a baseline, unrestricted flooding paradigm; Section 5.3.2 describes distributed and tunable forwarding policies to achieve a good tradeoff between the probability of decoding success at D and number of transmissions in the network. Section 5.3.3 extends the analytical model to a form of restricted flooding, which is leveraged to provide insight on the practical policies; finally, Section 5.3.4 comments on the fundamental interplay between the Galois field size employed for the fountain code and the network performance.

5.3.1 Unrestricted flooding model

We start by assuming that the forwarding paradigm adopted in the network is a baseline full flooding, i.e., a multihop forwarding process where every unique packet is transmitted once by the source, and forwarded exactly once by each node that correctly receives it. Possible collisions among the transmissions by two or more relays of either the same or different stages can be avoided in several ways. Feasible measures include channel sensing, fine-tuned backoff times and transmission pipelining [73], and typically introduce additional delays. However, we note that applications relying on flooding to convey data through a network are typically most likely concerned about the probability of successful delivery than they are about the time it takes to achieve it. Therefore, in this work, we focus mainly on the probability of success, discuss a model to measure it, and present effective distributed ways to achieve successful packet delivery while limiting the number of transmissions in the network. In this light, a collision avoidance mechanism that introduces some delay is not a restraining assumption.[2]

---

of error from stage $i-2$ to stage i is very high or b) the flooding process compensates for transmission errors and is thus the main reason behind successful reception at stage i.

[2] A possible alternative for collision modeling would be to incorporate the probability of collision into the link error probability p. Such an extension entails the analysis of a very interesting interplay between MAC- and routing-level performance, and is left as a future extension.

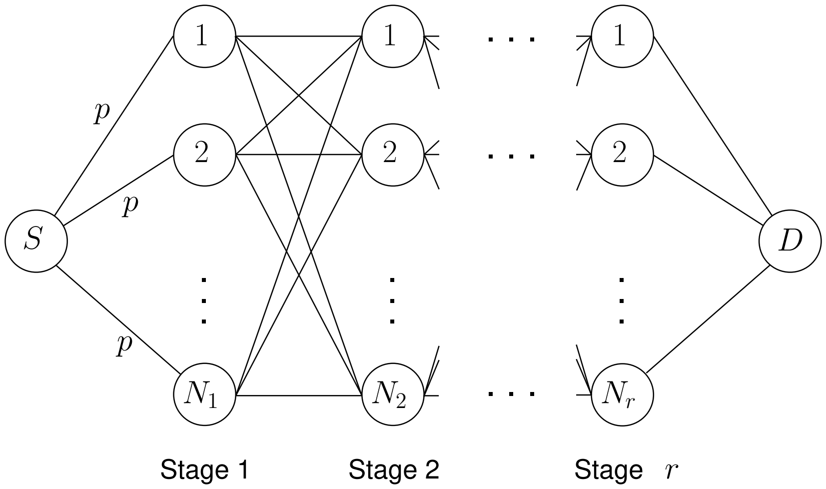

Fig. 5.1. *Reference scenario for a multihop network with multiple relay stages between a source S and a destination D. The probability of error over any link is equal to p.*

For reference, a scheme of our network scenario is provided in Fig. 5.1.

We recall that, assuming at least $\ell > m$ packets are received by D, the probability $P_{e,D}$ that D cannot decode the fountain code is upper-bounded by [99]

$$P_{e,D} \leqslant \eta^{-(\ell-m)} \tag{5.1}$$

where η is the size of the Galois field over which the code is designed, e.g., $\eta = 2$ for a binary fountain code.

Based on the assumptions above, we now derive the probability of decoding error at D. Assume for the moment that $r = 1$, i.e., there is only one relay stage between S and D, and the length of the path from S to D is 2 hops. The event that a given encoded packet does not reach D is equivalent to the event that no relay received the packet correctly, or that all relays that received the packet fail to forward it further to D. The probability of this event is

$$q_1 = \sum_{j_1=0}^{N_1} \binom{N_1}{j_1} p^{j_1}(1-p)^{N_1-j_1} p^{N_1-j_1} = \big(p + (1-p)p\big)^{N_1} , \tag{5.2}$$

where the argument of the summation means that if $N_1 - j_1$ nodes out of N_1 receive the packet correctly, all these nodes fail in forwarding it to D over the second hop. We note that the number of nodes that fail to receive a packet at the relay stage follows a binomial distribution $\mathcal{B}(N_1, p, j_1)$ with parameters N_1 and p, hence its average is $N_1 p$. Considering the degenerate cases where $p = 1, j_1 = N_1$ (all nodes fail to receive if the link error probability is identically 1) and when $p = 0, j_1 = 0$ (all nodes are successful if the link error probability is identically 0), we note that $\mathcal{B}(N_1, p, j_1)$ should be extended to yield 1 in both cases. We therefore define

$$\overline{\mathcal{B}}(N_1, p, j_1) = \begin{cases} 1 & \text{if } p = 1, j_1 = N_1 \text{ or } p = 0, j_1 = 0 \ , \\ \binom{N_1}{j_1} p^{j_1} (1-p)^{N_1 - j_1} & \text{otherwise} \ . \end{cases} \tag{5.3}$$

Assume now that $r = 2$, i.e., there are two relay stages between S and D, respectively containing N_1 and N_2 relays. The probability that a packet is correctly received by any given number of relays $1, 2, \ldots, N_2$ of stage 2 depends on the number of nodes that successfully received the same packet at stage 1. Hence the probability that the packet does not reach D is

$$q_2 = \sum_{j_1=0}^{N_1} \sum_{j_2=0}^{N_2} \overline{\mathcal{B}}(N_1, p, j_1)\, \overline{\mathcal{B}}(N_2, p^{N_1 - j_1}, j_2) p^{N_2 - k} \ , \tag{5.4}$$

where we note that, for a given number j_1 of nodes that failed the reception at stage 1, the probability of failing to advance a packet from stage 1 to any node in stage 2 is $p^{N_1 - j_1}$. In the general case, there are r relay stages between S and D, and the expression for the probability of error can be given as

$$q_r = \sum_{j_1=0}^{N_1} \sum_{j_2=0}^{N_2} \cdots \sum_{j_r=0}^{N_r} \overline{\mathcal{B}}(N_1, p, j_1)\, \overline{\mathcal{B}}(N_2, p^{N_1 - j_1}, j_2) \cdots \overline{\mathcal{B}}(N_r, p^{N_{r-1} - j_{r-1}}, j_r) p^{N_r - j_r} \ , \tag{5.5}$$

where $j_1, j_2, \ldots, j_r$ is the number of relays that failed the reception of the packet at the first, second, up to the rth stage, respectively, and therefore $(N_{r-1} - j_{r-1})$ is the number of nodes which received the packet correctly at the $(r-1)$th relay stage. Finally, by virtue of the approximate formula in (5.1), D fails to decode the fountain code with probability 1 whenever it correctly receives only m or fewer coded packets, whereas the probability of decoding failure is upper-bounded as in (5.1) whenever at least $m + 1$ packets are received. Therefore we have the following upper bound:

$$P_{\text{fail}}(q_r) \leqslant \sum_{k=0}^{m} \overline{\mathcal{B}}(M, 1 - q_r, k) + \sum_{k=m+1}^{M} \overline{\mathcal{B}}(M, 1 - q_r, k) \eta^{-(k-m)} \tag{5.6}$$

which in non-degenerate cases can be explicitly re-written as

$$P_{\text{fail}}(q_r) \leqslant \sum_{k=0}^{m} \binom{M}{k} q_r^{M-k}(1-q_r)^k + \sum_{k=m+1}^{M} \binom{M}{k} q_r^{M-k}(1-q_r)^k \, \eta^{-(k-m)} \,, \tag{5.7}$$

where $k-m$ is the number of extra packets received by D.

The formula in (5.5) does not admit a closed-form expression, and neither do (5.6) and (5.7). However, we will now illustrate a conveniently compact way to write (5.5), which will be leveraged to compute other metrics. We start by grouping the probabilities that a given number of relays fail packet reception at a given stage into matrices. To do so, we define the following probabilities

$$p_{S,j} = \overline{\mathcal{B}}\big(N_1, p, j\big) \tag{5.8}$$

$$p_{i,jk} = \overline{\mathcal{B}}\big(N_i, p^{N_{i-1}-j}, k\big) \tag{5.9}$$

$$p_{D,k} = p^{N_r-k} \,. \tag{5.10}$$

In particular, $p_{S,j}$ is the probability that j nodes at the first relay stage fail to receive the packet from S; $p_{i,jk}$ is the probability that k nodes at the ith relay stage fail to receive the packet given that j nodes failed to receive it at stage $i-1$; $p_{D,k}$ is the probability that the destination D fails to receive the packet, given that k nodes (out of the available N_r nodes at stage r) failed to receive it at stage r. We now define the column vectors $\mathbf{p}_S = [\, p_{S,0} \;\; p_{S,1} \;\; \cdots \;\; p_{S,N_1} \;\; \mathbf{0}_{1\times(N_{\max}-N_1)} \,]^T$, $\mathbf{p}_D = [\, p_{D,0} \;\; p_{D,1} \;\; \cdots \;\; p_{D,N_r} \;\; \mathbf{0}_{1\times(N_{\max}-N_r)} \,]^T$, and the matrix

$$\mathbf{P}_i = \left[\begin{array}{ccc|c} p_{i,00} & \cdots & p_{i,0N_i} & \\ \vdots & \ddots & \vdots & \mathbf{0}_{(N_{i-1}+1)\times(N_{\max}-N_i)} \\ p_{i,N_{i-1}0} & \cdots & p_{i,N_{i-1}N_i} & \\ \hline \multicolumn{4}{c}{\mathbf{0}_{(N_{\max}-N_{i-1})\times(N_{\max}+1)}} \end{array}\right] , \tag{5.11}$$

where $N_{\max} = \max_i N_i$, and we remark that $p_{i,N_{i-1}N_i} = 1$ is the only non-zero element of the N_ith row. Eq. (5.5) can now be rewritten as

$$q_r = \mathbf{p}_S^T \left(\prod_{i=1}^{r} \mathbf{P}_i \right) \mathbf{p}_D \,, \tag{5.12}$$

where it is intended that $\mathbf{P}_1 = \mathbb{I}_{N_{\max}+1}$, the $(N_{\max}+1) \times (N_{\max}+1)$ identity matrix. The value of q_r thus derived can be finally plugged into (5.7) to obtain an upper bound to the probability of decoding the fountain code at D.

We illustrate the formulation above via a simple example. Consider a scenario with $r = 2$ and $N_1 = N_2 = 1$. In this case, $\mathbf{p}_S = [1-p \;\; p]^T$, $\mathbf{p}_D = [p \;\; 1]^T$, and matrix $\mathbf{P}_1$ is given as

$$\mathbf{P}_2 = \begin{bmatrix} \overline{\mathcal{B}}(1,p,0) & \overline{\mathcal{B}}(1,p,1) \\ \overline{\mathcal{B}}(1,1,0) & \overline{\mathcal{B}}(1,1,1) \end{bmatrix} = \begin{bmatrix} (1-p) & p \\ 0 & 1 \end{bmatrix} . \tag{5.13}$$

Finally we have

$$q_2 = \mathbf{p}_S^T \mathbf{P}_1 \mathbf{P}_2 \mathbf{p}_D = \begin{bmatrix} 1-p & p \end{bmatrix} \begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p \\ 1 \end{bmatrix} = (1-p)^2 p + (1-p)p + p \tag{5.14}$$

For convenience, note that (5.5) for $r = 2$ can be rewritten as

$$q_2 = \sum_{j_1=0}^{N_1-1} \sum_{j_2=0}^{N_2} \binom{N_1}{j_1} p^j (1-p)^{N_1-j_1} \binom{N_2}{j_2} (p^{N_2-j_1})^k (1-p^{N_1-j_1})^{N_2-j_2} p^{N_2-j_2} + p^{N_1} \;, \tag{5.15}$$

where the last term conveys the fact that if all nodes at stage 1 fail to receive a packet from the source, the destination will also fail to receive it. For $N_1 = N_2 = 1$ and $r = 2$, (5.15) expands to

$$q_2 = (1-p)^2 p + (1-p)p + 0 + p \;, \tag{5.16}$$

where the four terms are computed for $(j_1, j_2) = (0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$, respectively.

We now observe that (5.14) equals (5.16) as expected, and both equal $1-(1-p)^3$, i.e., the probability of error over a 3-hop path, where the error probability over each link is p.

Along the same lines of (5.12), we now compute the average number of transmissions that all nodes carry out to advance a given packet towards the destination (regardless of whether the packet actually reaches the destination or not). Focus on relay stage i and assume that j_{i-1} nodes failed the reception of the packet at stage $i-1$. The conditional number of transmissions that take place at stage i is equal to the number of nodes that received the transmission correctly, and its average over the distribution of the successful nodes at stage

i can be found as

$$u_{i|j_{i-1}\cdots j_1} = \sum_{j_i=0}^{N_i} \overline{\mathcal{B}}\big(N_i, p^{N_{i-1}-j_{i-1}}, j_i\big)(N_i - j_i) = N_i(1 - p^{N_{i-1}-j_{i-1}}) \ , \tag{5.17}$$

where the latter equality is valid in non-degenerate cases. Eq. (5.17) must be averaged over $j_{i-1}, j_{i-2}, \dots, j_1$ to yield the unconditional average number of transmissions at the ith relay stage. Call this quantity u_i. To compute the average of (5.17), we can replicate the structure of (5.12) by first defining the column vectors

$$\mathbf{t}_i = \begin{bmatrix} N_i & N_i{-}1 & \cdots & 1 & 0 \end{bmatrix}^T . \tag{5.18}$$

At this point, $u_i = \mathbf{p}_S^T\big(\prod_{j=0}^{i-1}\mathbf{P}_j\big)\mathbf{t}_i$, and the average total number of transmissions is found as

$$T_{\text{avg}} = 1 + \sum_{i=1}^{r} u_i \ . \tag{5.19}$$

We remark that $u_1 = N_1(1-p)$ and $u_2 = N_2\big(1-(p+(1-p)p)^{N_1}\big)$, after which no closed-form expression can be obtained for the remaining u_is.

If we assume that each stage hosts the same number of relays N, then $\mathbf{P}_i = \mathbf{P}\ \forall i$, where $\mathbf{P}$ is an $N \times N$ matrix, and we can simplify (5.12) to yield

$$q_r = \mathbf{p}_S^T\mathbf{P}^{r-1}\mathbf{p}_D \ , \tag{5.20}$$

and $T_{\text{avg}} = 1 + \sum_{i=1}^{r}\mathbf{p}_S^T\mathbf{P}^{i-1}\mathbf{t}_i$.

5.3.2 Distributed Flooding Policies

In many cases, typically if the link error probability is sufficiently low, the multiplicity of the flooding process is excessive, and leads to several packet replicas being uselessly transmitted, without noticeably increasing the success probability at the destination. It is therefore appropriate to design practical flooding policies that attempt to reduce this number of transmissions, by having each node decide locally if it should retransmit a given packet or not, without explicit coordination mechanisms.

We achieve this by allowing the nodes to overhear nearby traffic to understand how many relays have retransmitted a given packet at each stage. We design two policies for this purpose:

the first is named Predetermined Restriction (PR), the second Adaptive Restriction (AR). PR is aimed at enforcing a maximum number of transmitters per stage. Conversely, the objective of AR is to keep the flooding process running by adapting to the network conditions: if many successful retransmissions are detected, fewer nodes will tend to transmit; if overheard traffic suggests that a given packet is not being correctly advanced to subsequent stages, more nodes will act to support the flooding process.

For all policies, we assume the presence of a MAC protocol that avoids the collision between retransmissions by different relays. In practice, this can be achieved with very high probability by having relays back off before performing a retransmission, where the length of the backoff interval is drawn within a sufficiently long backoff window, or by loosely synchronizing subsequent relay stages in order to make transmission pipelining possible [73]. To fix ideas, in the following we assume the case of a backoff-based MAC protocol. We also assume that each node listens to the channel during the backoff period and, when its own backoff timer expires, the decision to transmit or not is made based on the policies described in the following subsections.

5.3.2.1 Flooding with Predetermined Restriction (PR)

The PR flooding policy prescribes that a node at a given stage should avoid retransmitting a packet if it overhears at least one retransmission of the same packet by any other node at the same stage, and the SNR of this transmission is greater than a prescribed value θ_r. Therefore, we can approach some desired average maximum number of relays per stage, $\hat{N}^{\text{res}}$, by varying θ_r.

To formalize the above ideas, assume, without loss of generality, that the nodes are ordered and numbered increasingly in subsequent relay stages, i.e., those of the first stage are numbered from 1 to S_1, those of the second stage from $S_1 + 1$ to S_2, and those of the ith stage from $S_{i-1} + 1$ to S_i, where $S_k = \sum_{\ell=1}^{k} N_\ell$. As a packet is forwarded from stage $i-1$ to stage i, the nodes at stage i start entering backoff periods to relay the packet, and keep listening for retransmissions in the meantime. As a node, say j, is ready to retransmit a packet from stage i, it will check if there exists at least one additional relay $S_{i-1} + 1 \leqslant k \leqslant S_i$, for which $\gamma_{kj} \geqslant \theta_r$, where γ_{kj} is the SNR of node k's transmission as measured by node j. In this case the node will refrain from transmitting, otherwise it will relay the packet to the next stage.

We note that, by increasing θ_r, fewer overheard retransmissions will meet the SNR threshold; in turn, the probability that a node is silenced decreases, and the average number of relays per stage increases. Conversely, by decreasing θ_r, a node can be silenced on average by a larger number of retransmitted packets, hence the probability that the node will refrain from retransmitting increases, which in turn will limit the average number of relays per stage, like setting a value for N^{res} as in Section 5.3.3. For example, in our scenario (see Section 5.3), setting $\theta_r \approx 21$ dB corresponds to $N^{\text{res}} = 2$, whereas $\theta_r \approx 40$ dB corresponds to $N^{\text{res}} = 8$.

5.3.2.2 Flooding with Adaptive Restriction (AR)

The AR policy is designed to achieve a high probability of success while at the same time reducing the overhead of the full flooding mechanism. Focusing on a given packet with sequence number n, the AR policy proceeds over the following two steps.

Step 1: perform a relaying step according to the PR policy. At a generic stage i, this will result in some nodes having relayed the packets, whereas other nodes will have refrained from doing so according to the rules of the PR policy (see Section 5.3.2.1). Call $C_i^{(s)}(n)$ the set of the nodes silenced at stage i: we note that the set can vary for different packets, hence for different values of n. All nodes in $C_i^{(s)}(n)$ remain in a listening state and attempt to overhear the nodes at stage $i+1$ forward packet n further on to stage $i+2$.

Step 2: Any node that hears n being transmitted at least once from stage $i+1$ to stage $i+2$ exits the flooding process for packet n permanently. Conversely, the nodes that could not hear n being forwarded from stage $i+1$ to stage $i+2$ assume that the forwarding process might have been unsuccessful. Therefore, they exit silence and relay packet n from stage i to stage $i+1$. We note that this further transmission step may result in additional nodes that receive the packet successfully at stage $i+1$. In the meantime these nodes will have overheard other transmissions in their neighborhoods, and will also decide to transmit or not according to steps 1 and 2 of the AR policy. However, in no event will a node forward a received packet more than once. Therefore, the number of transmissions carried out under the AR policy will be always less than or equal to that of unrestricted flooding. Both the PR and AR policies work based on the idea that forwarding operations could take advantage of some inherent MAC-level mechanisms that are typically present in MAC protocols for wireless networks, including idle listening and backoff. Instead of restricting such operations to the role of informing the MAC logic, we leverage on them to optimize the flooding performance by achieving a good tradeoff between the probability of decoding success at D and the number of transmissions. Further insight on this aspect is provided in our performance evaluation in Section 5.4.

5.3.3 Restricted flooding model

Both the PR and the AR policies act by reducing the total number of relays per stage, either in a fixed (PR) or in an adaptive fashion (AR). An analytical framework to assess the performance of these policies can be provided by extending the model of Section 5.3.1 to include restricted flooding. Restricted flooding is generally defined as a flooding policy where some nodes refrain from forwarding a packet even though they received it correctly. In this work, we model restricted flooding as a limit N^{res} on the maximum number of nodes that relay a packet at each stage.[3] To simplify the analysis, we assume that the number of successful nodes is known to all relays of a given stage, and that some form of arbitration takes place, in order to make sure that up to N^{res} relays transmit. This assumption will be relaxed later in the design of our heuristic policies.

Under the assumptions above, the analysis carried out so far offers a straightforward way to model restricted flooding. Assume that we have the same number of relays at each stage, which is the case, e.g., in (5.20): the analysis can be easily extended to the more general case of Eq. (5.12). Restricting the maximum number of relays to N^{res} means that whenever more than N^{res} relays correctly receive a packet at a given stage, only N^{res} of them will actually retransmit. This can be modeled by setting the first $N - N^{\text{res}} + 1$ lines of matrix $\mathbf{P}$ in (5.20) to be equal to the vector

$$\mathbf{p}^{\text{res}} = [\ p_{(N-N^{\text{res}})0} \quad p_{(N-N^{\text{res}})1} \quad \cdots \quad p_{(N-N^{\text{res}})N}\ ]^T \tag{5.21}$$

---

[3] Note that, in this model, N^{res} denotes the (exact) maximum number of nodes that will forward a correctly received packet at any stage, as opposed to the *average* number of nodes allowed to forward a packet at each stage in the PR and AR policies, which was denoted as $\hat{N}^{\text{res}}$ in Section 5.3.2.

to yield

$$\mathbf{P}^{\mathrm{res}} = \begin{bmatrix} & (\mathbf{p}^{\mathrm{res}})^T & \\ & \vdots & \\ & (\mathbf{p}^{\mathrm{res}})^T & \\ p_{(N-N^{\mathrm{res}}+1)0} & \cdots & p_{(N-N^{\mathrm{res}}+1)N} \\ \vdots & \ddots & \vdots \\ p_{N0} & \cdots & p_{NN} \end{bmatrix} . \tag{5.22}$$

In the same vein, $\mathbf{p}_D$ becomes

$$\mathbf{p}_D^{\mathrm{res}} = [\ p_{D,N-N^{\mathrm{res}}} \ \cdots \ p_{D,N-N^{\mathrm{res}}} \ \ p_{D,N-N^{\mathrm{res}}+1} \ \cdots \ p_{D,N}\ ]^T , \tag{5.23}$$

so that $q_r^{\mathrm{res}} = \mathbf{p}_S^T \left(\mathbf{P}^{\mathrm{res}}\right)^{r-1} \mathbf{p}_D^{\mathrm{res}}$, and the probability that D fails to decode the fountain code is $P_{\mathrm{fail}}(q_r^{\mathrm{res}})$. To compute the average number of transmissions in the network, we first define the column vector

$$\mathbf{t}_D^{\mathrm{res}} = \left[N-N^{\mathrm{res}} \ \cdots \ N-N^{\mathrm{res}} \ \ N-N^{\mathrm{res}}-1 \ \cdots \ 0\right]^T , \tag{5.24}$$

and finally the average number of transmissions is found as

$$T_{\mathrm{avg}}^{\mathrm{res}} = 1 + \sum_{i=1}^{r} \mathbf{p}_S^T \left(\mathbf{P}^{\mathrm{res}}\right)^{i-1} \mathbf{t}_D^{\mathrm{res}} . \tag{5.25}$$

5.3.4 Selecting the Galois field size η

Before proceeding to present analysis and simulation results in Section 5.4, it is useful to choose the size of the Galois field over which the flooded fountain code is designed. We do this by evaluating the interplay between the Galois field size and $1 - P_{\mathrm{fail}}$. In fact, a typical fountain code packet contains an encoding vector that defines how many packets have been linearly combined, and with which coefficients. Designing the fountain code over a small Galois field (at the minimum, $\mathbb{F}_2$) keeps the encoding vector short at one bit per encoded packet; however, in this case Eq. (5.1) dictates that the probability of successful decoding at the destination becomes lower for a fixed number of redundancy packets E. Conversely, a large Galois field (e.g., $\mathbb{F}_{256}$ as typically assumed in rateless code design) provides a higher probability of success for equal E, but at the same time it increases the size of each coded

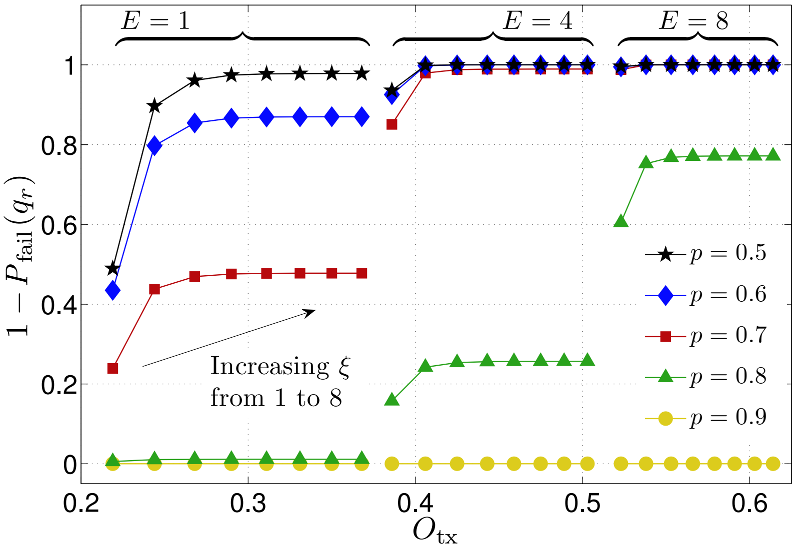

Fig. 5.2. $1 - P_{\text{fail}}(q_r)$ *vs.* O_{tx} *for different values of* ξ, E *and of the probability of link error* p*. For each value of* p*, a set of markers of the same color is shown; the set is spanned from bottom-left to top-right by increasing* ξ *from 1 to 8. The results show no significant improvement in* $1 - P_{\text{fail}}(q_r)$ *for* $\xi \geqslant 4$.

packet. Given that we consider the flooding of fountain codes through a network, where each packet is expected to be retransmitted several times by different nodes at each hop, we need to find a good tradeoff between the overhead yielded by the encoding vector size, and that yielded by the number of extra packets.

Assume that each packet has the following structure: a fixed overhead that represents generic protocol information (e.g., Time-To-Live and version fields, flags, etc.) of size α_{fix}; a sequence number that distinguishes different sequences of m source packets, of size α_{seq}; and an encoding vector of $m \times \xi$ bits, where $\xi = \log_2 \eta$, and η is defined in (5.1); finally, a payload of L bits. We calculate the transmission overhead affecting a sequence of m source packets as

$$O_{\text{tx}} = \frac{(m+E)(\alpha_{\text{fix}} + \alpha_{\text{seq}} + m\xi) + E \cdot L}{(m+E)(\alpha_{\text{fix}} + \alpha_{\text{seq}} + m\xi + L)} \,. \tag{5.26}$$

Fig. 5.2 shows a plot of $1 - P_{\text{fail}}(q_r)$ vs. O_{tx} for different values of ξ, E and $1 - p$. The packet parameters are $\alpha_{\text{fix}} = 24$ bits, $\alpha_{\text{seq}} = 8$ bits, $L = 256$ bits. For each value of $1 - p$, a set of markers of the same color is shown; the set is spanned from bottom-left to top-right by increasing ξ from 1 to 8. Three sets of curves are shown, respectively for $E = 1$, $E = 4$ and $E = 8$: a curly brace encompasses the horizontal span of each set. As expected, the probability of success increases with increasing ξ, and with increasing E for fixed ξ.

However, both improvements come at the cost of an increase in O_{tx}. While the increase of O_{tx} with E is expected and depends on the configuration of the code, the value of ξ can still be optimized. In particular, we observe that $1 - P_{\text{fail}}(q_r)$ remains almost constant after $\xi = 4$. Therefore, we will fix this value throughout the rest of this work.

5.4 Numerical results

5.4.1 Simulation scenario

We now present some results related to the flooding of fountain codes through a multihop network. We will show that our analytical model, albeit based on a simple uniform Bernoulli error process, matches simulations based on a more realistic Rayleigh fading propagation model with good accuracy, provided that an appropriate map between the error probability p of the Bernoulli model and the average probability of error in the Rayleigh fading scenario is used.

The simulations are set up as follows. We represent each relay stage as a set of nodes whose position is drawn at random within a cell of size 20 m × 60 m, where the centers of any two nearest cells are $\Delta = 60$ m apart. This scenario makes it possible to test that our analysis is still valid even after removing one of the assumptions of Section 5.3, namely that any two nodes belonging to subsequent relay stages are always connected. The source S and the destination D are located, respectively, at the center of the first and last cells. We assume that all nodes transmit with a power P_T. Therefore, if two nodes i and j are located at a distance d_{ij}, the Signal-to-Noise Ratio (SNR) γ_{ij} of the link between the two nodes can be computed as

$$\gamma_{ij} = \frac{\rho_{ij}\, P_T\, d_{ij}^{-k}}{P_N} \tag{5.27}$$

where k is the path-loss exponent, ρ_{ij} is an exponentially distributed power fading coefficient of average value 1 (which corresponds to a Rayleigh fading amplitude) representing the fading realization on the link from i to j, and P_N is the noise power. We assume that a transmission from i to j is successful only if $\gamma_{ij} \geqslant \theta$, where θ is a minimum SNR threshold.

In order to compare the Rayleigh fading simulations with the Bernoulli link error model presented in Section 5.3, we map the parameter p of the Bernoulli model to the average error probability computed via a stochastic geometry argument. Consider two nodes i and j, where

i belongs to stage ℓ, $\ell = 1, 2, \dots, r-1$ and j to stage $\ell + 1$. Without loss of generality, let $\ell = 1$, i.e., focus on the first and second relay stages. Call (x_i, y_i) and (x_j, y_j) the coordinates of nodes i and j, and assume that they take values in $[0, X_i]$, $[0, Y_i]$, $[\Delta, \Delta + X_j]$, $[0, Y_j]$, respectively. In our simulation scenario, $X_i = X_j = 20$ m, $Y_i = Y_j = 60$ m and $\Delta = 60$ m. The map is obtained as follows:

$$p = \int_0^{X_i} \mathrm{d}x_i \int_0^{Y_i} \mathrm{d}y_i \int_\Delta^{\Delta + X_j} \mathrm{d}x_j \int_0^{Y_j} \mathrm{d}y_j \int_0^{+\infty} \mathrm{d}\rho_{ij} \, \frac{\mathrm{e}^{-\rho_{ij}}}{A_i \, A_j} \, \mathbb{1}\left[\frac{\rho_{ij} \, P_T \, d_{ij}^{-k}}{P_N} < \theta \right] , \tag{5.28}$$

where $A_i = X_i Y_i$, $A_j = X_j Y_j$, $\mathrm{e}^{-\rho}/(A_i \, A_j)$ is the joint distribution of fading and of the locations of nodes i and j, and $\mathbb{1}[\,\cdot\,]$ denotes the indicator function, which returns 1 whenever the argument is true. We remark that (5.28) averages the value of the indicator function over all random parameters that concur to the computation of the link error probability. Such an average is by definition the probability of a Bernoulli event [100, Section 3.1], which is fully in line with our Bernoulli link error probability model. In the following, we set $k = 2$, $P_T = 34$ dBm and $P_N = -17$ dBm, we consider the network to be composed of $r = 5$ relay stages, and we assume that there are exactly 10 nodes in any stage, i.e., $N_1 = \cdots = N_r = 10$. Different probabilities of success are obtained by varying θ between 0 and 20 dB. All results are averaged over 2000 random draws of the nodes' positions, and over 10 different fading realizations for each position draw.

In the next subsection, we will discuss the performance of flooded fountain codes in the presence of both full and restricted flooding. In Subsection 5.5.1, we show the flexibility of our formulation by extending the analysis to the case of a "helper" node injecting additional redundancy in the network, and discuss the optimal placement of that node. In Section 5.3.2, we discuss practical policies for the distributed realization of full as well as restricted flooding.

5.4.2 Performance of fountain codes under unrestricted flooding

We start from Fig. 5.3, which shows $1 - P_{\text{fail}}(q_r)$ vs. p in the case of unrestricted flooding, by comparing our analytical model against Rayleigh fading simulation outcomes, for two different values of E. We observe that the probability of decoding success at the destination for $E = 8$ remains practically equal to 1 for $p \leqslant 0.75$, after which it sharply falls and becomes practically 0 for $p \geqslant 0.85$ due to the excessive number of forwarding errors. The anticipated effect of a lower number of redundancy packets, for $E = 2$, is that the probability of success

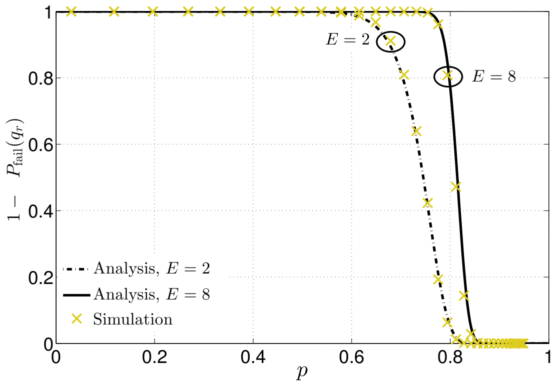

Fig. 5.3. *Success probability* $1-P_{\text{fail}}(q_r)$ *vs.* p *for* $m = 10$, $r = 5$ *and* $N = 10$, *for varying number of redundancy packets* E. *The probability that the destination can decode the fountain code, mainly driven by* E, *dominates the performance for sufficiently low values of* p.

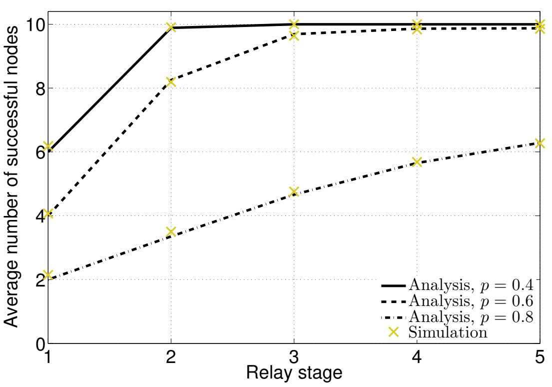

Fig. 5.4. *Average number of nodes that received a given packet correctly as a function of the relay stage (from 1 to 5) for different values of* p. *For sufficiently high link error probability* ($p = 0.8$), *in every stage there exist some nodes that fail to receive a given packet.*

is practically 1 only for $p \leqslant 0.5$, and the transition to 0 is also smoother. In both cases, Fig. 5.3 shows a very good agreement between the simulations and the analytical model.

A second metric of interest for characterizing the network behavior is the average number of nodes that successfully receive a given packet at each relay stage for different values of p. This metric is depicted in Fig. 5.4. The results show that the number of successful nodes at each stage increases for decreasing p and that, even for moderately high values of p (up to 0.6 in Fig. 5.4), practically all $N_r = 10$ nodes in the last relay stage have correctly received the packet. Conversely, $p = 0.8$ reduces the number of correct transmissions, and it is difficult for the flooding process to involve several nodes before the packet traverses the last relay stage. In all cases, the simulations match the analysis very well.

One of the main conclusions from the previous results is that $1 - P_{\text{fail}}(q_r)$ remains equal to 1 even for significantly high values of p. This further motivates the design of practical policies that achieve the probability of success of unrestricted flooding, while requiring fewer transmissions. The following section presents the performance evaluation of the PR and AR policies, introduced in Section 5.3.2.1 and 5.3.2.2.

5.4.3 Performance of the PR and AR policies

The behavior of the PR and AR policies is simulated by assuming the Rayleigh fading propagation model discussed in Section 5.4.1, and compared to the analytical model for

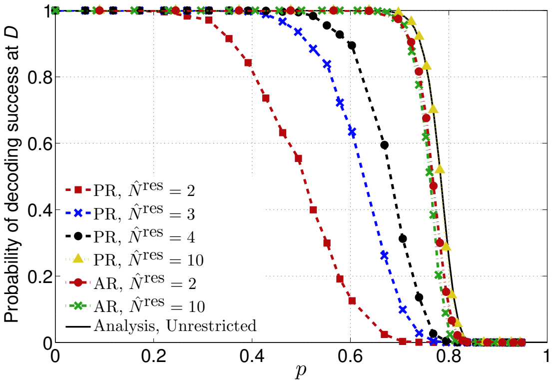

Fig. 5.5. *Probability of decoding success at D as a function of p for different combinations of $\hat{N}^{\text{res}}$ for the PR and the AR policies, compared to the analysis for unrestricted flooding.*

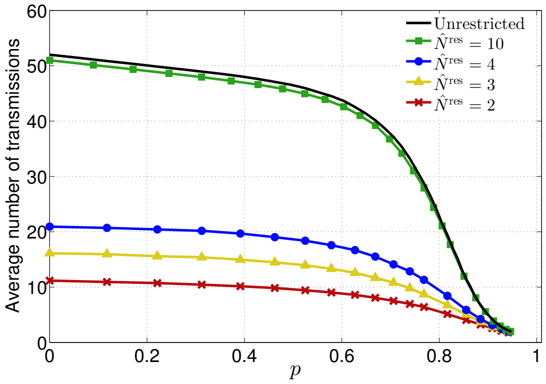

Fig. 5.6. *Average number of transmissions as a function of p for the PR policy for different values of $\hat{N}^{\text{res}}$, compared to the analysis for unrestricted flooding..*

unrestricted flooding via the map for p in (5.28).

Fig. 5.5 shows a plot of the probability of correct decoding at D against p obtained via the simulation of the PR and AR policies, compared to the unrestricted flooding analysis. For the latter, this is equivalent to $1-P_{\text{fail}}(q_r)$. The PR and AR curves are plotted for different values of $\hat{N}^{\text{res}}$. There are $N = 10$ nodes in all relay stages, and the number of redundancy packets transmitted by the source S is fixed to $E = 4$. Considering the PR policy, we observe that increasing the $\hat{N}^{\text{res}}$ parameter (by choosing θ_r as explained in Section 5.3.2.1) correspondingly increases the number of relays that are allowed to re-forward a packet at each stage. This progressively improves the probability of success as a function of p, until it finally matches that of unrestricted flooding for $\hat{N}^{\text{res}} = 10$. Contrary to the PR policy, the AR policy is adaptive. Posing restrictions on the number of forwarders via $\hat{N}^{\text{res}}$ makes the policy react by allowing more silenced relays to retransmit, upon failing to hear relays at the next stage re-forward a packet. Therefore, as we observe from Fig. 5.5, the success probability of the AR policy is always comparable to that of unrestricted flooding. This proves the effectiveness of the policy at compensating for forwarding errors.

It is interesting to evaluate the behavior of the PR and AR policies in terms of the number of transmissions performed per each packet carried through the network. This metric is depicted in Fig. 5.6 for PR, and shows the expected behavior that a higher value of $\hat{N}^{\text{res}}$ results in more relays being activated per stage, hence in more transmissions. For the same configurations shown in Fig. 5.6, we observe a steady increase in the number of transmissions. For $\hat{N}^{\text{res}} = 10$, PR transmits practically as many packets as unrestricted flooding. A different behavior is observed for the average number of transmissions per packet of the AR policy, shown in Fig. 5.7 as a function of p. When $p < 0.2$, the low probability of error makes the choice of $\hat{N}^{\mathrm{res}} = 4$ very inefficient, as in this condition it is highly likely that the nodes will hear the transmissions of their neighbors, refrain from transmitting, and hear the packet be correctly retransmitted by the nodes at the next stage. Instead, only a few relays will suffice to correctly convey a packet to D. In this respect choosing, e.g., $\hat{N}^{\mathrm{res}} = 2$ would achieve very good probability of success with about half the transmissions compared to the case $\hat{N}^{\mathrm{res}} = 4$. The ranking among the curves is progressively inverted when p increases, as in this case more relays are needed to guarantee success. For example, for $p = 0.5$, $\hat{N}^{\mathrm{res}} = 4$ yields fewer transmissions than even $\hat{N}^{\mathrm{res}} = 1$ (which is too restrictive, and often causes several silenced nodes to transmit, after they fail to hear the relays at the next stage forward the packets further on). Fig. 5.7 also suggests that $\hat{N}^{\mathrm{res}} = 2$ and $\hat{N}^{\mathrm{res}} = 3$ constitute good choices for almost all values of p, except perhaps if $p < 0.2$. Finally, note that the AR policy performs similar to unrestricted flooding in terms of probability of successful decoding at D (Fig. 5.5), but requires much fewer transmissions to attain this, as confirmed by Fig. 5.7.

5.4.4 Insight from the restricted flooding model

The restricted flooding model described in Section 5.3.3 offers a good means to understand the behavior of the PR and AR policy. In Fig. 5.8, we consider the interplay between the probability of decoding success at D and the number of transmissions performed in the whole network. The graph shows a comparison between restricted flooding (lines) and the PR policy (markers), for $p = 0.5$ and for different values of N^{res}, $\hat{N}^{\mathrm{res}}$ and E. The restricted flooding curves are spanned from left to right by increasing N^{res}, the PR curves by increasing $\hat{N}^{\mathrm{res}}$: in both cases the effect is to increase the probability of decoding success at the price of an increase in the number of transmissions. The figure marks four choices of the $(\hat{N}^{\mathrm{res}}, E)$ pairs that achieve a probability of success of at least 0.9. We observe that the PR policy is an effective distributed implementation of restricted flooding, and that it helps achieve the same performance in terms of number of transmissions, at the price of a very small decrease in the success probability, mainly due to the distributed implementation. In any event, the mismatch becomes negligible for $\hat{N}^{\mathrm{res}} \geqslant 3$.

Fig. 5.9 compares the performance of the AR policy against the analysis of restricted

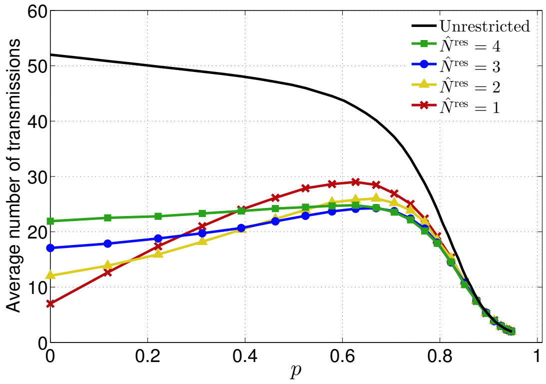

Fig. 5.7. *Average number of transmissions for a single packet carried out by the AR policy as a function of* p *for different values of* $\hat{N}^{\text{res}}$, *compared to the number of transmissions of unrestricted flooding.*

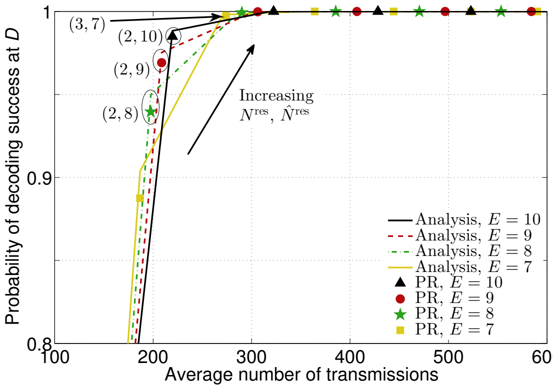

Fig. 5.8. *Interplay between the probability of decoding success at* D *and the average number of transmissions achieved by the PR policy, compared to restricted flooding, for* $p = 0.5$ *and for different values of* N^{res}, $\hat{N}^{\text{res}}$, *and* E. *Curves are spanned from left to right by increasing* N^{res} *(for restricted flooding) or* $\hat{N}^{\text{res}}$ *(for the PR policy).*

flooding for $p = 0.5$ and for different values of N^{res}, $\hat{N}^{\text{res}}$ and E. The plot suggests that the AR policy behavior typically leads to an increased number of transmissions. However, this is compensated by the advantage of a higher probability of success. This is especially the case for low values of E, which implies that the typical overhead achieved by the AR policy is lower (see also Fig. 5.2). Once the probability of success has achieved a value of about 1, a further increase of $\hat{N}^{\text{res}}$ yields a negligible improvement and, as expected, causes the number of transmissions to increase. In any event, the increase is limited with respect to plain restricted flooding, as the AR policy achieves a very high probability of success already for low values of E. For example, for $E = 2$ the AR policy already attains a probability of success close to 1, and requires fewer transmissions than plain restricted flooding would require to achieve the same result.

5.5 Extension: Helper nodes

The study presented in this work considers fountain-coded data flooded through a network up to a given destination, and provides fundamental insights on the interplay between the probability of success and the number of transmissions required to achieve it. In the process, two policies are proposed to optimize the flooding process by reducing the total number of transmissions without decreasing the probability of success. The study lends itself to several

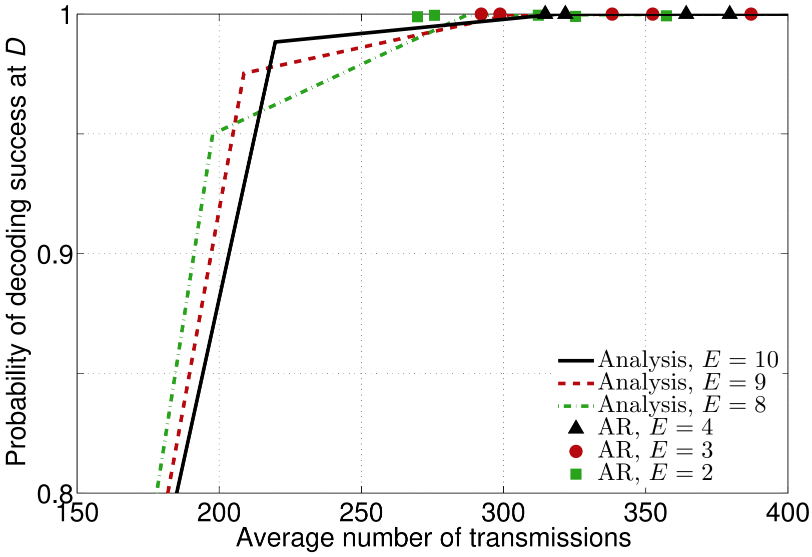

Fig. 5.9. *Interplay between the probability of decoding success at* D *and the average number of transmissions achieved by the AR policy, compared to restricted flooding, for* $p = 0.5$ *and for different values of* N^{res}*,* $\hat{N}^{\text{res}}$*, and* E*. Curves are spanned from left to right by increasing* N^{res} *(for restricted flooding).*

extensions, in terms of both functionality and modeling. In the following subsection, we extend the study by allowing relays that successfully decoded the fountain-coded packets from the source to become "helper nodes," and inject additional redundancy. In general, this redundancy will be different from that sent by the source, and thus will help the destination decode the source packets.

5.5.1 Additional redundancy from intermediate "helper" nodes

In the following, we assume that there exists one node that received enough packets to decode the fountain code and reconstruct the original information sent by the source S. This node becomes able to inject additional redundancy packets, different from those generated by S, in order to favor the correct decoding of the m source packets at D. We call this node a helper node, and refer to it via the subscript h. We will now elaborate on the performance of the network in the presence of a helper, we provide design guidelines for where this node should be placed, and discuss the utility yielded by the presence of additional helper nodes.

We start by noting that node h must belong to one of the relay stages between the source S and the destination D. In particular, we assume that out of the r stages that separate S from D, there are u stages between S and h, and v stages between h and D, so that

$r = u + v + 1$. After successful decoding, h transmits E_h additional redundancy packets, in order to increase the probability of successful decoding at D. For this analysis, we assume that only one helper node exists in the network.

We start from the probabilities q_u that a packet fails to reach node h after u relay stages, and q_v that D fails to receive an extra redundancy packet transmitted by h; using the same formulation of (5.12), we can write

$$q_u = \mathbf{p}_S^T \left(\prod_{i=1}^{u} \mathbf{P}_i \right) \mathbf{p}_D \ , \quad q_v = \mathbf{p}_S^T \left(\prod_{i=1}^{v} \mathbf{P}_i \right) \mathbf{p}_D \ . \tag{5.29}$$

The probability of decoding failure at the relay stage of the helper node is found via q_u as

$$P_{\text{fail}}(q_u) \leqslant \sum_{j=0}^{m} \binom{M}{j} q_u^{M-j}(1-q_u)^j + \sum_{j=m+1}^{M} \binom{M}{j} q_u^{M-j}(1-q_u)^j \, \eta^{-(j-m)} \ . \tag{5.30}$$

The probability that at least one node successfully decodes the fountain code (out of N nodes belonging to the relay stage of node h) is $1 - P_{\text{fail}}(q_u)^N$. Now, a fountain decoding failure occurs at D in one of the following two cases: *i)* h fails to decode the fountain code and thus cannot help: in this case, D may fail to decode after the M packets transmitted by S according to the same arguments in Section 5.3.1; *ii)* h successfully decodes the fountain-coded packets received from S and floods E_h extra redundancy packets towards D, but D still fails to decode. The total probability of failure is finally upper-bounded by

$$\left(1 - P_{\text{fail}}(q_u)^{N_{u+1}}\right) P_{\text{fail}}^{\text{hlp}}(q_r, q_v) + P_{\text{fail}}(q_u)^{N_{u+1}} P_{\text{fail}}(q_r), \tag{5.31}$$

where $P_{\text{fail}}^{\text{hlp}}(q_r, q_v)$ is the probability of fountain decoding failure at D, given the probability q_r that D fails the reception of a packet from S, and the probability q_v that D fails a reception from h. As in (5.7), we distinguish between two cases, namely that D receives up to m packets and that D receives more than m packets. In both cases, the packets may either come only from S or be a combination of packets from S and h. We have:

$$\begin{aligned} P_{\text{fail}}^{\text{hlp}}(q_r, q_v) \leqslant & \sum_{j=0}^{m} \sum_{k=0}^{\min\{E_h, m-j\}} \binom{M}{j} q_r^{M-j}(1-q_r)^j \binom{E_h}{k} q_v^{E_h-k}(1-q_v)^k \qquad (5.32) \\ & + \sum_{j=\max\{0,m+1-E_h\}}^{M} \sum_{k=\max\{0,m+1-j\}}^{E_h} \binom{M}{j} q_r^{M-j}(1-q_r)^j \binom{E_h}{k} q_v^{E_h-k}(1-q_v)^k \eta^{-j+k-m} \ . \end{aligned}$$

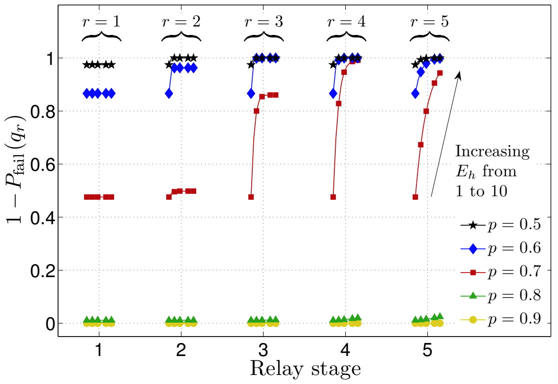

Fig. 5.10. *Analysis and simulation of* $1 - P_{\text{fail}}(q_r)$ *vs relay stage in the case* $N_i = N = 10\ \forall i$, *for* $m = 10$, $r = 5$, *and* $E = 1$, *for varying number of redundancy packets* E_h *from the extra transmitting node. For higher values of the link error probability* p, *a helper node located at stage* $r = 4$ *yields better probability of success, whereas for lower values of* p $r = 3$ *is a better choice.*

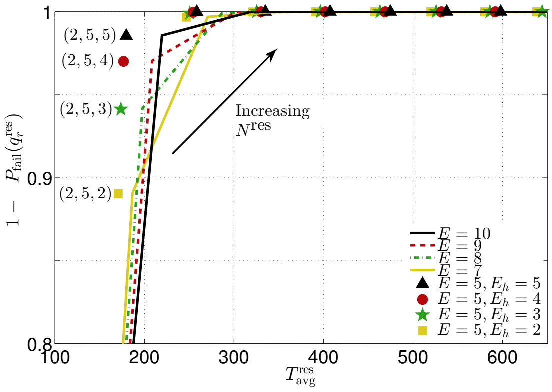

Fig. 5.11. *Interplay between* $1 - P_{\text{fail}}(q_r^{\text{res}})$ *and* $T_{\text{avg}}^{\text{res}}$ *for the case where only the source* S *transmits compared to the case with an extra helper node. For the former, we consider the analysis in Section 5.3.3 for different values of* E*; for the latter, we show simulation results for different* (N^{res}, E, E_h) *triples, for* $p = 0.5$*. Curves are spanned from left to right by increasing* N^{res}.

In the computation of the average number of transmissions $T_{\text{avg}}^{\text{hlp}}$ in the presence of h, we need to account for the extra effort paid by the network to forward h's packets. We have

$$T_{\text{avg}}^{\text{hlp}} = M T_{\text{avg}}(S \to D) + E_h T_{\text{avg}}(h \to D) \ , \tag{5.33}$$

where

$$T_{\text{avg}}(S \to D) = 1 + \sum_{i=1}^{r} \mathbf{p}_S^T \left(\prod_{j=0}^{i-1} \mathbf{P}_j \right) \mathbf{t}_i \ , \quad T_{\text{avg}}(h \to D) = 1 + \sum_{i=1}^{v} \mathbf{p}_S^T \left(\prod_{j=0}^{i-1} \mathbf{P}_j \right) \mathbf{t}_i \tag{5.34}$$

respectively represent the average number of transmissions from S to D and the average number of transmissions from h to D.

Fig. 5.10 plots the probability of successful decoding at the destination in the presence of a helper node, as a function of the relay stage where the helper is located, and for different values of the link error probability p, where $N = 10$, $r = 5$ and $E = 1$. We observe that the actual value of the success probability increases with the number of redundancy packets sent by h, E_h, and that the best performance for each value of p depends on the relay stage where h is located.

For example, consider the case of $p = 0.7$. Choosing to have h at stage 4 would make the success probability achieve a value of ≈1 for $E_h = 8$ packets. However, this is not the case if the helper node h is placed at any other relay stage, especially at the 1st or 5th stage. Similarly, for $p = 0.6$ and, say $E_h = 3$, having the helper node at stage 3 achieves a probability of success very close to 1, whereas a placement in any other stage would be less effective.

While the position of node h that maximizes the success probability depends also on N and E, the example above shows our point that h should not be located too close to S or D. The intuition behind the above result is that if h were too close to S, it would not benefit from the flooding of source packets and the higher decoding probability that results. Conversely, if h were too close to D, any sufficiently high probability of error would make the packet flooding process die out before reaching h. In turn, h would not be able to reconstruct the source information and inject new redundancy packets. Moreover, even if h could decode the message, extra redundancy packets would not benefit from the flooding process on the way to the destination.

Assuming that h is located at stage 4, in Fig. 5.11 we show the interplay between the probability of decoding success at D and the number of transmissions in the network in the presence of h. We consider restricted flooding (see Section 5.3.3) and plot different curves, each for a different value of E. All curves are spanned from left to right by increasing N^{res}. In addition to these curves, we plot a set of points described by the triplets (N^{res}, E, E_h), which correspond to the cases where a helper node is present. We observe that the inclusion of h helps reduce the average number of transmissions (where the saving is larger for higher values of E). In fact, the best probability of success for the same total redundancy $E + E_h$ would be achieved if S sent all the redundancy itself. However, this would also lead to the largest possible number of transmissions because of the flooding process. Having a helper node between S and D helps save considerable resources by reducing the total transmissions while not harming the probability of success. This demonstrates the feasibility of the helper node solution.

Having assessed the impact of one helper and discussed how to optimize its location in the network, we now consider the possibility of having either limited or unlimited helper nodes, and to allow each helper to inject either a limited or an unlimited number of redundancy packets.

More specifically, in the same simulation scenario considered so far, we allow any intermediate node that correctly received a sufficient number of packets from S to become a *potential* helper. When a helper injects additional redundancy packets, these may be received by intermediate nodes and thereby generate additional helpers in a sort of avalanche effect. Helpers act only if D failed to decode the fountain code via the source packets relayed by the network. In this case, we pick a helper node at random and let it transmit one redundancy packet each time. This packet will be flooded through the network between the helper and D and, if received by D, it will help decoding the fountain coded packets of S. The duration of this process and the set of helper nodes are regulated by the following four policies:

1) ***Limited*** *number of helpers,* ***limited*** *redundancy per helper*: nodes that decoded the fountain code using only the packets received from S can become helpers (no avalanche). Only one redundancy packet can be transmitted by each helper.

2) ***Unlimited*** *number of helpers,* ***limited*** *redundancy per helper*: any node that decoded the fountain code using any set of packets, including those sent by other helpers, can become a helper itself (avalanche allowed). Only one redundancy packet can be transmitted by each helper.

3) ***Limited*** *number of helpers,* ***Unlimited*** *redundancy per helper*: helper avalanche not allowed, helpers can transmit an unlimited number of redundancy packets, until the destination decodes the packet successfully.

4) ***Unlimited*** *number of helpers,* ***Unlimited*** *redundancy per helper*: helper avalanche allowed, helpers can transmit an unlimited number of redundancy packets.

We compare the four policies above with the case of a single helper node chosen at random in the network, and allowed to transmit unlimited redundancy packets, akin to the case discussed for Figs. 5.10 and 5.11.

Fig. 5.12 shows the probability of successful decoding at D for $N^{\text{res}} = 6$ and $E = 2$. The leftmost curve correspond to the case where only S transmits. Any helper node policy provides better performance. Specifically the policies allowing only limited redundancy per helper tend to fail for $0.7 \leqslant p \leqslant 0.8$, as transmission failures prevent D from receiving the helper packets; moreover it becomes increasing less likely that there are any helpers at all. Conversely, the policies that allow unlimited redundancy per helper achieve a very good performance. In particular the policy allowing unlimited helpers and redundancy per helper

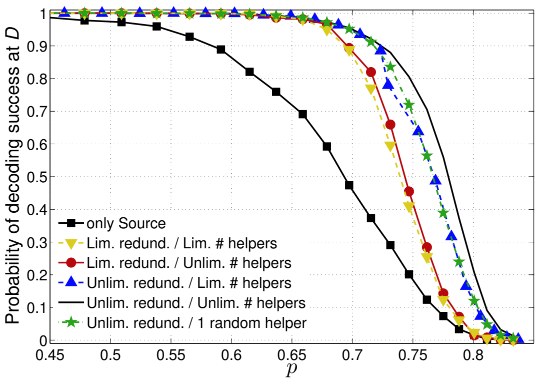

Fig. 5.12. *Probability of decoding success at D against p for different helper policies. The best performance is achieved when one or more helper nodes are allowed to send an unlimited number of redundancy packets.* $N^{\text{res}} = 6$, $E = 2$.

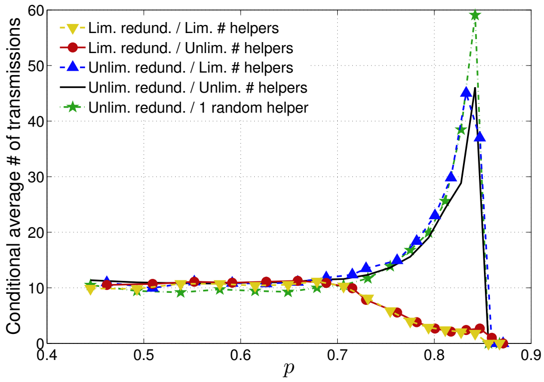

Fig. 5.13. *Conditional average number of transmissions per packet against p, for the only cases where helper nodes are needed to achieve correct decoding at D. The most aggressive helper policies experience failure only if p is sufficiently high to avoid that any helpers actually exist.* $N^{\text{res}} = 6$, $E = 2$.

only fails in those cases where no helper is present in the network.

The cost of the improvement in the probability of success is measured by the number of transmissions carried out to relay each packet in the network. For those cases where the helper nodes actually transmit, this metric is shown in Fig. 5.13. The policies with limited redundancy per helper progressively decrease their number of transmissions for $p \geqslant 0.7$, supporting the discussion above. Conversely, the policies allowing unlimited redundancy keep increasing the number of transmissions until p is sufficiently high to prevent the generation of helpers in the network. The total average number of transmissions in Fig. 5.14 (taken over both the cases where helpers act and those where they do not), further confirms the above discussion, and shows that the largest number of transmissions is reached for very high values of p, as expected.

We remark that a helper node injects redundancy packets only when it has successfully decoded the source packets. Removing the latter constraint would mean that the injected redundancy could only help recover the set of source packets originally encoded into the fountain-coded packets received by the helper. The evaluation of this case is left as a future extension.

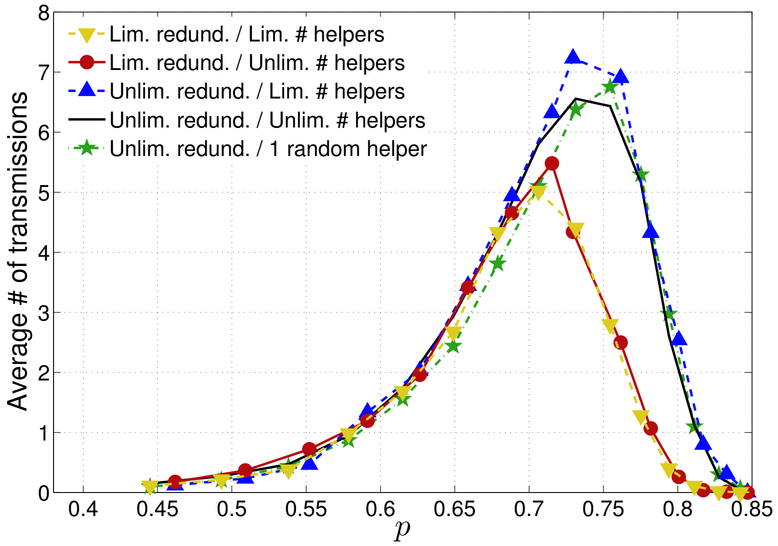

Fig. 5.14. *Average total number of transmissions per packet against* p*, including both the cases where helpers node are required and the cases where they are not.* $N^{\mathrm{res}} = 6$*,* $E = 2$*.*

5.6 Cross-Layer Optimization

In this work, we particularly focused on the routing layer and studied a trade-off between the required number of transmissions and the probability of successful decoding at the destination. However, MAC layer issues, particularly interference and the associated retransmission delay has not been addressed. As mentioned earlier, flooding based networks are more popular for delay tolerant networks, where the main goal is the tranmission reliability rather than the associated delay, and therefore our network layer based approach fulfills the requirements of such networks. However, we do point out that the cross-layer optimization, i.e., optimizing both reliability and delay at the same time would be an interesting extension to our work.

Another important aspect would be to investigate the performance of our proposed restricted flooding protocols jointly with the MAC protocol, i.e., analysis of transmission delay of predetermined and adaptive restriction policies considering a particular MAC layer protocol. This may also provide an interesting trade-off between energy efficiency and achievable throughput of the network. Moreover, a scenario where both adaptive and predetermined can be selected interchangeably at a particular relay stage to maximize energy efficiency or throughput could also be worth investigating.

Therefore, one possible future extension of our work would be the optimization of cross-layer design, i.e., network optimization considering both MAC and routing layer protocols.

5.7 Conclusions

In this chapter we analyzed a multi-hop flooding network, where the source is transmitting fountain coded packets towards the destination. We devised a probabilistic model to evaluate the probability of successful decoding as a function of the link error probability, the network topology, the number of forwarding nodes, and the amount of redundancy introduced in the network. We proposed two practical restricted flooding policies, i.e., 1) predetermined restriction which performs similar to our analysis of restricted flooding, and 2) adaptive restriction which performs similar to unrestricted flooding but saves energy by reducing the required number of transmissions. We further included helper nodes (any intermediate relay in the network) in our model which can decode and inject extra redundancy packets in the network. Results showed that the inclusion of helper nodes can further reduce the required number of transmissions.

5.8 Low complexity approximations to q_r

Since the computation of q_r in (5.5) involves several binomial coefficients, it is convenient to discuss a few ways to approximate it. First of all, observe that $\big(p^k+(1-p^k)p\big)^N > p^N$, where the lower bound is tighter for smaller p. Using this argument, the sum over j_r in (5.5) is lower-bounded by p^{N_r}, after which all other sums are identically 1. Hence a first approximation to (5.5) is the lower bound $q_r > p^{N_r}$. This is equivalent to assuming that the flooding process gets started with probability 1 (i.e., that at least one node at the first stage receives the packet correctly), that any given packet percolates through the network, and that all relays at stage r receive the packet correctly from any of the relays that correctly received it at stage $r-1$. In this case, the packet fails to reach D only if all nodes at stage r fail their transmission, an event which has probability p^{N_r}.

The approximation above is accurate only for small values of p. A better approximation is achieved by considering that at least one relay at stage 1 must receive the packet correctly for the flooding process to get started. This corresponds to the following lower bound

$$q_r > (1-p^{N_1})p^{N_r} + p^{N_1} \ , \tag{5.35}$$

where if all relays at stage 1 are erroneous (with probability p^{N_1}) the delivery of the packet to D fails with probability 1.

A further improvement of the approximation above requires to deal separately with the following three cases: a) no nodes, b) one node, or c) more than one node receives a packet correctly at a given stage. Call $f_i(p)$ the probability that the delivery of a given packet from relay stage i to destination D fails, and focus on the first relay stage, which includes N_1 relays. The probability that all nodes fail the reception of a packet transmitted from the source, case a), is p^{N_1}, and in this case $f_1(p) = 1$. Case b) occurs with probability $N_1 p^{N_1-1}(1-p)$, and in this case we compute the probability that the packet does not reach D depending on the outcome of the receptions at the second stage, i.e., $f_1(p) = f_2(p)$ in case b). Finally, in case c) we assume that if two or more nodes receive the packet correctly, then the flooding process will lead to a progressive increase of the number of relays that receive the packet correctly at each subsequent stage, until ultimately all N_r nodes at the rth stage will have a copy of the packet and will be able to transmit it to the destination. Therefore, this event occurs with

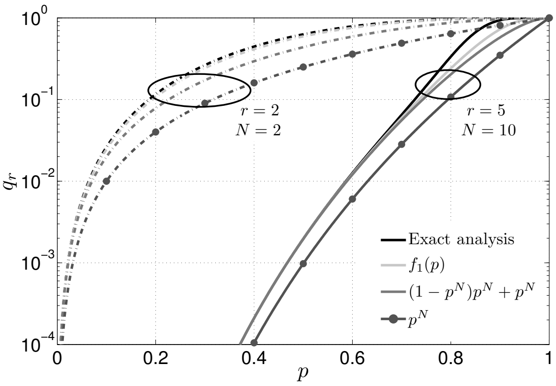

Fig. 5.15. *Approximations of* q_r *as a function of* p *for the case* $N_i = N$ $\forall i$. *All approximations improve for* $p \to 0$. *The approximation* $q_r \approx f_1(p)$ *achieves the best accuracy.*

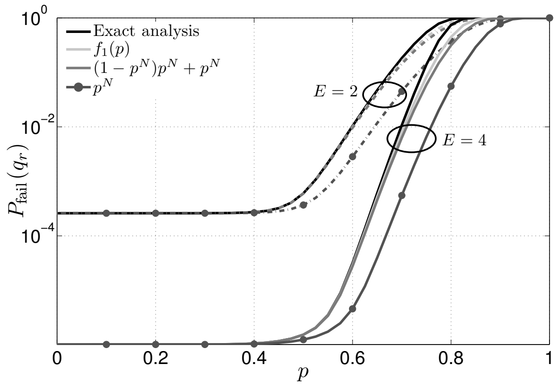

Fig. 5.16. *Analysis and simulation of* $P_{\text{fail}}(q_r)$ *as a function of* p *in the case* $N_i = N$ $\forall i$ *for* $m = 10$, $r = 5$ *and* $N = 10$, *for different numbers of redundancy packets* E. *The probability of that the destination can decode the fountain code, mainly driven by* E, *dominates the performance for sufficiently low values of* p.

probability $1 - p^{N_1} - N_1 p^{N_1-1}(1-p)$, and $f_1(p) = p^{N_r}$ in this case. In the generic case of stage i we have

$$f_i(p) = p^{N_i} + f_{i+1}(p) N_i p^{N_i-1}(1-p) + p^{N_r}\left(1 - p^{N_i} - N_i p^{N_i-1}(1-p)\right) , \tag{5.36}$$

and at the the final stage r we have

$$f_r(p) = p^{N_r} + p N_r p^{N_r-1}(1-p) + \xi\left(1 - p^{N_r} - N_r p^{N_r-1}(1-p)\right) , \tag{5.37}$$

where we note that the probability of error in case b) is exactly p, since the rth stage is the last one. The probability of error in case c), ξ, can be approximated by observing that if the most likely event is that two nodes received a packet correctly, the delivery of the packet to D fails if these two nodes fail their transmission to D, hence $\xi \approx p^2$. Finally, we can approximate the probability of error at D as $q_r \approx f_1(p)$. Note that $\xi \approx p^2$ does not provide a lower bound to the error probability. To obtain a lower bound, we must replicate the argument employed for case c) in $f_i(p)$, see (5.36), and set $\xi = p^{N_r}$. However, this results in a larger maximum deviation from the actual value of q_r.

In Fig. 5.15, we compare the accuracy of the three approximations proposed above by assuming that $N_i = N$ $\forall i$, and by considering two scenarios: $r = 2$, $N = 2$ and a second network with more hops but higher connectivity ($r = 5$, $N = 10$). All approximations except the looser p^N function achieve a satisfactory degree of accuracy for sufficiently low values of p. The meaning of "low" in this context depends on how likely it is for the flooding process to get started and involve an increasing number of relays at each stage, ultimately all relays at stage r. For both pairs of r and N, we observe that the simpler p^N function is substantially improved by the $(1 - p^N)p^N + p^N$ approximation, as the latter models more accurately the probability of error over the hop from the source to the first relay stage. Finally, the recursive approach involving $f_1(p)$ approximates q_r even better, and therefore represents a valid alternative to the exhaustive computation of q_r in (5.5).

Fig. 5.16 extends the analysis of the approximations q_r to the computation of the decoding probability $1 - P_{\text{fail}}(q_r)$, for $E = 2$ and $E = 4$. We set $m = 10$, $r = 5$ and $N = 10$. Also in this case, the $f_1(p)$ approximation to q_r makes it possible to correctly reproduce the probability of failure at D with good accuracy for $p \leqslant 0.7$, after which the deviation becomes larger. Higher values of p would be anyway of little interest for practical network analysis.

6

Conclusions

This thesis studied two main aspects of energy efficiency in wireless communication. Firstly, we studied and analyzed the energy efficiency of a receiver equipped with large antenna arrays, where we primarily focused on quantized millimeter wave receiver architectures and showed that there are regimes where digital combining can be an energy efficient option compared to other beamforming schemes. Secondly, we analyzed a multi-hop flooding network where the source transmits fountain coded packets towards the destination. To improve the energy efficiency of such networks, we investigated the minimum number of transmissions required to forward a packet successfully to the destination. We also studied a trade-off between the amount of redundancy and the number of transmissions while maintaining a certain required probability of success at the destination. We next summarize the main outcomes of these works.

6.1 Millimeter Wave Receiver Efficiency

We studied millimeter wave receiver efficiency from two main aspects (Chapters 2-4). Firstly, focusing on data plane communication, we investigated power, energy and spectral efficiency of analog, digital and hybrid beamforming schemes. Secondly, focusing on control plane communication, we studied the performance of ABF, DBF, HBF and PSN architectures in the context of directional initial cell discovery.

Results showed that contrary to general perception, DBF with low resolution ADCs and up to certain maximum bandwidth (suitable to support requirements of future communication) may achieves better power, energy and spectral efficiency than other beamforming schemes. The main highlights of these works are summarized as follows:

6.1.1 General Conclusions

We first investigated power consumption of analog, digital and hybrid beamforming millimeter wave receiver architectures. We showed there is a certain range of number of receive antennas, number of ADC bits and bandwidth upto which DBF results in a lower power consumption than HBF. Secondly, we investigated spectral and energy efficiency of ABF, DBF and HBF schemes with low resolution ADCs. Results showed that there is an optimal number of bits which maximizes energy efficiency. We also showed that in contrast to general perception regarding high energy consumption of a fully digital architecture, there are many use-cases where DBF achieves better spectral efficiency vs energy efficiency trade-off compared to ABF and HBF architectures. Thirdly, to further reduce the power consumption of a fully digital scheme, we proposed a receiver architecture with variable resolution ADCs. We showed that the proposed architecture can provide power savings in the range of 20-80%, depending on the operating SNR.

We then investigated the performance different beamforming architectures from an initial cell discovery perspective. The performance is analyzed based on probability of access error, delay, power consumption and energy consumption. Firstly, to reduce delay, we proposed a context information based cell search approach, where location information of the millimeter wave BS positioning is available to the MS. We showed that when accurate CI is available then ABF performs better than other schemes, however, the performance degrades in case of erroneous CI. To address the issue of erroneous CI, we proposed a low power PSN architecture, which combines and compares multiple beams (in analog) simultaneously but requires only a single RF chain. Secondly, we studied the impact of bandwidth, and showed that the delay and the energy consumption decrease whereas the power consumption increases with an increase in the bandwidth. We showed that ABF performs better when CI is available without delay, whereas in other scenarios (i.e., no CI or CI with delay) DBF results in a lower energy consumption and a lower delay than other schemes. However, results showed that DBF is power efficient only up to certain range of bandwidth, and therefore at very high bandwidths DBF may not be a feasible option for power constrained devices (such as the MS). Finally, to further reduce the initial cell search delay, we proposed a new signaling structure, where the control plane sub-carrier is composed of higher sub-carrier bandwidth compared to the data plane sub-carrier.

Detail outcomes of each chapter are summarized below:

6.1.2 Power, Spectral and Energy Efficiency

In Chapter 2, we argued that a comparison of different beamforming schemes solely based on ADC power consumption does not give full insight, and the results can be quite different when the total power consumption of the mmWave receiver is considered. Based on a detailed power consumption analysis, we showed that there are cases in which DC can still be a viable option compared to the other beamforming schemes. In particular, considering both a high power and a low power ADC for a mmWave receiver model, we showed that for a certain range of N_{ANT}, b and B, DC may result in a lower power consumption than HC. The results also showed that with a low number of antennas (e.g., as in case of the MS), DC power consumption may be comparable to (or only slightly higher than) AC power consumption. Moreover, with SNR as a figure of merit, we showed that the number of ADC bits for DC which results in a similar power consumption as in HC is large enough to avoid any loss in SNR.

We further studied the spectral and energy efficiency trade-off for analog, digital and hybrid combining schemes. Firstly, we showed that there is an optimal number of ADC bits which results in maximum energy efficiency. Moreover, the results show that, AC achieves the best EE only in low-SNR or low-rank channels. Whereas for any other mmWave channel and hardware scenario, DC and HC alternate depending critically on the system parameters input to the model and the preference of the receiver designer between maximizing EE or SE. There is a range of components that favor the usage of DC with a low number of bits, which does not seem to be accurately represented by the popular claim that DC always consumes more power, based only on ADC power consumption with a high number of bits.

The ideal scenario for AC is a receiver with very tight power constraints in a mmWave rank-1 channel. The former may be relevant for future low-power systems such as machine-type communications, but the latter may not be entirely plausible, considering that with a < 10 mm wavelength all small objects in the surroundings can become potentially relevant reflectors.

We have shown that the conventional wisdom that hybrid combining architectures are preferable over fully digital ones is not universally true. Rather, the mmWave channel parameters and component power consumption parameters, which can vary up to a couple of orders of magnitude between different papers in the literature, determine critically the relationship between DC with few bits and HC. In this work, we provided a comprehensive power-consumption comparison method and propose a performance chart technique that allows to choose between HC and DC depending on the given component parameters. We also provided a multi-objective optimization interpretation to list a collection of points in the chart that are optimal under different preference weights assigned by the system operator to the utility of SE and EE.

We have also shown that if the receiver has a smaller antenna array, or if a subset of the antennas may be turned off to save power, HC does not have an advantage over DC architectures. Particularly, we showed that in the Downlink where the receiver is equipped with a low number of antennas, DC achieves better performance than HC, whereas in the point to point Uplink, HC offers better EE while DC with fewer bits offers a balance between SE and EE.

Moreover, even with high power ADCs, we have shown that if the scattering environment is not-so-sparse (e.g., with $(N_c, N_p) = (Poisson(1.8), 20)$), the spatial multiplexing gains of DC in the Downlink achieve either a higher rate than HC (if the hybrid scheme uses fewer RF chains) or a similar rate with better energy efficiency (if the hybrid scheme increases the number of RF chains). We showed that there is an optimal number of RF chains which maximizes EE in HC, but this value changes with SNR and a single implementation of a receiver cannot have the optimal value for all distances occurring in a typical network. Recall that the accurate model has an average of 1.8 clusters with 20 paths per cluster, with ranks 1-10 being frequent, and thus the mmWave channel model is the best possible scenario for HC to compete against DC.

The analysis is also more in favor of HC in the uplink scenario, due to the higher number of receive antennas in the BS. However, even though HC achieves the highest EE, DC with fewer bits still achieves a good EE and a greater SE, potentially being preferable when both characteristics are equally desired by the system operator.

While in our work we focused on spectral and energy efficiency, size and cost are other important dimensions that need to be considered, although we may expect that these factors will become less important as technology becomes more mature and 5G mmWave devices are produced and deployed in very large numbers.

6.1.3 Variable Bits ADCs

In Chapter 3, we proposed variable resolution ADC in fully digital receiver architectures with large antenna arrays. This variable-resolution ADC approach can be seen as a generalization of antenna selection and other 1-bit ADC proposals.

We discussed models for a mmWave uplink scenario with a scalar transmitted signal and beamforming over a sparse scattering channel matrix where a single dominant eigenvalue is responsible for most of the energy transfer of the system. We have noted the usual power-consumption model for ADCs, and proposed a simple "first-attempt" type of two-level resolution ADC system design. We have also designed two algorithms to operate in our model, one that merely alternates between high and low resolution states for the ADCs, and one that adds a third off state inspired by proposals of antenna-selection techniques.

We have studied the capacity and power consumption of the mmWave link under this variable resolution model, and shown that there can be very significant power gains up to 80% in variable-resolution quantization schemes, depending on the link SNR pre-quantization. Our results also show that the benefits are greater with not-so-low and not-so-high numbers of bits for the low and high resolution levels of the ADCs, respectively. This approach seems to outperform the existing literature proposing variable resolution systems with 1 bit for the lowest resolution ADCs.

We would also like to point out that, even though the power savings obtained in our results are not so impressive in the lower SNR regimes, such low-SNR systems do not usually benefit from the kind of spatial multiplexing gains that are the staple of Digital Combining. Therefore, instead of choosing variable-vs-fixed resolution ADCs, at lower SNRs a more radical switch to Analog Combining would make more practical sense.

In the future, we will extend the analysis of variable resolution ADCs design with full spatial multiplexing and optimize the choice of the variable bits based on the power and capacity constraints.

6.1.4 Directional Cell Discovery

In Chapter 4, we investigated the directional initial cell discovery related to millimeter wave communication. We studied how the availability of CI at the MS can reduce the inherent search delay of ABF. However, the performance of ABF starts degrading with an increase of the angular error in the available CI. We also showed that the optimal number of receiver antennas is related to the angular error. In addition, considering small angular errors in CI, ABF with lower power consumption is the best option for initial access. Moreover, we presented an analog beamforming based PSN architecture with a single RF chain to mitigate the effect of the angular error. Simulation results validate that this solution has equivalent performance to HBF, while exhibiting lower power consumption. This makes PSN a viable approach for initial cell search in mmWave 5G cellular networks.

We then extended our analysis and studied the directional initial cell discovery in the context of an energy consumption comparison of ABF, DBF, HBF and PSN schemes. We considered both context information (CInD, CID) and non context information (nCI) based scenario. We showed that the perception regarding the higher energy consumption of DBF is not true during the directional initial cell discovery phase. Rather, DBF has a lower energy consumption than ABF and HBF with much lower angular search delay for both CID (i.e., context information with delay) and nCI scenarios. We would also like to point out that above a certain sub-carrier bandwidth a fully digital architecture starts consuming very high power. This suggests that at a very high sub-carrier bandwidth, DBF may not be a feasible option for power constrained devices (such as a mobile station). Moreover, our proposed PSN architecture has a lower energy consumption than other beamforming schemes at higher sub-carrier bandwidth for the nCI scenario, whereas ABF has minimum energy consumption in the CInD (i.e., CI with no delay) scenario. We also showed that energy consumption results presented in this work (especially related to ABF, DBF and HBF) are valid irrespective of the number of ADC bits.

In the future, we will study the bandwidth vs noise power trade-off and identify the regimes where the energy consumption and the detection probability are optimized. In the future, we will also extend our work to multipath scenarios, and study how different beamforming schemes perform in the presence of multiple transmitting BS. We will also evaluate the optimal number of receiver antennas based on the statistics of the angular error.

6.2 Energy Efficient Flooding of Fountain Codes

In Chapter 5, we considered the controlled flooding of fountain-coded packets throughout a multi-hop network, and analytically determined the probability that the destination can recover the original data packets as a function of the number of nodes in the network, of the link error probability and of the amount of redundancy generated by the code. We showed that our model matches simulation results very well. Based on the conclusions drawn from the model, we designed practical, distributed policies that achieve the same success probability performance while requiring fewer (re)transmissions to advance a packet through the network. We measured the performance of these policies and justified their behavior in light of an analytical model for flooding under a restriction on the number of relays per hop. Among several possible extensions to this work, we considered the case of intermediate relays being able to decode the source packets and thereby inject additional redundancy in the network. We thoroughly developed and evaluated this case.

Future extensions may involve, e.g., the inclusion of plain retransmissions (without decoding the fountain code) or the explicit modeling of MAC-level issues such as backoff mechanisms, interference and collisions.

List of Publications

The work presented in this thesis has appeared or submitted in the articles reported below.

Journal papers

[J1] **W. b. Abbas**, P. Casari M. Zorzi, "Controlled Flooding of Fountain Codes", submitted to *IEEE Transactions on Wireless Communications.*

Contribution

- Partial contribution in probabilistic modeling of the scenario.
- Matrix based representation of the probabilistic model.
- Proposed practical restricted flooding policies.
- Analysis through monte-carlo simulation.

[J2] **W. b. Abbas**, F. Gomez-Cuba and M. Zorzi, "Millimeter Wave Receiver Efficiency: A Comprehensive Comparison of Beamforming Schemes with Low Resolution ADCs", preparing draft for submission in *IEEE Transactions on Wireless Communications.*

Contribution

- Main theme and idea of the paper.
- Partial design of the mathematical model.
- Analysis through monte-carlo simulation.

Conference papers

[C1] P. Casari, **W. b. Abbas** and M. Zorzi, "On the number of transmissions vs. redundancy tradeoff for flooded fountain codes", in *Proc. of IEEE International Workshop on*

Computer-Aided Modeling Analysis and Design of Communication Links and Networks (CAMAD), December 2014, Athens, Greece

Contribution

- Partial contribution in probabilistic modeling of the scenario.
- Matrix based compact representation of the probabilistic model.
- Results evaluation through monte-carlo simulations.

[C2] **W. b. Abbas** and M. Zorzi, "Context information based initial cell search for millimeter wave 5G cellular networks", in *Proc. of European Conference on Networks and Communications (EuCNC)*, June 2016, Athens, Greece

Contribution

- Main theme and idea of the paper.
- System modeling.
- Analysis through monte-carlo simulations.

[C3] **W. b. Abbas** and M. Zorzi, "Towards an Appropriate Receiver Beamforming Scheme for Millimeter Wave Communication: A Power Consumption Based Comparison", in *Proc. of European Wireless*, May 2016, Oulu, Finland

Contribution

- Main theme and idea of the paper.
- System modeling.
- Mathematical analysis.

[C4] **W. b. Abbas**, F. Gomez-Cuba and M. Zorzi, "Bit Allocation for Increased Power Efficiency in 5G Receivers with Variable-Resolution ADCs", preparing draft for submission

Contribution

- Main theme and idea of the paper.
- Partial contribution in system analysis.

  - Proposed two greedy algorithms
  - Analysis through monte-carlo simulations

[C5] **W. b. Abbas** and M. Zorzi, “Towards an Energy Efficient Beamforming Scheme for Initial Cell Discovery in mmW 5G Cellular Networks ”, preparing draft for submission

Contribution

- Main theme and idea of the paper.
- Mathematical analysis and design evaluation.
- New control plane frame structure proposal with higher sub-carrier bandwidth compared to sub-carrier bandwidth of data plane signaling.

Additional work done during the PhD but not reported in this thesis.

[J3] **W. b. Abbas**, N. Ahmed, A. Syed, “Design and evaluation of a low-cost, DIY-inspired, underwater platform to promote experimental research in UWSN”, *Ad Hoc Networks*, 2015

Contribution

- Design of a low cost underwater acoustic transducer.

Bibliography

[1] D. Feng, C. Jiang, G. Lim, L. J. Cimini, G. Feng, and G. Y. Li, "A survey of energy-efficient wireless communications," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 167–178, 2013.

[2] GSMA Intelligence, "Understanding 5G: Perspectives on future technological advancements in mobile," Dec. 2014. [Online]. Available: https://www.gsmaintelligence.com/research/?file=141208-5g.pdf&download

[3] NTT DOCOMO, "DOCOMO 5G white paper, 5G radio access: Requirements, concept and technologies," Jul. 2014. [Online]. Available: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_5g/DOCOMO_5G_White_Paper.pdf

[4] Samsung, "5G vision," Feb. 2015, http://www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-0.pdf.

[5] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 20142019," Feb. 2015, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf.

[6] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and

M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, May 2014.

[7] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[8] Nokia, "Looking ahead to 5G," in *White Paper*, May. 2014. [Online]. Available: http://networks.nokia.com/file/28771/5g-white-paper

[9] Huawei, "5G: A technology vision," in *White Paper*, Jan. 2014. [Online]. Available: http://www.huawei.com/5gwhitepaper/

[10] C. X. Wang, F. Haider, X. Gao, X. H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, Feb. 2014.

[11] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[12] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, Jun 2014.

[13] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[14] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[15] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[16] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, “MIMO for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both?” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 110–121, Dec. 2014.

[17] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, “Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[18] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, “MIMO precoding and combining solutions for millimeter-wave systems,” *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[19] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially sparse precoding in millimeter wave MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[20] J. Mo and R. W. Heath, “Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information,” *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[21] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, “One-bit massive MIMO: Channel estimation and high-order modulations,” in *2015 IEEE International Conference on Communication Workshop (ICCW)*, Jun. 2015.

[22] H. Lim and C. Kim, “Flooding in wireless ad hoc networks,” *Computer Communications*, vol. 24, pp. 353 – 363, 2001.

[23] I. Stojmenovic, M. Seddigh, and J. Zunic, “Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 1, pp. 14–25, Jan. 2002.

[24] W. Lou and J. Wu, “On reducing broadcast redundancy in ad hoc wireless networks,” *IEEE Transactions on Mobile Computing*, vol. 1, no. 2, pp. 111–122, Apr. 2002.

[25] D. J. C. MacKay, “Fountain Codes,” *IEE Proceedings – Communications*, vol. 152, no. 6, pp. 1062–1068, 2005.

[26] M. Luby, “LT Codes,” in *The 43rd Annual IEEE Symposium on Foundations of Computer Science*, Vancouver, B.C., Canada, Nov. 2002.

[27] A. Shokrollahi, “Raptor codes,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.

[28] S. Rangan, T. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[29] J. A. Nossek and M. T. Ivrlac, “Capacity and coding for quantized MIMO systems.” in *Proceedings of the International Conference on Wireless Communications and Mobile Computing, (IWCMC 2006)*, 2006.

[30] B. M. Murray and I. B. Collings, “AGC and quantization effects in a zero-forcing MIMO wireless system,” in *2006 IEEE 63rd Vehicular Technology Conference*, vol. 4, May 2006, pp. 1802–1806.

[31] A. Mezghani and J. A. Nossek, “Capacity lower bound of MIMO channels with output quantization and correlated noise,” in *IEEE International Symposiaum on Information Theory Proceedings (ISIT)*, 2012.

[32] J. Singh, O. Dabeer, and U. Madhow, “On the limits of communication with low-precision analog-to-digital conversion at the receiver,” *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.

[33] Q. Bai, A. Mezghani, and J. A. Nossek, “On the optimization of ADC resolution in multi-antenna systems,” in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS 2013)*, Aug. 2013.

[34] O. Orhan, E. Erkip, and S. Rangan, “Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance,” in *Information Theory and Applications Workshop (ITA)*, Feb. 2015.

[35] J. Mo, A. Alkhateeb, S. Abu-Surra, and R. W. Heath Jr, “Hybrid architectures with few-bit ADC receivers: Achievable rates and energy-rate tradeoffs.” [Online]. Available: https://arxiv.org/abs/1605.00668

[36] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, “Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[37] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. Heath, “Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?” *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.

[38] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, “Uplink achievable rate for massive MIMO systems with low-resolution ADC,” *IEEE Communications Letters*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[39] J. Zhang, L. Dai, S. Sun, and Z. Wang, “On the spectral efficiency of massive MIMO systems with low-resolution ADCs,” *IEEE Communications Letters*, vol. 20, no. 5, pp. 842–845, May 2016.

[40] W. T. Li, Y. C. Chiang, J. H. Tsai, H. Y. Yang, J. H. Cheng, and T. W. Huang, “60-GHz 5-bit phase shifter with integrated VGA phase-error compensation,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 3, pp. 1224–1235, Mar. 2013.

[41] Y. Yu, P. Baltus, A. de Graauw, E. van der Heijden, C. Vaucher, and A. van Roermund, “A 60 GHz phase shifter integrated with LNA and PA in 65 nm CMOS for phased array systems,” *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1697–1709, Sep. 2010.

[42] R. Mendez-Rial, C. Rusu, A. Alkhateeb, N. Gonzalez-Prelcic, and R. Heath, “Channel estimation and hybrid combining for mmwave: Phase shifters or switches?” in *Information Theory and Applications Workshop (ITA)*, Feb. 2015.

[43] H.-S. Lee and C. Sodini, “Analog-to-digital converters: Digitizing the analog world,” *Proceedings of the IEEE*, vol. 96, no. 2, pp. 323–334, Feb. 2008.

[44] O. Orhan, E. Erkip, and S. Rangan, “Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance,” in *Information Theory and Applications Workshop (ITA)*, Feb. 2015.

[45] M. Kraemer, D. Dragomirescu, and R. Plana, “Design of a very low-power, low-cost 60 GHz receiver front-end implemented in 65 nm CMOS technology,” *International Journal of Microwave and Wireless Technologies*, vol. 3, pp. 131–138, Apr. 2011.

[46] C. Barati, S. Hosseini, S. Rangan, P. Liu, T. Korakis, S. Panwar, and T. Rappaport, “Directional cell discovery in millimeter wave cellular networks,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6664–6678, Dec. 2015.

[47] A. Capone, I. Filippini, and V. Sciancalepore, “Context information for fast cell discovery in mm-wave 5G networks,” in *Proceedings of 21th European Wireless Conference*, May 2015.

[48] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[49] W. b. Abbas, F. Gomez-Cuba, and M. Zorzi, “mmwave receiver beamforming comparison tool.” [Online]. Available: http://enigma.det.uvigo.es/~fgomez/mmWaveADCwebviewer/

[50] T. Rappaport, R. Heath, R. Daniels, and J. Murdock, *Millimeter Wave Wireless Communications*, ser. Prentice Hall Communications Engineering and Emerging Technologies Series. Pearson Education, 2014.

[51] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter wave channel modeling and cellular capacity evaluation,” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[52] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, “Robust predictive quantization: Analysis and design via convex optimization,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 618–632, Dec 2007.

[53] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, “Designing structured tight frames via an alternating projection method,” *IEEE Trans. Inf. Theor.*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[54] L. Kong, "Energy-efficient 60 GHz phased-array design for multi-Gb/s communication systems," *Ph.D. dissertation, EECS Department, University of California, Berkeley,* Dec. 2014. [Online]. Available: http://digitalassets.lib.berkeley.edu/techreports/ucb/text/EECS-2014-191.pdf

[55] Y.-H. Lin and H. Wang, "A low phase and gain error passive phase shifter in 90 nm CMOS for 60 GHz phase array system application," in *2016 IEEE MTT-S International Microwave Symposium (IMS)*, May 2016, pp. 1–4.

[56] B. Murmann, "ADC performance survey 1997-2015." [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[57] B. Nasri, S. P. Sebastian, K.-D. You, R. RanjithKumar, and D. Shahrjerdi, "A 700W 1GS/s 4-bit folding-flash ADC in 65nm CMOS for wideband wireless communications." [Online]. Available: https://arxiv.org/abs/1612.04855

[58] W. b. Abbas and M. Zorzi, "Towards an appropriate receiver beamforming scheme for millimeter wave communication: A power consumption based comparison," in *Proceedings of the 22nd European Wireless Conference, 2016.*

[59] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 983–997, Apr. 2016.

[60] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. Jun., pp. 1164–1179, Apr. 2013.

[61] S. Veeramachanen, A. M. Kumar, V. Tummala, and M. B. Srinivas, "Design of a low power, variable-resolution flash ADC," in *22nd International Conference on VLSI Design, 2009*, Jan. 2009.

[62] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[63] A. Capone, I. Filippini, V. Sciancalepore, and D. Tremolada, "Obstacle avoidance cell discovery using mm-waves directive antennas in 5G networks," in *IEEE 26th An-*

nual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Aug. 2015.

[64] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmwave communications," in *48th Asilomar Conference on Signals, Systems and Computers*, Nov. 2014.

[65] M. Giordani, M. Mezzavilla, C. N. Barati, S. Rangan, and M. Zorzi, "Comparative analysis of initial access techniques in 5G mmwave cellular networks," in *50th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2016.

[66] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 180–185, Jan. 2015.

[67] C. N. B. Nt., S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi, "Initial access in millimeter wave cellular systems," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–14, 2016.

[68] J. Wang, Z. Lan, C.-S. Sum, C.-W. Pyo, J. Gao, T. Baykas, A. Rahman, R. Funada, F. Kojima, I. Lakkis, H. Harada, and S. Kato, "Beamforming codebook design and performance evaluation for 60 GHz wideband WPANs," in *IEEE 70th Vehicular Technology Conference Fall (VTC)*, Sep. 2009.

[69] H. L. Van Trees, *Detection, estimation, and modulation theory. Part IV. Optimum array processing.* New York: Wiley-Interscience, 2002.

[70] J. Litva and T. K. Lo, *Digital Beamforming in Wireless Communications*, 1st ed. Norwood, MA, USA: Artech House, Inc., 1996.

[71] F. V. Diggelen, *A-GPS: Assisted GPS, GNSS, and SBAS.* Artech House, 2009.

[72] M. Mitzenmacher, "Digital fountains: a survey and look forward," in *Proc. IEEE ITW*, San Antonio, TX, Oct. 2004.

[73] M. Rossi, N. Bui, G. Zanca, L. Stabellini, R. Crepaldi, and M. Zorzi, "SYNAPSE++: Code dissemination in wireless sensor networks using fountain codes," *IEEE Trans. Mobile Comput.*, vol. 9, no. 12, pp. 1749–1765, Dec. 2010.

[74] Z. Zhou, H. Mo, Y. Zhu, Z. Peng, J. Huang, and J.-H. Cui, "Fountain code based adaptive multi-hop reliable data transfer for underwater acoustic networks," in *Proc. IEEE ICC*, Ottawa, Canada, Jun. 2012.

[75] A. James, A. S. Madhukumar, E. Kurniawan, and F. Adachi, "Performance analysis of fountain codes in multihop relay networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4379–4391, Nov. 2013.

[76] R. Budde, S. Nowak, and R. Kays, "Reliable broadcast transmission in vehicular networks based on fountain codes," in *Proc. IEEE VTC Spring*, Budapest, Hungary, May 2011.

[77] N. Kadi and K. Al Agha, "MPR-based flooding with distributed fountain network coding," in *Proc. IFIP Med-Hoc-Net*, Juan-les-pins, France, Jun. 2010.

[78] A. Apavatjrut, K. E. Jaffrés-Runser, C. Goursaud, and C. Lauradoux, "Overflow of fountain codes in multi-hop wireless sensor networks," in *Proc. IEEE PIMRC*, Toronto, Canada, Sep. 2011.

[79] E. Altman and F. de Pellegrini, "Forward correction and fountain codes in delay-tolerant networks," *IEEE/ACM Trans. Networking*, vol. 19, no. 1, pp. 1–13, Feb. 2011.

[80] M. Gerami and M. Xiao, "Repair for distributed storage systems with erasure channels," in *Proc. IEEE ICC*, Budapest, Hungary, Jun. 2013.

[81] P. Casari, W. bin Abbas, and M. Zorzi, "On the number of transmissions vs. redundancy tradeoff for flooded fountain codes," in *Proc. IEEE CAMAD*, Athens, Greece, Dec. 2014.

[82] A. James and A. Madhukumar, "Performance limits of rateless codes in delay constrained multihop relay networks," in *Proc. IEEE ICC*, Sydney, Australia, Jun. 2014.

[83] A. Crismani, U. Schilcher, S. Toumpis, G. Brandner, and C. Bettstetter, "Packet travel times in wireless relay chains under spatially and temporally dependent interference," in *2014 IEEE International Conference on Communications (ICC)*, Jun. 2014.

[84] G. Sim, J. Widmer, and B. Rengarajan, “Opportunistic finite horizon multicasting of erasure-coded data,” *IEEE Trans. Mobile Comput.*, vol. 15, no. 3, pp. 705–718, Mar. 2016.

[85] P. Casari and M. Rossi and M. Zorzi, “Towards Optimal Broadcasting Policies for HARQ based on Fountain Codes in Underwater Networks,” in *Proc. IEEE/IFIP WONS*, Garmisch-Partenkirchen, Germany, Jan. 2008.

[86] R. Diamant and L. Lampe, “Adaptive error-correction coding scheme for underwater acoustic communication networks,” *IEEE J. Oceanic Eng.*, vol. 40, no. 1, pp. 104–114, Jan. 2014.

[87] R. Ahmed, M. Stojanovic, and M. Chitre, “Random linear packet coding for broadcast networks,” in *Proc. MTS/IEEE OCEANS*, St. John’s, NL, Canada, Sep. 2014.

[88] A. James and A. Madhukumar, “Enhanced packet delivery in delay limited rateless coded multihop networks,” in *Proc. IEEE ICCS*, Macau, China, Nov. 2014.

[89] Y. Sasson, D. Cavin, and A. Schiper, “Probabilistic broadcast for flooding in wireless mobile ad hoc networks,” in *Proc. IEEE WCNC*, New Orleans, LA, Mar. 2003.

[90] S. Crisostomo, U. Schilcher, C. Bettstetter, and J. Barros, “Analysis of probabilistic flooding: How do we choose the right coin?” in *Proc. IEEE ICC*, Dresden, Germany, Jun. 2009.

[91] H. AlHazza, “Enhancing dynamic probabilistic broadcasting flooding scheme in MANETs,” in *Proc. IEEE CHUSER*, Penang, Malaysia, 2011.

[92] A. Hanashi, I. Awan, and M. Woodward, “Performance evaluation based on simulation of improving dynamic probabilistic flooding in MANETs,” in *Proc. WAINA*, Bradford, UK, May 2009.

[93] C. Betoule, T. Bonald, R. Clavier, D. Rossi, G. Rossini, and G. Thouenon, “Adaptive probabilistic flooding for multipath routing,” in *Proc. NTMS*, Istanbul, Turkey, May 2012.

[94] Y. Mylonas, M. Lestas, A. Pitsillides, P. Ioannou, and V. Papadopoulou, "Speed adaptive probabilistic flooding for vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 1973–1990, May 2015.

[95] X. Shen, Y. Chen, Y. Zhang, J. Zhang, Q. Ge, G. Dai, and T. He, "OppCode: Correlated opportunistic coding for energy-efficient flooding in wireless sensor networks," vol. 11, no. 6, pp. 1631–1642, Dec. 2015.

[96] T. Kunz, S. Paul, and L. Li, "Efficient broadcasting in tactical networks: Forwarding vs. network coding," in *Proc. IEEE MILCOM*, San Jose, CA, Oct. 2010.

[97] H. Xi, X. Wang, Y. Zhao, and H. Zhang, "A reliable broadcast transmission approach based on random linear network coding," in *Proc. IEEE VTC Spring*, Yokohama, Japan, May 2012.

[98] B. Vellambi, N. Rahnavard, and F. Fekri, "FTS: A distributed energy-efficient broadcasting scheme using fountain codes for multihop wireless networks," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3561–3572, Dec. 2010.

[99] G. Liva, E. Paolini, and M. Chiani, "Performance versus overhead for fountain codes over $\mathbb{F}_q$," *IEEE Commun. Lett.*, vol. 14, no. 2, pp. 178–180, Feb. 2010.

[100] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed. Academic Press, 1998.