Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Real–time Wireless Networks for Industrial Control Systems

Ph.D. candidate
Michele Luvisotto

Advisor
prof. Stefano Vitturi

Director & Coordinator
prof. Andrea Neviani

Ph.D. School in
Information Engineering

Department of
Information Engineering
University of Padova
2017

Abstract

The next generation of industrial systems (Industry 4.0) will dramatically transform many productive sectors, integrating emerging concepts such as Internet of Things, artificial intelligence, big data, cloud robotics and virtual reality, to name a few. Most of these technologies heavily rely on the availability of communication networks able to offer nearly–istantaneous, secure and reliable data transfer. In the industrial sector, these tasks are nowadays mainly accomplished by wired networks, that combine the speed of optical fiber media with collision–free switching technology.

However, driven by the pervasive deployment of mobile devices for personal communications in the last years, more and more industrial applications require wireless connectivity, which can bring enormous advantages in terms of cost reduction and flexibility. Designing timely, reliable and deterministic industrial wireless networks is a complicated task, due to the nature of the wireless channel, intrinsically error–prone and shared among all the devices transmitting on the same frequency band.

In this thesis, several solutions to enhance the performance of wireless networks employed in industrial control applications are proposed. The presented approaches differ in terms of achieved performance and target applications, but they are all characterized by an improvement over existing industrial wireless solutions in terms of timeliness, reliability and determinism. When possible, an experimental validation of the designed solutions is provided.

The obtained results prove that significant performance improvements are already possible, often using commercially available devices and preserving compliance to existing standards. Future research efforts, combined with the availability of new chipsets and standards, could lead to a world where wireless links effectively replace most of the existing cables in industrial environments, as it is already the case in the consumer market.

Sommario

La prossima generazione di sistemi industriali (Industria 4.0) rivoluzionerà molti settori produttivi, grazie all'integrazione di concetti emergenti quali Internet delle Cose, intelligenza artificiale, "big data", robotica nel cloud e realtà virtuale, tra gli altri. La maggior parte di queste tecnologie necessita di reti di comunicazione in grado di offrire un trasferimento di informazione quasi–istantaneo, sicuro e affidabile. Nel settore industriale, ad oggi, ciò è reso possibile principalmente da reti cablate, in grado di combinare la velocità della fibra ottica con l'assenza di collisioni data dalla tecnologia Switched Ethernet.

Tuttavia, spinte dall'intensa diffusione, negli ultimi anni, di dispositivi mobili per le comunicazioni interpersonali, sempre più applicazioni industriali richiedono connettività wireless, che porterebbe enormi vantaggi in termini di riduzione dei costi e flessibilità. Progettare reti industriali wireless puntuali, affidabili e deterministiche è molto complicato, a causa della natura del canale radio, incline agli errori e intrinsecamente condiviso tra tutti i dispositivi che trasmettono nella stessa banda di frequenza.

In questa tesi sono proposte diverse soluzioni per migliorare le prestazioni di reti wireless impiegate in applicazioni di controllo industriale. Gli approcci presentati si differenziano per prestazioni raggiunte e scenari di utilizzo, ma sono accomunati da un miglioramento rispetto alle soluzioni wireless industriali esistenti in termini di puntualità, affidabilità e determinismo. Una verifica sperimentale delle soluzioni progettate è stata effettuata, quando possibile.

I risultati ottenuti dimostrano che si possono già ottenere importanti miglioramenti nelle prestazioni, spesso utilizzando dispositivi commerciali e mantenendo la conformità agli standard esistenti. Le ricerche future, assieme alla disponibilità di nuovi dispositivi e standard, potranno rendere possibile un mondo in cui i link radio rimpiazzeranno con successo la maggior parte dei cavi esistenti negli ambienti industriali, com'è già avvenuto in ambito consumer.

Acknowledgments

First and foremost, I would like to thank my PhD advisor, Prof. Stefano Vitturi. Ever since I started my master five years ago, he was a constant presence and an endless source of support and advice, helping me to grow not only as a researcher, but also as a man. He constantly ensured that I found realization and fun in what I was doing, always putting my happiness as the top priority. I wish him all the best for the future and I strongly hope to continue working with him.

I would like to also thank the other professors that I had the pleasure to work with at the University of Padova, particularly Prof. Angelo Cenedese and Prof. Michele Zorzi, from whom I have learned a lot.

A special thank goes to my supervisor at ABB Corporate Research, Dr. Zhibo Pang. He has involved me in a project as ambitious and exciting as I could hope for, giving me a lot of freedom in its development and always motivating me to work hard. I am happy to be a colleague of his now and look forward to continue this succesful work together. I also thank the other colleagues at ABB, from Dacfey Dzung, who has been an unbeatable source of knowledge, to Roger Jansson and my manager Linus Thrybom.

Going back to UniPD, I would like to thank my colleagues from the EMC lab, especially Federico, who has worked side by side with me for most of my projects and has taught me a countless amount of things, and Guglielmo, who has been an unbelievable source of fun.

A deep thank goes to my colleagues from office 330: Giulia, Chiara, Nicoletta, Giacomo, Irene, Yutao, Diego, Giulia, Andrea and Marco. Thanks to you, the PhD life has been really enjoyable, both inside and outside the office.

Moving to a more personal side, I must mention my two great flat mates, Pierre and Zimbo. My Padova years (arguably the best years of my life so far) will be forever linked with you.

It is impossible to forget all my buddies in Oderzo (and surrounding areas): Francesco, Marwy, Gabriele, Fabio, Mattia, Alessandro, Anna, Sara, Lara, Alessandro, Alberto, Marco, Giulia and many others. We have known each other since quite a long time now,

yet going back home and meet with you is always a heart-warming experience.

I would like to thank deeply all the people I met in Västerås: Yuhei, Giorgios, Ida, Konstantina, Riccardo, Fahimeh, Meha, Karthi, Andreas, Sotiris, Federico, Marta and many others. Whenever I spend time with you, Sweden is not as cold and dark as it might seem.

The warmest and most heartfelt thanks goes to my parents. Actually, saying thank you is not enough at all. You have always been there for me, through bad and good times, constantly giving me everything I needed and much more. You supported my decisions and prevented me from taking bad ones. You have always shown interest and passion for what I was learning, from kindergarden to PhD. You taught me the most important things in life and I can only hope to be one day as extraordinary as a parent as you have been with me. I also thank the rest of my family, from my grandmothers, to all my uncles, aunts and cousins.

Last, but definitely not least, a special thank goes to the person that has changed my life over the last three years. It is really impossible to describe how much joy you bring in my life everyday and how much a better person I have become since being together with you. I thank you so much for always being so supportive and proud of me and for never stopping to be curious and passionate about my work. I am really grateful for the neverending patience, strength and trust you have shown during the periods while I was abroad. Thankfully, these times are over now and I am strongly looking forward to start a new adventure together with you. I love you, Sophie.

Contents

ACRONYMS

2G Second Generation

3GPP Third Generation Partnership Project

3G Third Generation

4G Fourth Generation

5G Fifth Generation

6LoWPAN IPv6 over Low power WPAN

6LoWPAN IPv6 over Low power WPAN

ACK Acknowledgement

ADC Analog–to–Digital Converter

ADR Adaptive Data Rate

AEC Average Energy Consumed

AGC Automatic Gain Control

AGV Automatic Guided Vehicle

AIFS Arbitration Inter Frame Space

AMCA Asynchronous Multi–channel Adaptation

AP Access Point

API Application Programming Interface

ARF Automatic Rate Fallback

ARQ Automatic Repeat Request

ASK Amplitude Shift Keying

AV Audio–Video

AWGN Additive White Gaussian Noise

BA Building Automation

BACK Block Acknowledgement

BACK2F Backoff to Frequency

BER Bit Error Rate

BLE Bluetooth Low Energy

BLINK Radio Frequency Identification Blink

BPSK Binary Phase–Shift Keying

BS Base Station

CAN Controller Area Network

CBR Constant Bitrate

CCA Clear Channel Assessment

CCK Complementary Code Keying

CDMA Code–Division Multiple Access

CER Chunk Error Rate

CFO Carrier Frequency Offset

CIM Computer–Integrated Manufacturing

CIP Common Industrial Protocol

CP Cyclic Prefix

CPS Cyber Physical Sytem

CRC Cyclic Redundancy Check

CSMA Carrier Sense Multiple Access

CSMA/CA Carrier Sense Multiple Access with Collision Avoidance

CSS Chirp Spread Spectrum

CTS Clear–to–Send

CW Contention Window

D2D Device–to–Device

DAC Digital–to–Analog Converter

DC Direct Current

DCF Distributed Coordination Function

DIFS Distributed Coordination Function Inter Frame Space

DoS Denial–of–Service

DSME Deterministic and Synchronous Multi–channel Extension

DSSS Direct–sequence Spread Spectrum

DS-UWB Direct–sequence Ultrawide Band

ECDF Empirical Cumulative Distribution Function

ED End Device

EDGE Enhanced Data rates for GSM Evolution

EDR Enhanced Data Rate

EEE Energy–Efficient Ethernet

EIRP Equivalent Isotropically Radiated Power

eMBB Enhanced Mobile Broadband

ERP Enterprise Resource Planning

ERP–OFDM Extended Rate PHY OFDM

ESINR Equivalent Signal–to–Interference plus Noise Ratio

ESIR Equivalent Signal–to–Interference Ratio

ET Exposed Terminal

EWMA Exponential Weighted Moving Average

FA Factory Automation

FARF Fast reduction rate ARF

FBMC Filter Bank Multicarrier

FCS Frame Check Sequence

FD Full Duplex

FDD Frequency–Division Duplex

FDMA Frequency–Division Multiple Access

FEC Forward Error Correction

FFD Full–Function Device

FFT Fast Fourier Transform

FHSS Frequency–Hopping Spread Spectrum

FIFO First–in First–out

FIP Factory Instrumentation Protocol

FMS Flexible Manufacturing System

FN False Negative

F-OFDM Filtered Orthogonal Frequency–Division Multiplexing

FPGA Field–Programmable Gate Array

GDP Gross Domestic Product

GF Greenfield

GFDM Generalized Frequency–Division Multiplexing

GFSK Gaussian Frequency–Shift Keying

GI Guard Interval

GIPT Global Interpacket Time

GPRS General Packet Radio Service

GSM Global System for Mobile communications

GTS Guaranteed Time Slot

GW Gateway

HARQ Hybrid Automatic Repeat Request

HCF Hybrid Coordination Function

HD Half–Duplex

HR-WPAN High Rate–Wireless Personal Area Network

HSDPA High Speed Downlink Packet Access

HSI High–Speed Interface

HSPA+ Evolved High Speed Packet Access

HSUPA High Speed Uplink Packet Access

HT High Throughput

ICN Industrial Communication Network

ICPS Industrial Cyber Physical Sytem

ICT Information and Communication Technology

IEC International Electrotechnical Commission

IEEE Institute of Electrical and Electronic Engineers

IETF Internet Engineering Task Force

IGBT Insulated Gate Bipolar Transistor

IIoT Industrial Internet–of–Things

IoT Internet–of–Things

IP Internet Protocol

IPT Interpacket Time

IPv6 Internet Protocol version 6

ISI Inter–Symbol Interference

ISM Industrial Scientific and Medical

ISO International Organization for Standardization

ITU International Telecommunication Union

IWSN Industrial Wireless Sensor Network

LAN Local Area Network

LBT/AFA Listen–Before–Talk/Adaptive–Frequency–Agility

LDPC Low–Density Parity–Check

LLDN Low–Latency Deterministic Network

LOS Line–of–Sight

LPSC Low–Power Single Carrier

LPWAN Low–Power Wide–Area Network

LR-WPAN Low Rate–Wireless Personal Area Network

LTE Long–Term Evolution

LTE-A Long–Term Evolution Advanced

MAC Medium Access Control

MAP Mobile Application Part

MB-OFDM Multiband Orthogonal Frequency–Division Multiplexing

MCS Modulation and Coding Scheme

MES Manufacturing Execution System

MF Mixed Format

MIMO Multiple–Input Multiple–Output

MIPT Mean Interpacket Time

mMTC Massive Machine–Type Communications

mmWave millimeter–wave

MRC Maximum–ratio Combining

MRS Multi–rate Support

MU-MIMO Multiuser Multiple–Input Multiple–Output

NAV Network Allocation Vector

NB-IoT Narrowband Internet of Things

NCS Networked Control System

NFC Near Field Communication

NFV Network Function Virtualization

NLOS Non Line–of–Sight

NS Network Server

OFDM Orthogonal Frequency–Division Multiplexing

OFDMA Orthogonal Frequency–Division Multiple Access

OOBE Out–of–band Emissions

O-QPSK Orthogonal Quadrature Phase-Shift Keying

OSI Open Systems Interconnection

P2P Point–to–point

PA Process Automation

PAN Personal Area Network

PAPR Peak–to–Average Power Ratio

PC Personal Computer

PCF Point Coordination Function

PEC Power Electronics Control

PER Packet Error Rate

PHY Physical layer

PLC Programmable Logic Controller

PoS Probability of Success

PPDU Physical layer Packet Data Unit

PRP Parallel Redundancy Protocol

PSA Power Systems Automation

PSDU Physical layer Service Data Unit

PSMP Power–Save Multi–Pol

PSK Phase–Shift Keying

PSSS Parallel–sequence Spread Spectrum

PT Primary Transmitter

QAM Quadrature Amplitude Modulation

QoS Quality–of–Service

QPSK Quadrature Phase–Shift Keying

RA Rate Adaptation

RBAR Received Based Autorate

RCFD RTS/CTS in the Frequency Domain

RDP Reverse Direction Protocol

RF Radio Frequency

RFD Reduced–Function Device

RFID Radio Frequency Identification

RPMA Random Phase Multiple Access

RR RTS Receiver

RSI Residual Self–Interference

RSIN Rate Selection in Industrial Networks

RSIN–E Enhanced Rate Selection in Industrial Networks

RSSI Received Signal Strength Indicator

RT-CPS Real–time Cyber Physical Sytem

RTE Real–Time Ethernet

RTS Request–to–Send

SARF Static retransmission rate ARF

SC Subcarrier

SCADA Supervisory Control and Data Acquisition

SC-FDMA Single Carrier Frequency Division Multiple Access

SCM Supply Chain Management

SDM Spatial Division Multiplexing

SDR Software Defined Radio

SE Signal Extension

SF Spreading Factor

SI Self–Interference

SIC Self–Interference Cancellation

SIFS Short Inter Frame Space

SIG Special Interest Group

SINR Signal–to–Interference plus Noise Ratio

SIR Signal–to–Interference Ratio

SMS Short Message Service

SNR Signal–to–Noise Ratio

STA Station

STBC Space–Time Block Coding

SU Scheduling Unit

TCP Transmission Control Protocol

TDD Time Division Duplexing

TDMA Time–Division Multiple Access

TPC Transmit Power Control

TSC Time Stamp Counter

TSCH Time–Slotted Channel Hopping

TSF Timing Synchronization Function

TTI Transmission Time Interval

TxBF Transmit Beamforming

TXOP Transmit Opportunity

UDP User Datagram Protocol

UE User Equipment

UEP Unequal Error Protection

UFMC Universally Filtered Multicarrier

UMTS Universal Mobile Telecommunication System

UNB Ultranarrow Band

URLLC Ultra–Reliable and Low–Latency Communications

US United States

USRP Universal Software Radio Peripheral

UTRAN Universal Mobile Telecommunication System Terrestrial Radio Access Network

UWB Ultrawide Band

WAN Wide Area Network

WIA-FA Wireless networks for Industrial Automation – Factory Automation

WIA-PA Wireless networks for Industrial Automation – Process Automation

WiMAX Worldwide Interoperability for Microwave Access

WirelessHP High–performance Wireless

WISA Wireless Interface for Sensors and Actuators

WLAN Wireless Local Area Network

WNIC Wireless Network Interface Card

WPAN Wireless Personal Area Network

WSN Wireless Sensor Network

1 Introduction

1.1 The fourth industrial revolution

The introduction of computers and networking in industrial automation, which started roughly in the mid-1970s, completely transformed all economic sectors (Sauter, 2007). This breakthrough is often referred to as the *third industrial revolution*, after the introduction of steam power at the end of the eighteenth century and the emergence of mass production at the end of the nineteenth one (Wollschlaeger et al., 2017). Today, the industrial world is on the verge of another radical change, the so-called *fourth industrial revolution* or, with a term originally created for a German national program, Industry 4.0 (Industrie 4.0).

This new vision involves the integration in industries of concepts originally developed in the Information and Communication Technology (ICT) world, such as the Internet–of–Things (IoT), Cyber Physical Sytems (CPSs) and tactile Internet. The first concept, defined as early as in 1999 (Ashton), envisages the massive spread of devices with sensing, processing, and communication capabilities, all interconnected to each other through the Internet. The CPS paradigm, instead, refers to the ever increasing number of physical (analog) systems whose operations are monitored or controlled by a computing (digital) core (Wolf, 2009). Finally, the tactile Internet concept was defined by the International Telecommunication Union (ITU) in 2014 as the possibility of accessing Internet with

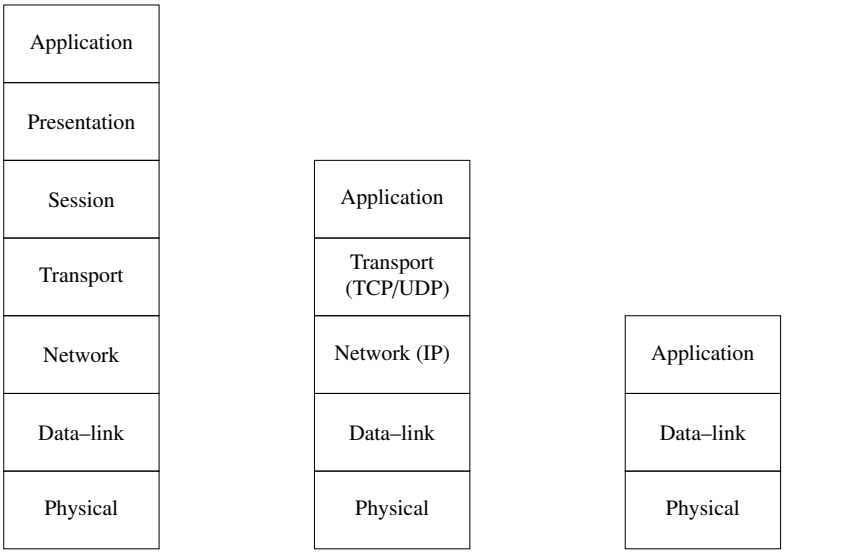
a very low latency, a very high availability, high reliability and a high level of security (Fettweis, 2014), unlocking a series of emerging applications that require extremely-low reaction times, such as virtual/augmented reality (Ong and Nee, 2013) and haptic control (Lee et al., 2002).

All these concepts, already quite popular in the ICT world, are now ready to shock the world of industrial manufacturing, effectively starting the fourth industrial revolution (Wollschlaeger et al., 2017). Adapting to the different assumptions and requirements of industrial applications would require slight modifications in these visions, and a new vocabulary is starting to be developed. For example, the deployment of IoT in industries is referred to as Industrial Internet–of–Things (IIoT) (Wan et al., 2016), while CPSs that are designed according to the stringent performance requirements of industrial applications are called Industrial Cyber Physical Sytems (ICPSs) (Colombo et al., 2016) or Real–time Cyber Physical Sytems (RT-CPSs) (Pang et al., 2017).

1.2 Communication networks in industry

The success of Industry 4.0 in all its different aspects, from IIoT to RT-CPS, depends in large part on the availability of high–performance communication networks. Indeed, the core of any automated industrial system is represented by the reliable and timely exchange of information among distributed entities, such as sensors, controllers and actuators (Wollschlaeger et al., 2017). Moreover, as more and more applications involving mobility arise, the use of wireless networks to connect distributed nodes becomes crucial also in the industrial sector, after having changed the landscape of personal communications in the last decades. The need for high–performance wireless networks to be deployed in industrial control and monitoring systems represents actually the motivation of this thesis.

The introduction of communication networks in industrial automation is not a new concept. Indeed, as soon as the third industrial revolution began in the 1970s, new communication interfaces started to be developed to connect computers that were now deployed to control and monitor industrial processes (Sauter, 2010). These first interfaces, generally termed as fieldbuses and developed in the 1980s, were based on the serial transmission of control and data over a wired bus and on a rigid master–slave architecture. The main goal of these networks was not to achieve the highest possible spectral efficiency, as in traditional telecommunication networks, but to exchange data in a deterministic, reliable and timely way, minimizing the occurrence of delay and losses in the communication. Only in this way, indeed, control applications previously relying on



(a) ISO/OSI stack (b) Internet stack (c) Industrial communications stack

Figure 1.1: Different protocol stacks adopted in communication networks.

point–to–point wiring could work seamlessly.

The need for deterministic communication prevented the use of general–purpose wired networks, already employed in home and offices, such as Ethernet, until the end of the century, where the Real–Time Ethernet (RTE) networks started to appear in the industrial sector. These solutions, made possible by the introduction of switching and full–duplex technologies, impose additional control on data exchange with respect to traditional Ethernet networks, to improve the performance in terms of, particularly, timeliness and reliability.

Only recently, wireless networks have started to carve out a role in this picture, driven by the increasing demand for mobile connectivity. However, although some industrial wireless solutions are available and satisfactorily deployed, they are still not as performing as wired networks, at least in an industrial context. Indeed, the intrinsic error–prone and shared nature of the wireless communication channel severely hampers the achievement of reliable, timely and deterministic data exchange, as demanded by industrial applications (Willig et al., 2005).

The typical protocol stack of an Industrial Communication Network (ICN), be it wired or wireless, considerably differs from that adopted in traditional ICT networks, generally referred to as International Organization for Standardization (ISO)/Open Systems Interconnection (OSI) stack, which can be seen in Fig. 1.1. Indeed, the latter is formed by seven different layers, which are briefly discussed in the following (Rappaport,

1996).

1. *Physical layer (PHY)*: it is the closest to hardware and it is in charge of transmitting and receiving raw bit streams over a physical medium.

2. *Data–link layer*: it is in charge of ensuring reliable data exchange between two end points connected by a physical link.

3. *Network layer*: it is in charge of a multi–point network, formed by several links, and it must handle routing and addressing.

4. *Transport layer*: it is in charge of ensuring reliable data exchange between any two nodes in a network (not necessarily linked directly).

5. *Session layer*: it is in charge of managing a communication session between any two nodes in a network.

6. *Presentation layer*: it is in charge of translating data between the networking services and the application that actually requires data exchange.

7. *Application layer*: it is the closest to the user, providing the Application Programming Interfaces (APIs) to realize any particular task which requires data exchange.

The ISO/OSI model is quite heavy and, in particular, the presentation and session layers are not always needed. Indeed, in the majority of telecommunication networks, the simpler Internet stack (also visible in Fig. 1.1) is adopted, which is limited to the other five layers.

The protocol stack usually adopted in industrial communications is even more condensed and comprises only three layers, namely PHY, data–link and the overlying application involved with the control or monitoring of an industrial process (IEC 61158-2003). Typically, the PHY is unvaried with respect to general–purpose networks, and the peculiarities of industrial communications are implemented in the data–link layer, which serves as an interface with the control application and ensures deterministic and efficient communication (Sauter, 2007). However, more recently, the idea of customizing also the lowest layer of the stack has started to being considered (Wollschlaeger et al., 2017; Luvisotto et al., 2017a), with the goal of further improving the determinism, latency and reliability of the communication. Some of the topics discussed in this thesis follow this trend, while others pursue a more traditional approach, building on top of consolidated lower layers taken from general–prupose wireless standards.

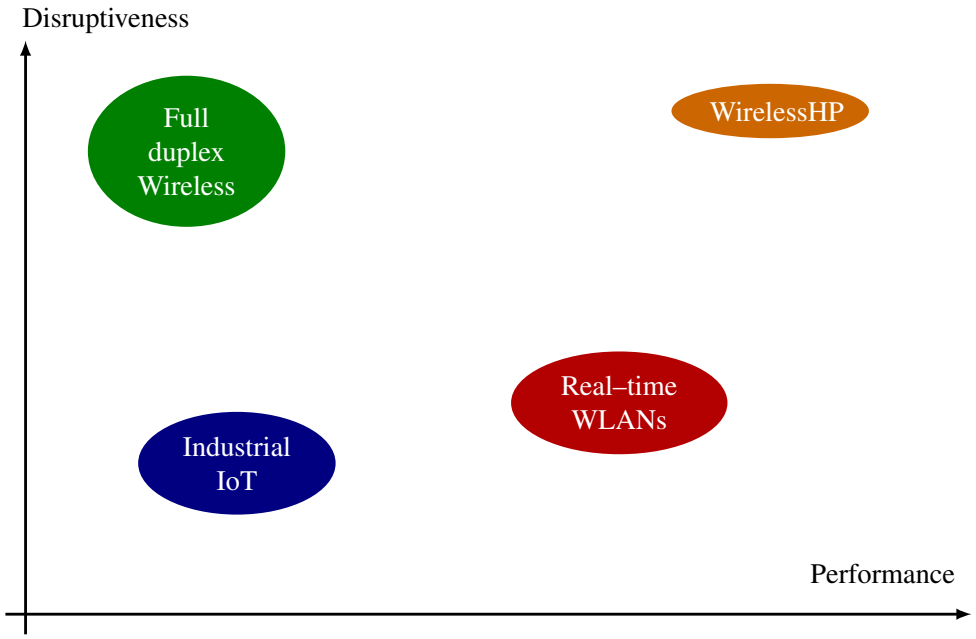


Figure 1.2: Main topics of this thesis, classified according to performance and disruptiveness.

1.3 Main topics of the thesis

The focus of this thesis is the design of wireless networks able to provide the low–latency, high–reliability and deterministic data exchange that is required by industrial control applications. In this regard, four main topics are discussed in this thesis and they are classified in a qualitative way in Fig. 1.2.

The classification is carried out according to two different indicators, namely *performance* and *disruptiveness*. The term "performance" does not refer to common performance metrics for communication networks (e.g., spectral efficiency), but rather to the key features required to industrial networks: timeliness, reliability and determinism. The performance that a network is able to provide also impacts its possible applications: a high–performing network can be used for fast–dynamics closed–loop industrial control, whereas a low–performing one may be used only for monitoring purposes. The "disruptiveness metric", instead, refers to the layers of the protocol stack in which each solution takes place: a more disruptive solution will be characterized by customized bottom layers, whereas less disruptive ones will only act as applications. Performance and disruptiveness usually go hand in hand, in the sense that a more disruptive solution, relying on customized bottom layers, is more likely to offer higher performance than a middleware–based one. However, this is not the case for all the topics in this thesis, as

detailed in the following.

The first topic discussed is real–time Wireless Local Area Networks (WLANs). These networks, commonly known as "Wi-Fi", are defined in the Institute of Electrical and Electronic Engineers (IEEE) 802.11 standard (IEEE 802.11-2016) and represent arguably the most common wireless solution for home/office applications. Their usage in the industrial environment is being proposed since quite a few years (Moraes et al., 2007), even if many issues have to be overcome, the most important being the IEEE 802.11 channel access strategy, based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA), which does not provide sufficient determinism. In this thesis some possible enhancements to reliability and determinism of WLANs are discussed, mostly related to the use of Multiple–Input Multiple–Output (MIMO) architectures and Rate Adaptation (RA) algorithms. The proposed enhancements do not modify the two layers defined in the IEEE 802.11 standard, namely data–link and PHY, and hence they imply a low disruptiveness.

The second topic deals with Full Duplex (FD) wireless networks. Around 2010, several research groups started to develop PHY techniques for the cancellation of Self–Interference (SI), effectively allowing a wireless node to transmit and receive simultaneously in the same frequency band (Duarte and Sabharwal, 2010). In this thesis, a channel access protocol for FD wireless networks is proposed and discussed in detail. Despite being highly disruptive, the proposed solution has only being investigated so far in traditional ad hoc networks, hence its performance figures in an industrial sense are still unknown. Nonetheless, some considerations for the effective usage of FD wireless in industrial applications are drawn.

The most disruptive topic in this thesis is represented by High–performance Wireless (WirelessHP), a new proposal for wireless networks characterized by ultra–low latency, ultra–high reliability and high determinism. These network target critical industrial control use cases, such as robotics, mining and power systems automation, and thet are based on a complete redesign of the protocol stack, to achieve the required high performance.

Finally, the last topic addressed in this thesis deals with the IIoT and proposes the use of a Low–Power Wide–Area Network (LPWAN), namely LoRaWAN, for the monitoring of indoor industrial processes. This relatively new kind of networks offers great communication range and extremely low power consumption, at the cost of a reduced data rate. The low speed combined with band–specific regulations do not allow high sampling rates, hence this solution can only be used for slow–dynamics monitoring applications. For these applications, however, it can represent an interesting opportunity,

thanks to its highly reliable data exchange and very long battery life.

1.4 Structure of the thesis

The remainder of this thesis is organized as follows.

Chapter 2 introduces the use of communication networks in industrial control applications in an exhaustive way. The features of industrial Networked Control Systems (NCSs) are described in detail and the requirements for communication networks adopted in this scenario are derived. Then, an overview of different solutions adopted over the years is given, from fieldbuses to industrial wireless networks. Focusing on the latter solution, the main problems arising from their usage are presented and some possible solutions are briefly discussed.

Chapter 3 is dedicated to the description of the most important international standards for wireless networks, starting from the IEEE 802.11 standard for WLANs and the IEEE 802.15.4 standard for Wireless Personal Area Networks (WPANs). Other wireless standards, less used in industrial applications, are also discussed, such as cellular networks and Bluetooth.

The first main topic of the thesis, real–time WLANs, is discussed in *Chapter 4*. The first part is dedicated to the use of the IEEE 802.11n amendment in industrial communications, then an original algorithm for industrial rate adaptation is presented.

Chapter 5 is dedicated to FD wireless. After a brief introduction on this technology, a channel access protocol for ad hoc FD wireless networks is presented. Finally, some considerations on possible industrial applications are drawn.

The design of WirelessHP networks for critical control applications is the subject of *Chapter 6*. The application scenarios are first described and performance requirements are derived. Then, the performance offered by the most advanced wireless network standards are presented, concluding that they are not suitable for the discussed applications. Consequently, directions for the development of a new solution, based on a completely new protocol stack, are presented, stemming from new trends present in the literature. The first step in this direction, with the design of a low–latency PHY, is presented and validated through experimental measurements.

The last topic of the thesis, namely the use of LoRaWAN for IIoT applications, is discussed in *Chapter 7*. After a general discussion on IIoT, the most common LPWANs are presented, with a focus on LoRaWAN. The indoor industrial monitoring scenario is then introduced and an accurate model for simulating LoRaWAN performance in this context is provided. Finally, the performance of this standard are assessed and compared

with those of IEEE 802.15.4.

Chapter 8 concludes this thesis, summing up the main results and mentioning some possible future research directions.

2

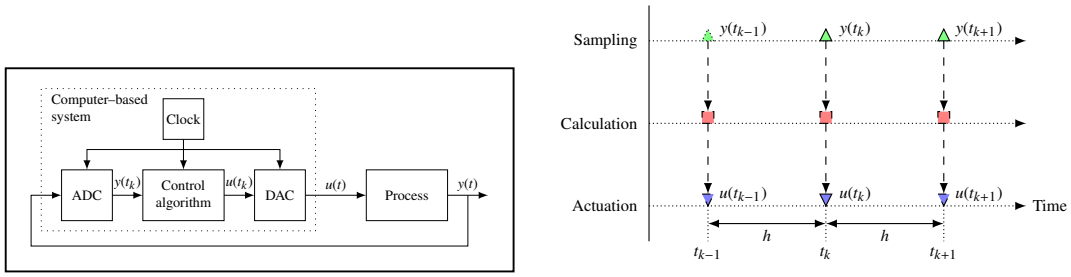
Communication Networks for Industrial Control

In modern control systems, real–time communication networks represent a fundamental element, which can dictate both the controller design and its performance (Wittenmark et al., 1995). This chapter explores the centrality of communication networks in control systems design as well as the most widespread networking technologies currently employed in industrial control applications.

2.1 Industrial networked control systems

A control system can generally be defined as an interconnection of components forming a system configuration that will provide a desired system response (Dorf and Bishop, 2011). To achieve this goal, closed–loop (feedback) systems are adopted, in which the process outputs, whose behavior needs to be controlled, are measured through some *sensors* and compared with desired output values. The result of this comparison is fed into a *controller* which, through the execution of a control algorithm, computes the appropriate values of some control variables, which are applied to the controlled systems through appropriate *actuators*.

In most modern industrial applications, the controller is typically implemented on a computing platform, thus exploiting results from discrete control theory, even when the variables to be controlled are continuous (Åström and Wittenmark, 1997). An overview of



(a) Block diagram of a discrete–time control system

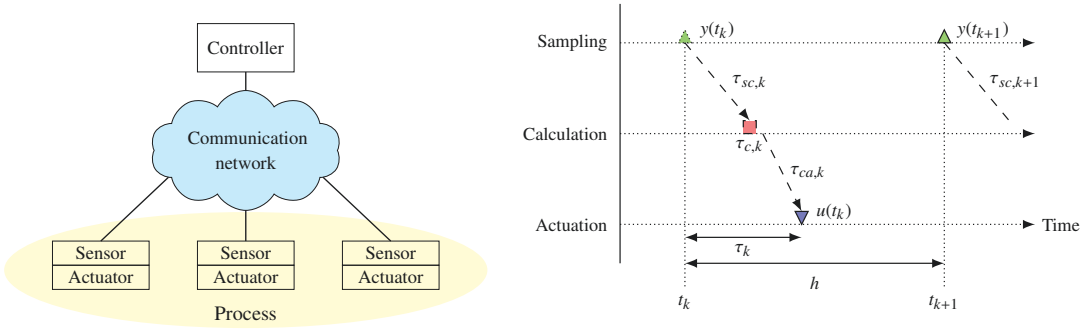
(b) Timing diagram of a discrete–time control system

Figure 2.1: Architecture of a discrete–time, closed–loop control system, adapted from (Marti et al., 2005).

a computer–based system used to control a continuous–time physical process is provided in Fig. 2.1. It can be observed that the control action is divided in three consecutive steps: (i) *sampling*, i.e., the measure of the process outputs to control, that are converted in digital form through an Analog–to–Digital Converter (ADC); (ii) *calculation*, i.e., the execution of a control algorithm which allows to derive the (discrete) values of the control variables; (iii) *actuation*, i.e., the conversion of control variables in analog form through a Digital–to–Analog Converter (DAC) and their application to the system. These steps are repeated in a periodic way, and the interval between two consecutive sampling instants (indicated with h in Fig. 2.1b) is called *sampling period*.

In Fig. 2.1b it can be noticed that the sampling, calculation and actuation phases are performed simultaneously. While this is always an approximation, in traditional control systems architectures it was a realistic one. Indeed, the earliest version of industrial control systems were based on point–to–point architectures, were sensors and actuators are directly wired to the controller using, for example, a 4 to 20 mA current loop (Marti et al., 2005). In these architectures, the delays between the different phases were deterministic and very short, especially if compared to the large sampling periods of traditional control applications. However, point–to–point architectures offer little flexibility, being difficult to reconfigure if new sensors/actuators need to be integrated in the system, and also do not allow remote diagnosis and maintenance. For this reason, NCSs have been introduced (Ray, 1989) and are nowadays the predominant choice for industrial control applications (Chow and Tipsuwan, 2001).

The typical architecture of NCSs can be seen in Fig. 2.2a: sensors, controller and actuators are all interconnected through a communication network, which is used to distribute sampling data from the sensors to the controller and control values from the



(a) NCSs architecture **(b)** NCSs timing

Figure 2.2: Networked Control Systems: schematic architecture and timing diagram.

controller to the actuators. In this sense, the NCS architecture is a completely distributed and flexible one, where components can be removed and added seamlessly by simply connecting and disconnecting them from the network. Moreover, the network can allow also supervisory devices to connect, thus permitting remote diagnosis and maintenance, possibly without stopping the operation of the system (a key feature for many industrial applications).

A typical drawback of the NCS architecture is that it is no longer safe to assume that sampling, calculation and actuation happen almost simultaneously. Indeed, these operations depend on the exchange of data packets, which may be delayed, arrive out–of–order or not arrive at all, according to the properties of the communication network and to contingent situations, such as the surrounding environment. The delay and loss of packets can in turn lead to a situation in which the temporal and spatial consistency of control variables is no longer ensured (Willig et al., 2005) and to the ultimate failure of control strategies, no matter how robust they are.

To evaluate the impact of a communication network on the performance of a control system, it can be useful to refer to Fig. 2.2b where, for the sake of simplicity, a system with one sensor, one controller and one actuator is considered. Taking into account the k–th sampling period, $\tau_{sc,k}$ indicates the delay between the sampling instant at the sensor and the delivery of the corresponding packet containing the sampled data at the controller, whereas $\tau_{ca,k}$ indicates the delay between the generation of the control variable value at the controller and its successful delivery at the actuator. These two delays are mostly related to the communication network performance, unlike the controller delay $\tau_{c,k}$, which represents the execution time of the control algorithm. Together, these delays

form the *time delay* at the k–th sampling period

$$\tau_k = \tau_{sc,k} + \tau_{c,k} + \tau_{ca,k} \tag{2.1}$$

It must be stressed that the time delay τ_k is in general time–variant, as the process of delivering a packet over a communication network is hardly a deterministic one. Moreover, the absolute value of time delay as well as its variance increase if the number of sensors/actuators is high, as it is the case of many industrial NCSs.

2.2 Communication requirements for industrial NCSs

Following the description of NCSs provided in Sec. 2.1, three fundamental requirements for the communication network used to interconnect sensors, controller and actuators may be stated:

1. The time delay τ_k must be kept as low as possible at each sampling period. Ideally, it should be $\tau_k \simeq 0$, to approach closely the performance of point–to–point control architectures. In any case, in practice it must be guaranteed that the time delay does not exceed the sampling period, i.e.

$$\tau_k \leq h, \quad \forall\, k \tag{2.2}$$

in order to avoid the overlapping of different sampling procedures, as it can be observed in Fig. 2.2b. In general, the longer the time delay, the higher the deviation between the desired system response and the actual one (Marti et al., 2005). From the perspective of the communication network, this requirement is equivalent to demand *low–latency packet delivery*.

2. The variance of the time delay τ_k across different sampling periods must be as reduced as possible. Ideally, it should be

$$\tau_k \simeq \tau, \quad \forall\, k \tag{2.3}$$

i.e., the time delay should be time–invariant. If this is the case, this quantity can be integrated in the controller design and its effects can be practically canceled. A similar result is obtained if τ_k varies with time but its deviation from a nominal value τ is bounded: in this case, robust control methodologies can be applied (Zhong, 2006). Conversely, if the deviation of τ_k from its nominal value is unbounded, the controller performance are degraded significantly, possibly leading to instability

(Marti et al., 2005). From the perspective of the communication network, this requirement is equivalent to demand *deterministic packet delivery*, whereas the deviation of time delay from the nominal value is often referred to as *jitter* on periodic operations.

3. All packets containing sampled data and actuation commands must be successfully received. Indeed, if a packet is lost, the corresponding data control variable is not updated, having an effect comparable to a time delay of one sampling period. Again, robust control strategies can be designed if the packet loss probability across the network is bounded (Xiong and Lam, 2007). From the perspective of the communication network, this requirement is equivalent to demand *reliable packet delivery*.

Low–latency, determinism and reliability are hence the key properties required to a communication network employed in industrial control applications. The term ICNs is used to indicate networks that satisfy these properties (Decotignie and Pleinevaux, 1993). Very often, the same networks are also referred to as "Real–time Networks", a term which emphasizes the deterministic requirement but that is often adopted to indicate a network meeting also the low–latency and reliability requirements and used in industrial applications (Decotignie, 2005a).

Example requirements for specific scenarios

While the above mentioned requirements of low–latency, determinism and reliability are typical of all industrial control applications, the degree to which they must be respected varies greatly among different application scenarios. Specifically, the maximum time delay is related to the sampling period, as highlighted in Eq. (2.2), which in turn is linked to the natural frequency of the physical process to control (Ogata, 1995). Similarly, the maximum tolerable jitter is often defined as a percentage of the nominal time delay and, hence, also the required level of determinism strongly depend on the underlying application. Finally, the reliability level (i.e., the percentage of packet loss tolerated) may also vary significantly, depending on the robustness of the control system and on the criticality of the application.

In order to provide some examples of how these requirements can vary, five different categories of applications can be defined (Luvisotto et al., 2017a): Building Automation (BA), Process Automation (PA), Factory Automation (FA), Power Systems Automation (PSA), and Power Electronics Control (PEC). The first one is relevant to control operations performed within houses and public/private buildings, such as lighting, heating,

Table 2.1: Example requirements for different industrial control applications, adapted from (Pang et al., 2017).

Scenario	Typical sampling period	Number of nodes	Reliability level
BA	10 s	10^3	medium
PA	100 ms	10^4	medium
FA	1 ms	10^2	high
PSA	100 μs	10^2	high
PEC	10 μs	10^2	very high

surveillance, etc. (Zhu et al., 2016). PA is instead involved with process industries, such as chemical, mining, oil and gas, and others (Yu et al., 2014). FA generally indicates all the procedures carried out in a production line, including assembling, packaging, and palletizing (Orfanus et al., 2013; Vitturi et al., 2011). The term PSA, instead, refers to controlling the generation, transmission and distribution of electrical power (Feliciano et al., 2014; Yang et al., 2012; Ngo and Yang, 2016). Finally, in PEC the focus is on the synchronized control of power electronics devices (Cottet et al., 2015; Zheng et al., 2012; Toh and Norum, 2013).

Tab. 2.1 reports some examples of communication requirements for the aforementioned application scenarios. In detail, the sampling period requirement can be easily translated in a constraint on time delay, according to Eq. (2.2), and it can be seen that it ranges from several seconds in BA applications down to some μs in PEC. In terms of determinism, it is safe to consider a jitter constraint to be in the order of 10% of the sampling period. The reliability level is indicated in a qualitative way, as it is difficult to state absolute values given that the required level of Packet Error Rate (PER) is strongly dependent on how much the overlying control application is robust to the occurrence of communication faults. However, to give a practical perspective, PEC applications that demand a very high reliability are typically served by optical fiber links, whose PER is in the order of 10^{-9} (Gerlach-Erhardt, 2009). Finally, the number of nodes (that can be sensors, actuators or both) is also an important parameter to evaluate the complexity of the network infrastructure, and it may range from 10-100 nodes in FA, PSA and PEC applications to large installation with 10000 nodes in PA.

Peculiarities of industrial communication networks

ICNs are defined by the low–latency, determinism and reliability requirements, that complicate their design significantly with respect to traditional communication networks employed in home/office environments. At the same time, though, the task of designing

communication networks to be employed in industrial control applications takes benefit from some peculiar assumptions that are generally verified in these applications, unlike in traditional communication networks.

A first, important, assumption is related to the nature of the exchanged data. As reported in Sec. 2.1, ICNs are generally used to transfer the sampled values of process outputs, as measured by the sensors, and the updated values of control variables, as computed by the controller. These quantities can generally be written in a few bytes (Willig et al., 2005), leading to very *short packet sizes* compared to traditional communication applications (e.g., video streaming, Dykstra (1999)).

Another important feature of ICNs can be derived from the description of NCSs in Sec. 2.1: since the sampling instants are fixed, the type and amount of data exchanged is constant over time and the computation time τ_c is also constant, the *traffic flows of the network are to a large extent predictable*. This means that, before the network is deployed, the instants at which the different nodes will have data to send are known with precision, unlike traditional networks where traffic flows often depend on human behavior and can only be approximated through statistical distributions (Chlebus and Divgi, 2007).

Furthermore, a closer look at the architecture of NCSs reveals that not only the packet flows are predictable, but *most of the traffic is cyclic*: the pattern of data exchanged over the network repeats itself with a periodicity given by the sampling period h. This observation simplifies significantly the design of an ICN and, in particular, of its Medium Access Control (MAC) layer protocol. Indeed, a very common choice is to adapt a Time–Division Multiple Access (TDMA) scheme, where time is divided in slots, each slot is assigned to the transmission of one packet and this schedule is repeated periodically. An example of such a scheme can be observed in Fig. 2.3 for a NCS composed by n sensors and n actuators. The network schedule is repeated periodically, with a period (often called cycle time) equal to the sampling period of the NCS. It can be noticed that the schedule is not entirely filled with transmission slots, but includes some idle time. This time can be used to perform retransmissions of lost packets (when needed) or to transmit acyclic data. Indeed, even if the majority of traffic is cyclic, an ICN can be characterized also by the presence of *important acyclic packets* like alarms, that need to be transmitted reliably and with bounded latencies (Willig et al., 2005). The presence of some idle slots in the schedule guarantees that possible alarms can be transmitted within one network cycle, i.e., with a latency bounded by the sampling period h.

A further key difference between ICNs and some kind of traditional communication networks (e.g., mobile networks) is that the *deployment is typically static*, i.e., the nodes

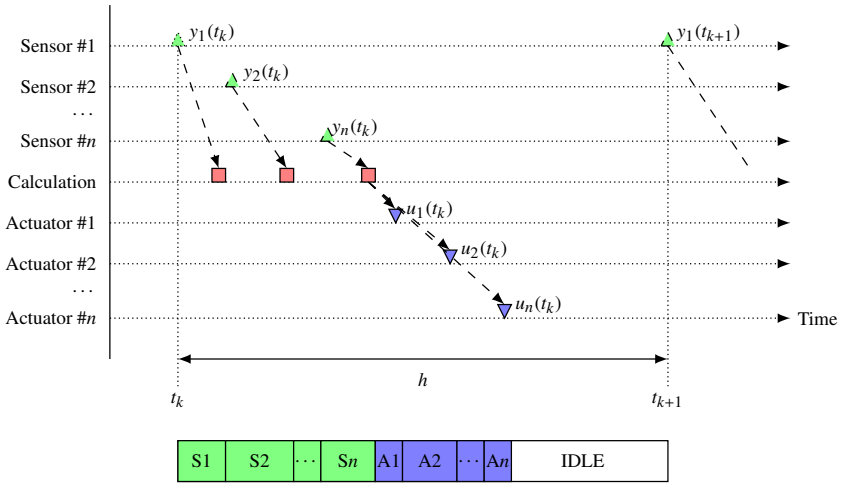


Figure 2.3: Example of TDMA schedule in a NCS with n sensors/actuators.

do not move during network operations.[1] This feature allows to avoid many significant problems that are encountered in mobile networks, such as the handover between cells and the need to equalize a rapidly time–varying communication channel (Dahlman et al., 2013). A statical deployment also means that, in general, the number of the nodes in the network is fixed, and there is no need to deal with frequent joining and leaving procedures by the nodes, as it may happen in traditional networks. However, ICNs must allow the possibility to insert or remove nodes if needed, since this flexibility is a key feature of NCSs, as discussed in Sec. 2.1.

Finally, the architecture of NCSs, as visible in Fig. 2.2a, determines another key feature of the underlying network, i.e., its logical topology. Indeed, since the system includes a controller and some distributed sensors/actuators, the network will always have a *logical star topology*, even if the physical topology can be of a different type (e.g., ring or bus). The central controller will generally take care also of all the functions related to the control and management of the network (Luvisotto et al., 2017a), such as the generation, update and broadcasting of the traffic schedule. In large–scale applications there might be multiple controllers, hence the network can be split in several subnetworks, each one handled by a single controller, with a backbone network connecting the different controller and possibly a master controller serving as central control point.

Tab. 2.2 summarizes the main features of ICNs discussed in this section. To conclude this analysis, it must be observed that low–latency, determinism and high reliability

[1] In some industrial applications (e.g., motion control and robotics) the nodes can indeed move; however, the movements are generally constrained along some axes and are limited in total distance.

Table 2.2: Main features of ICNs.

Requirements	Assumptions
Low latency	Short packet size
Determinism	Predictable traffic flows
High reliability	Static deployment
	Logical star topology

are not the only key requirements when evaluating an ICN. Indeed, there are many other factors that, although not dealt with specifically in this thesis, often determine the success or failure of ICN solutions. A brief outline is given in the following.

- *Cost*: the major incitement for the adoption of one network solution over another one in industrial automation is how much it costs to a company (Åkerberg et al., 2011). The design, commissioning, installation and maintenance costs of an ICN are hence key factors that should be quantified and optimized carefully to guarantee its success in the competitive industrial market. Furthermore, since industrial control systems are designed to be operative continuously and to last for many years, the robustness over time of a network, and its ability to be repaired or upgraded without interrupting its operations, are key factors that result in significant economic advantages.

- *Security*: with the continuous integration between ICNs and other networks, their security is at risk, exposing them to eavesdropping and Denial–of–Service (DoS) attacks among other threats (Dzung et al., 2005). Several solutions can be designed to enhance security, be it at the device level (e.g., encryption and authentication) or at system level (e.g., transition of the system to a safe state whenever an attack is detected) (Willig, 2008). However, almost every solution introduces an overhead and hence should be traded off with performance requirements, such as low–latency and reliability.

- *Safety*: safety of humans, environment and property is a top priority in industrial workplaces (Åkerberg et al., 2011). Consequently, automation equipment, including networks, must be designed to reduce the risk of uncontrolled or dangerous situations. In this sense, it is crucial that the components of a control system can detect if a communication fault has happened and eventually transition to a safe state to prevent hazards and damages.

- *Energy efficiency*: all the nodes involved in an ICN consume a great deal of

energy in transmitting and receiving data, besides the one they spend for other purposes such as sensing, computing and actuating. Although the energy used by communication components can be significantly lower than that employed by other industrial devices (e.g., motors and drives), the need for green communications has started to emerge recently, with the application of Energy–Efficient Ethernet (EEE) solutions to ICNs (Tramarin and Vitturi, 2015). When it comes to wireless ICNs, energy efficiency is even more important, as the removal of cables means that nodes are typically battery–powered and, hence, should be parsimonious with energy consumption, in order to enhance their lifetime (Willig, 2008). Several solutions can be designed, ranging from low–power protocols characterized by small duty cycles (Karl and Willig, 2007) to enhancements of battery life through wireless power transmission (Hirai et al., 1999) and energy harvesting (Kansal et al., 2007).

2.3 Wired industrial communication networks

The introduction of NCSs in industrial automation has been linked since the beginning with the idea of making the process data available across all company levels (Sauter, 2010). Consequently, ICNs have never been considered as a standalone solution, but rather integrated with several other (already existing) layers. Several attempts to model the different enterprise layers and their interactions have been made since the 1970s, with the emergence of Computer–Integrated Manufacturing (CIM), when the enterprise was modeled as a pyramid (Sauter, 2007).

A complete representation of the so–called "automation pyramid" as initially developed can be seen in Fig. 2.4. It can be observed that the pyramid comprises several levels: company, factory, shop floor, cell, process and field level. Nowadays, this complex structure has been mainly reduced to three levels: company level, cell level and field level (Sauter, 2010). The highest layer is the place where business systems like Enterprise Resource Planning (ERP) and Supply Chain Management (SCM) reside and is generally better served by Wide Area Networks (WANs), able to interconnect different premises of a company distributed around the world and/or companies with suppliers and customers. The cell level is dedicated to the supervision and optimization of production processes, through tools like Manufacturing Execution Systems (MESs) and Supervisory Control and Data Acquisition (SCADA) systems, and communication is generally carried out through Local Area Networks (LANs) deployed within a single industrial building. Finally, the field level includes sensors, actuators and controllers that allow to actually control the industrial process and communicate through the ICNs that represent the subject of

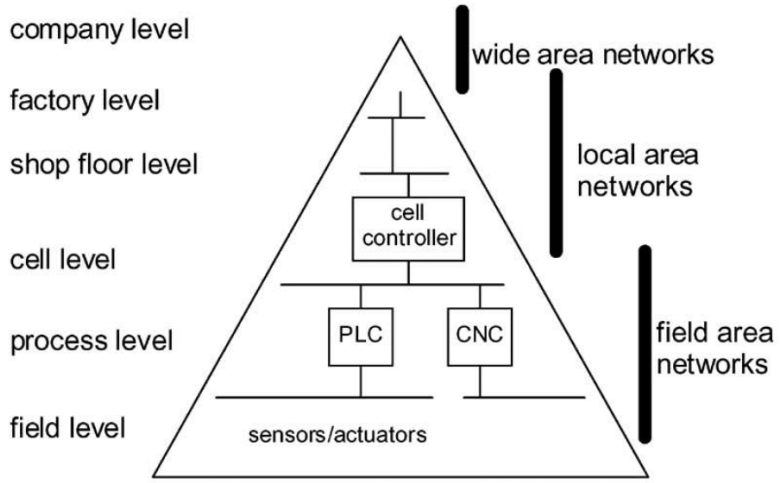


Figure 2.4: A schematic representation of the automation pyramid as initially developed in the 1970s, taken from Sauter (2010).

this thesis. It must be observed that the closer an application is to the bottom of the pyramid, the higher the low–latency, determinism and reliability constraints apply to it; conversely, the amount of data transferred increases with the level in the pyramid (Decotignie, 2005b).

Focusing on field–level networks, the initial approach was to develop dedicated solutions that offered the required properties, which were generally termed as *fieldbuses* (Thomesse, 2005). Subsequently, networks originally developed for home/office communications have started to being used also at the lowest levels of the automation pyramid, although with specific modifications to guarantee the aforementioned requirements, giving birth to a wide range of *real–time Ethernet* solutions (Decotignie, 2005a).

Fig. 2.5 reports a schematic representation of the most important networking solutions for industrial applications in the last 50 years, together with parallel trends in related fields, such as traditional computer networks. Fieldbus systems, represented in yellow, and RTE networks, represented in cyan, will be discussed in details in the rest of this section, whereas wireless networks (depicted in green) will be treated in the following section.

Fieldbus systems

The term "fieldbus" is defined in the International Electrotechnical Commission (IEC) 61158 standard in the following way: "A fieldbus is a digital, serial, multidrop, data bus for communication with industrial control and instrumentation devices such as, but not

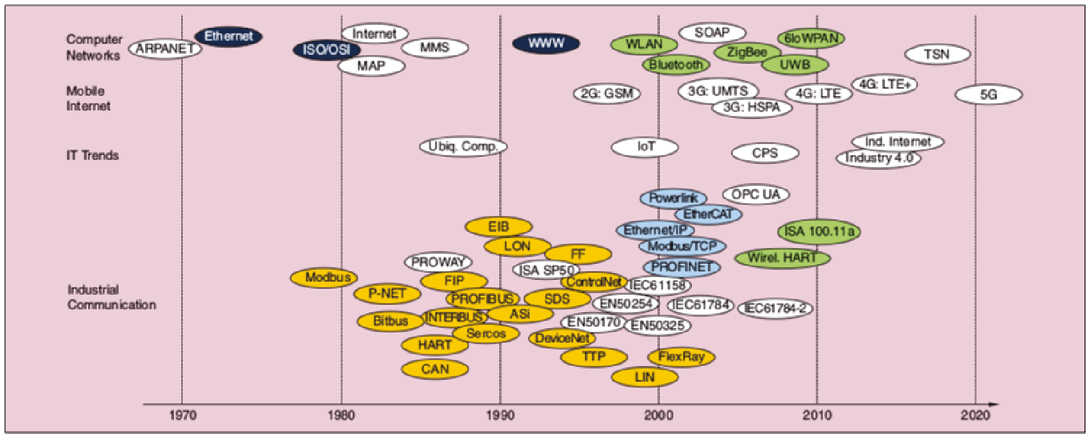


Figure 2.5: Milestones in the evolution of industrial communications and related technology fields, taken from Wollschlaeger et al. (2017).

limited to, transducers, actuators and local controllers" (IEC 61158-2003).

Fieldbuses were initially developed to replace the point–to–point connections between controllers and sensors/actuators with a single bus, thus marking the beginning of NCSs in industry, yielding obvious advantages as flexibility, modularity and easier system configuration and maintenance. The introduction of fieldbuses at the bottom level of the automation pyramid filled the gap with the upper layers, which already had dedicated networks while the field level was still relying on point–to–point connections (Sauter, 2010).

Although the first "real" fieldbus is the Military Standard 1553 bus, released in 1970, the most successful solutions were developed in the 1980s, with the introduction of Modbus and industrial protocols based on Controller Area Network (CAN) (such as, for example, DeviceNet), among the others. Characteristic properties of these systems were the serial transmission of control and data over the same line, a master–slave structure and a focus on the two lowest layers of the ISO/OSI protocol stack (namely physical and MAC layers), to guarantee the fulfillment of the required properties for industrial communications (Sauter, 2010).

During the 1980s, the golden age of fieldbuses, several different systems were developed, all tailored to specific application scenarios and with peculiar features. The most important example is the approach used by different fieldbus systems to access the shared communication channel, which range from polling protocols (e.g., Modbus), token–passing schemes (e.g., P-NET and ControlNet), TDMA strategies (e.g., INTERBUS and Sercos) and random access methods (e.g., CAN) (Sauter, 2010).

Unfortunately, all these different solutions offered little to none interoperability,

leading to integration problems and ultimately compromising the key advantage brought by NCSs (flexibility), with the result of possibly losing customers trust and market share. In order to avoid these drawbacks, an effort began in the end of the 1980s to develop and promote universal definitions and profiles to be shared by different fieldbus systems. An international standardization effort also began, with the creation of the IEC technical subcommittee SC65C on fieldbuses in 1985. The goal of this committee was to develop a single fieldbus standard based on the two most promising approaches back then, namely PROFIBUS and Factory Instrumentation Protocol (FIP). However, the enormous investments for existing systems already in place as well as the different requirements of specific application fields crippled these efforts, with the result that IEC 61158 was released as a multiprotocol standard (IEC 61158-2003) that encompasses several different fieldbus systems without defining a universally accepted solution (Sauter, 2010).

Real–Time Ethernet networks

RTE networks, often termed also as "Industrial Ethernet" networks, were introduced towards the turn of the century with a clear goal in mind: allow an easier integration between the field level (served by dedicated fieldbus systems) and the upper levels in the automation pyramid (served by Ethernet–based LANs) (Sauter, 2010). Moreover, emerging application scenarios in industrial automation, characterized by ever lower sampling periods, required faster communication technologies with respect to legacy fieldbus systems, and the very high data rates offered by Ethernet looked appealing to this regard. Indeed, the IEEE 802.3 standard that defines Ethernet is always increasing the available transmission rate, from the 10 Mbps of 10BASE-T physical layer (1990) to the 10 Gbps of 10GBASE (2006), and it is currently working to increase it over 100 Gbps (IEEE 802.3-2015).

However, high data rate alone is not sufficient to guarantee the low–latency and determinism required by NCSs and, initially, the use of Ethernet in this context was prevented by its usage of a non–deterministic channel access method, namely Carrier Sense Multiple Access (CSMA). A breakthrough was the introduction of switching and full–duplex technology, that eliminated the collision problem almost completely and allowed simultaneous transmission and reception of frames (Skeie et al., 2006). Building on these features, the development of field–level networks based on Ethernet began, with the hope of overcoming the painful division between multiple solutions that characterized the fieldbus landscape.

Unfortunately, the ideal of a unique RTE solution was soon defeated by a variety

of approaches and solutions, that barely shared the use of IEEE 802.3 as a low–layer technology, given that some networks use dedicated switches and controllers (Sauter, 2010). Several other differences exist in the data–link, network, transport and application layers, with the result that full interoperability between different RTE solutions is actually a chimera. On the other hand, the compatibility between new RTE networks and legacy fieldbuses is almost always guaranteed, since often the producers of the former were also the companies that developed the latter, and the provision of a seamless migration path to their customers was a top priority. Several approaches to reach this compatibility can be pursued: full–compatibility of high layer protocols (e.g., Ethernet/IP uses the Common Industrial Protocol (CIP) application protocol, common to ControlNet and DeviceNet), compatibility of data models and objects (e.g., PROFINET adopts proxy solutions to incorporate legacy PROFIBUS devices), usage of application layer profiles (e.g., EtherCAT uses the CANopen application layer) (Sauter, 2010).

While the lack of interoperability between different RTE solutions can represent a big problem, a key achievement of this technology has been the introduction of the Transmission Control Protocol (TCP)/Internet Protocol (IP) suite in modern industrial networks, that effectively allowed an easier integration across the different levels of the automation pyramid. Indeed, almost all RTE solutions consider the presence of a non real–time channel where configuration or diagnostic information can be exchanged through transport layer protocols such as TCP or User Datagram Protocol (UDP). The coexistence of this type of traffic with the traffic generated from sensors, controller and actuators and subject to low–latency and real–time constraints can be achieved through dedicated slots in a TDMA schedule or through the frame prioritization feature of IEEE 802.3 (Sauter, 2010). In both cases, if the system is properly configured, the performance figures of the NCS are not affected by the non real–time traffic.

2.4 Wireless industrial communication networks

The next logical step in the evolution of field–level networks, after fieldbus systems and RTE networks, is represented by wireless networks (Sauter, 2010). Indeed, in the world of home/office communications, wireless solutions have replaced wired ones in all the cases where wiring was problematic (e.g., mobile communications) and several different wireless networks to cover all possible application scenarios have been developed, offering performance in pair with those of wired ones in terms of data rate.

When it comes to industrial communications, the introduction of wireless networks would bring several advantages to their users. First, a great cost reduction is envisioned,

mainly due to the lower cost of materials: considering a green field installation, deploying wires requires roughly $200 per meter indoor and $1000 per meter outdoor (Åkerberg et al., 2011). Moreover, the replacement of cables with wireless links would imply easier design, installation and maintenance, causing further economic savings (Luvisotto et al., 2017a). A second benefit, strongly related to the first one, is that the lower costs of installation would allow to connect sensors that were previously unwired for economical reasons, thus enhancing the overall performance of the underlying NCSs, as well as unlocking a whole new range of applications not feasible with wires, such as those characterized by large heights, high temperatures, mobile environments and rotating parts (Luvisotto et al., 2017a). Moreover, temporary measurements of specific process values would be feasible without significant effort (Åkerberg et al., 2011). Finally, although wireless links are more likely to encounter temporary failures, they offer a higher long–term reliability with respect to wires, that will age and break, especially if employed in motion control (because of wear and tear) or power electronics (because of high potential differences) applications (Luvisotto et al., 2017a).

Despite all these potential advantages, the road towards the introduction of wireless networks in industrial automation has been bumpy and, to this day, wireless ICNs are not yet established in the same way as RTE networks are. The reasons behind these struggles are manifold, but they can all be traced back to one "original sin", i.e., the error–prone nature of the radio channel (Willig et al., 2005). The poor reliability of this transmission medium is related to a series of physical phenomena, which are briefly described in the following.

- *Path loss*: the signal strength of a radio signal in free space decays with the distance according to a power–law, thus making long–range communications very difficult. Several parameters influence the path loss trend and a general model can be represented as (Willig et al., 2005)

$$P_{rx}(d) \sim P_{tx} \cdot \left(\frac{d_0}{d}\right)^{\gamma} \tag{2.4}$$

  where P_{tx} and P_{rx} are the transmitted and received power, respectively, d is the communication distance, d_0 a reference distance and γ is the path loss exponent, which is environment–related and generally ranges between two and four (Rappaport, 1996).

- *Multipath fading*: while path loss can be attenuated by increasing the transmitted power, as visible in Eq. (2.4), there are other impairments related to the fact that multiple copies of the transmitted waveform are received at different times due to

reflections, diffraction or scattering in the environment (Rappaport, 1996). The constructive and destructive interference of these waveforms result in fluctuations in the received power known as fading, which can also vary extremely fast in time if the nodes or the environment moves. Moreover, when the delay between different copies of the same waveform (called "delay spread") is very large, it may happen that consecutive transmitted symbols overlap at the transmitter, causing Inter–Symbol Interference (ISI) (Willig et al., 2005).

- *Shadowing*: another typical impairment is represented by obstacles that block the main signal path between transmitter and receiver. Depending on the nature of the obstacle and on the frequency band, the signal can be completely blocked or partially penetrate. In the former case, the received power will be altered with respect to the nominal value and this phenomenon is called shadow fading or shadowing (Rappaport, 1996). The power loss also changes over time, but slower than what happens with multipath fading.

- *External interference*: a key difference between wireless and wired networks is that the former uses a medium (the radio channel) which is shared among everyone that transmits at a given frequency. For this reason, different networks deployed in the same portion of the frequency spectrum may interfere among them, resulting in a mutual disturb or in a prevalence of one over the other depending on the transmitted powers and modulation properties (Rappaport, 1996).

All these phenomena result in bit errors (i.e., bits that are flipped at the receiver) or packet losses, with the latter that can be caused either by a burst of bit errors or (more likely) by failures in the detection and time synchronization of packets. The resilience of wireless transmission to channel impairments is strongly influenced by the system characteristics (e.g., transmit power, antenna gains, etc.) as well as by the modulation properties. A general rule of thumb is that the higher the data rate, the less robust the transmission.

The first negative consequence of bit errors and packet losses is, clearly, a drop in the high reliability required by ICNs, that can lead to loss of space/temporal consistency of process variables and to the ultimate instability of the system under control (Marti et al., 2005). However, these errors can indirectly affect also the other requirements. For example, the presence of a noisy channel forces the use of long preamble sequences to acquire carrier/bit synchronization and to estimate the channel response in order to equalize the received data, thus causing a high overhead which impacts on the low–latency requirement (Willig, 2008). Moreover, packet losses are often dealt with Automatic Repeat

Request (ARQ) mechanisms, i.e., retransmitting the same packet until it is received, which can lead to both an higher latency and uncertainties in packet delivery time, thus impairing determinism. Several other techniques have been developed to cope with channel errors, such as Forward Error Correction (FEC), Hybrid Automatic Repeat Request (HARQ), spatial diversity and cooperative diversity, to name a few, but they all impact latency and determinism to a certain degree (Luvisotto et al., 2017a).

Finally, the shared and error–prone nature of the wireless channel affects also the design of MAC protocols. For example, in token–passing protocols the token can be lost due to a deep fading situation, excluding stations from the logical ring (Willig, 2008). Random access protocols, such as CSMA, are often used in shared mediums but they suffer from the occurrence of collisions that can compromise communication, and the usage of traditional deterministic collision–resolution strategies, such as the one used by CAN, is often not feasible due to the half–duplex constraint of wireless transceivers (i.e., they can not simultaneously transmit and receive in the same frequency band) (Willig, 2008). Another significant issue of CSMA–based wireless networks is the well–known hidden terminal effect (Tobagi and Kleinrock, 1975): if two nodes, that are outside each other's sensing range, want to send a packet to a node that is in the range of both at the same time, they sense the channel as idle and hence transmit, leading to a collision.

In spite of all these issues, the research on wireless networks to be used in industrial applications is stronger than ever and several solutions have been already deployed successfully.

Home/office wireless standards

Similarly to what happened with Ethernet, the dominant trend towards the realization of wireless ICNs has been to avoid a complete redesign of the protocol stack and rather reuse proven standards from home/office communications for the lower layers (Sauter, 2010). From there, two different approaches can be followed depending on the application scenario: i) these standards can be used "standalone" with just an industrial application layer on top that carries out the required control task; ii) industrial wireless networks can be realized by designing a set of dedicated middle layers, that can extend from the application down to the MAC layer. Some solutions that follow the latter approach are discussed in the next subsection, whereas here the most interesting home/office wireless standards are briefly listed (for a detailed overview on some of these standards please refer to Chap. 3).

- *IEEE 802.11*: this standard (IEEE 802.11-2016) defines WLANs, also known in the consumer market with the term "Wi-Fi", and it is regarded as the wireless

equivalent of Ethernet.

- *IEEE 802.15.4*: this standard (IEEE 802.15.4-2015) defines Low Rate–Wireless Personal Area Networks (LR-WPANs), also known in the consumer market and scientific literature as Wireless Sensor Networks (WSNs). It is part of the broader IEEE 802.15 family of standards, which defines WPANs in general.

- *IEEE 802.15.1*: this standard (IEEE 802.15.1-2005), also part of the WPANs, is known in the consumer market and scientific literature as "Bluetooth".

- *IEEE 802.15.3*: this standard (IEEE 802.15.3-2016) defines High Rate–Wireless Personal Area Networks (HR-WPANs) and it is often referred to as Ultrawide Band (UWB), even though this term can be used also for other technologies that make use of very large bandwidth and high frequencies.

- *Mobile networks*: these networks, often referred to as "cellular networks", are standardized by the Third Generation Partnership Project (3GPP) group, which publishes a new release almost every year. The releases are then grouped in "generation" of networks (almost one every decade): the Second Generation (2G) included Global System for Mobile communications (GSM), General Packet Radio Service (GPRS) and Enhanced Data rates for GSM Evolution (EDGE); the Third Generation (3G) included Universal Mobile Telecommunication System (UMTS) and other related standards; the Fourth Generation (4G) included the Long–Term Evolution (LTE) standard; and, currently, the Fifth Generation (5G) is being standardized.

It must be observed that all these networks are not specifically designed for industrial communications and, hence, additional design efforts are required to ensure that the low–latency, determinism and reliability requirements are met.

Dedicated industrial wireless networks

Some complete solutions for the realization of wireless ICNs have been proposed over the years and started to be used in the market. Although they are always based on one of the discussed wireless standards, they come with their own recognizable name and are often standardized as well.

The most significant example is arguably that of *WirelessHART* (WirelessHART) and *ISA 100.11a* (ISA-100.11a-2009), the two leading standards in the context of Industrial Wireless Sensor Networks (IWSNs). The protocol stack of these two standards is quite

similar (Petersen and Carlsen, 2011), as they are both based on the IEEE 802.15.4 standard for WPANs, deployed in the 2.4 GHz Industrial Scientific and Medical (ISM) band with a data rate of 250 Kbps. Both networks are based on TDMA, with a fixed slot length in WirelessHART (10 ms) and a variable one (also lower bounded by 10 ms) in ISA 100.11a. Slots can be dedicated to a specific node or shared, and CSMA/CA is adopted to regulate access in the latter case. Besides TDMA, both standards adopt a pseudorandom frequency hopping scheme, in which nodes change the transmission channel at any given slot (possibly with simultaneous slots being used in different channels) to improve the communication robustness against external interference and fading. Moreover, retransmissions in case of failures are carried out in the next available slot (Vitturi et al., 2013). Finally, both standards implement a form of spatial diversity known as "path diversity", meaning that, for each link between two nodes, a graph of possible paths is defined offline and different paths can be used in case of failures.

Another well–known industrial wireless solution is *Wireless Interface for Sensors and Actuators (WISA)* (Scheible et al., 2007), based on the IEEE 802.15.1 physical layer. In this system ,up to 120 slaves can be connected to a WISA Base Station (BS), forming a WISA cell, and different cells can be connected via a fieldbus or RTE backbone. At the MAC layer, WISA also uses a combination of frequency hopping and TDMA, with a 2.048 ms long cycle containing 30 slots, each allowing up to 4 uplink transmissions (from slaves to BS) and 1 downlink transmission (from BS to slaves) in parallel frequency channels. Lost packets are retransmitted in the following frames and reliability can be improved by means of spatial diversity techniques through the use of MIMO at the BS (Vitturi et al., 2013).

Besides these solutions, several other wireless ICNs are standardized and commercially available, such as the Wireless networks for Industrial Automation – Process Automation (WIA-PA) and Wireless networks for Industrial Automation – Factory Automation (WIA-FA), proposed by the chinese Shenyang Institute of Automation and recently approved as IEC standards (IEC 62601-2015; IEC 62948-2017). Moreover, a plethora of different proposals can be found in scientific literature but are not yet realized in the form of commercial products. As a very important example, the RT-WiFi protocol (Wei et al., 2013) is a real–time protocol for NCSs based on the IEEE 802.11 standard, that implements a TDMA where the slot time can be as low as 200 μs and whose good performance figures have been validated through experiments.

Hybrid wired/wireless networks

The difficulties encountered by industrial wireless networks in their establishment lead to think that completely wireless architectures will never be the dominant solution in industrial applications, due to their reliability concerns. In this context, the idea of hybrid networks, able to combine the reliability of wired links with the flexibility offered by wireless, may represent a reasonable compromise and is perhaps better suited as the future of ICNs (Sauter, 2010).

In an hybrid network, a wired field–level backbone (likely based on Ethernet) is deployed and connects several wired nodes but also some Access Points (APs) to which other nodes are connected via wireless links. The interaction between the two networks (wired and wireless) is easy if wireless segments are autonomous islands that do not need to exchange data in real–time with the wired nodes or other wireless segments and an interconnection based on gateways is likely sufficient (Cena et al., 2008). However, if real–time data exchange is needed, the AP should act as a bridge and the operations on the two segments should be carefully synchronized, properly dimensioning the slots reserved to wireless nodes in the wired network cycle.

Several works concerning the wireless extension of wired networks can be found as scientific papers, addressing the technical feasibility of this task and assessing the achievable performance, as well as commercial products. Among the several technologies that can be used for wireless segments, IEEE 802.11 seems to be the preferred one (Cena et al., 2008). For example, IEEE 802.11–based extensions of fieldbus systems such as PROFIBUS (Lee et al., 2002) and DeviceNet (OMRON WD30-2002) have been considered. Analogously, wireless extensions of RTE networks have been studied, including PROFINET (Santandrea, 2006) and Ethernet POWERLINK (Luvisotto et al., 2017c).

3

Wireless Network Standards

The majority of industrial wireless solutions are based on an international standard, typically developed for home/office networks, that defines the lowest layers of the protocol stack (generally physical and MAC layer). It is then worth spending some time in reviewing the major features of some of these standards, with a specific focus on the IEEE 802.11 standard for WLANs and the IEEE 802.15.4 standard for LR-WPANs.

3.1 The IEEE 802.11 standard for WLANs

The IEEE 802.11 standard, released in 1997, deals with the PHY and MAC layer of WLANs. The original standard has been integrated by numerous amendments, each one identified by one or two letters. Every few years, all the amendments are merged into a new version of the standard, which supersedes the previous one. The last version of the standard has been released in 2016 (IEEE 802.11-2016), integrating amendments up to IEEE 802.11ad. Backward compatibility with earlier versions of the standard is always required when discussing amendments, hence there are some key features of IEEE 802.11 that are common to all the amendments. These features are discussed first, then a brief overview of the most important amendments is given.

General features of IEEE 802.11

Several network topologies are possible with IEEE 802.11. The most used ones are the *infrastructure* mode, where several nodes or Stations (STAs) are connected to an AP, and the *ad–hoc* mode, in which there is no AP and two or more STAs are connected together. The IEEE 802.11s amendment introduced a new topology, the *mesh* mode, in which a STA can be connected to its AP through other APs situated nearby.

The IEEE 802.11 MAC layer provides three main functionalities, namely access to the medium, fragmentation/defragmentation and Multi–rate Support (MRS). As far as the first functionality is concerned, although several access modes are defined in the standard and in some amendments, the most used one by far is the Distributed Coordination Function (DCF), which corresponds to an implementation of CSMA/CA and must be implemented by all IEEE 802.11-compliant devices.

In DCF, a STA that wishes to transmit first listens to the channel for an amount of time called Distributed Coordination Function Inter Frame Space (DIFS). If the channel is idle it can transmit, otherwise it waits until the end of the busy period. After that, the STA has to check again if the channel remains idle for the duration of a DIFS, then the Contention Window (CW) begins, when all STAs that have data to transmit can attempt to access the channel. To avoid collisions, each STA initializes (or resumes) a timer, whose duration is a random variable, called *random backoff time*. While this timer decreases, the STA continues to sense the channel, stopping the timer whenever it detects activity. When the timer expires, the STA gains access to the channel and can finally start to transmit.

The random backoff time is obtained by multiplying a fixed quantity (called slot time) for a a pseudo–random integer drawn from a uniform distribution over the interval $[0, CW]$. The parameter CW is set to CW_{min} initially and updated to $2CW+1$ after each failed transmission attempt. In order to verify the outcome of a transmission attempt, the receiver should send an Acknowledgement (ACK) frame upon the correct reception of a data frame (after a short time interval called Short Inter Frame Space (SIFS)). If the ACK is not received within a certain timeout, the value of CW is updated and the frame is retransmitted. This procedure is repeated for a maximum amount of 7 times, then, in case the transmission has not yet succeeded, the frame is permanently discarded.

Another important feature of the IEEE 802.11 channel access scheme is the Request–to–Send (RTS)/Clear–to–Send (CTS) mechanism, an optional feature that can be used to cope with the hidden terminal issue (Tobagi and Kleinrock, 1975). When this mechanism is activated, a STA that wishes to transmit sends a RTS frame to the potential receiver, which replies with a CTS frame. If the CTS frame is not received within a certain time

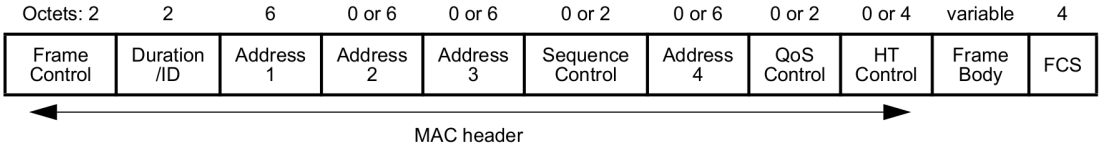


Figure 3.1: General format of a IEEE 802.11 MAC frame (IEEE 802.11-2016).

interval, the RTS frame is retransmitted using a random backoff procedure. Conversely, if the RTS/CTS exchange is successful, the transmission of data frame starts after a SIFS. The STAs not involved in the communication receive the RTS and CTS frames and update a Network Allocation Vector (NAV), which stores information about the next time instant in which the channel will be available. Clearly, this mechanism adds a significant overhead to the transmission of frames and it is therefore not indicated for low–latency applications such as most ICNs (Willig, 2008).

The other two key features of the IEEE 802.11 MAC layer are fragmentation and MRS. The first functionality allows to split a large information units in smaller frames, that can be sent consecutively separated by a SIFS, to enhance the overall probability of success.[1] MRS, on the other hand, stems from the fact that multiple modulation and coding options, each with a specific data rate and robustness level, are generally available at IEEE 802.11 STAs. Therefore, at each transmission attempt the STA can adapt the transmission rate to the channel conditions, although no default algorithm is defined in the standard, leaving its implementation to the producers of IEEE 802.11 devices.

Three types of MAC layer frames are specified in the IEEE 802.11 standard: data, control and management. A MAC frame, whose general format is reported in Fig. 3.1, is composed by an header, a frame body and a trailer. The frame body is variable in size and it is present only in data frames. The various fields of the header are not all mandatory and their presence depends on the amendment as well as on some configuration options (for a detailed overview please refer to the standard). Finally, the trailer contains a 32–bit Frame Check Sequence (FCS), used to check the integrity of received frames.

The IEEE 802.11a/b amendments

IEEE 802.11a and IEEE 802.11b are the two first amendments to the standard, published in 1999 and still supported by many devices currently on the market.

IEEE 802.11a offers eight different modulation and coding combinations, with data rates ranging from 6 to 54 Mbps. It is deployed on the 5 GHz ISM band, where

[1] The fragmentation feature is generally not used in industrial applications, as the original size of the information unit is already small, as discussed in Sec. 2.2.

several channels are available, each with a 20 MHz bandwidth. Through the use of Orthogonal Frequency–Division Multiplexing (OFDM), the channel bandwidth is split in 64 subcarriers, 48 of which are used for data transmissions. OFDM symbols are 4 μs long, including a Guard Interval (GI) of 800 ns, and each of them can carry a different amount of data (from 48 to 216 bits), depending on the selected modulation and coding.

IEEE 802.11b, instead, is deployed in the 2.4 GHz ISM band, which offers 14 channels, each one also 20 MHz wide. In this case, a combination of Complementary Code Keying (CCK) modulation with a Direct–sequence Spread Spectrum (DSSS) PHY allows to offer 4 different data rates, namely 1, 2, 5.5 and 11 Mbps.

The IEEE 802.11g and IEEE 802.11e amendments

The IEEE 802.11g amendment was released in 2003 and utilizes basically the same OFDM PHY of IEEE 802.11a, but deployed in the 2.4 GHz band. Despite offering the same data rates of the older standard, from 6 to 54 Mbps, it managed to obtain broader commercial success due to the fact that IEEE 802.11g products were cheaper and often compatible with existing IEEE 802.11b networks already deployed in the same frequency band.

IEEE 802.11e, despite preceding IEEE 802.11g in alphabetical order, was released later, in 2005, and, rather than introducing a new PHY layer, it upgrades the original IEEE 802.11 MAC layer. Specifically, it introduces a new channel access mode, named Hybrid Coordination Function (HCF), which aims at enhancing the Quality–of–Service (QoS) of IEEE 802.11 users. In HCF, different priorities could be assigned to different STAs or even to different traffic flows within a STA and the channel sensing time can be changed from a DIFS to an Arbitration Inter Frame Space (AIFS), whose value reflects the priority (higher–priority traffic flows have a lower AIFS value). Moreover, a STA or a traffic flow can also be granted a Transmit Opportunity (TXOP) period of a specified length during which it has unlimited access to the channel. This possibility is interesting in industrial applications, where nodes with different real–time requirements can be assigned to different classes (Cena et al., 2010).

The IEEE 802.11n and IEEE 802.11ac amendments

IEEE 802.11n, released in 2009, was the first amendment to introduce MIMO in WLANs, borrowing this technology from mobile communication networks (Perahia and Stacey, 2013). Besides MIMO, IEEE 802.11n introduces a set of other enhancements targeted at improving the data rate, that can reach 600 Mbps (almost an order of magnitude higher than the previous amendments. IEEE 802.11n does not limit to data rate). other new

features are targeted at improving reliability (e.g., Low–Density Parity–Check (LDPC) channel coding) and QoS (e.g., frame aggregation and Block ACK). The possibilities offered by IEEE 802.11n are of great interest for industrial applications (Tramarin et al., 2016b) and they are discussed in detail in Ch. 4 of this thesis.

The IEEE 802.11ac amendment was released in 2013 as an upgrade of IEEE 802.11n, although it is restricted only to the 5 GHz band, whereas IEEE 802.11n can work also in the 2.4 GHz one. The use of more efficient modulation and coding, coupled with wider channels and more complex MIMO architectures, pushes the maximum data rate up to 6.93 Gbps. Moreover, this amendment also includes Multiuser Multiple–Input Multiple–Output (MU-MIMO) and a more refined version of the Transmit Beamforming (TxBF) feature already introduced in IEEE 802.11n.

The IEEE 802.11ad amendment

With the IEEE 802.11ad amendment, released in 2012, the IEEE 802.11 standard opened to a new ISM frequency band, namely the one between 57 and 66 GHz. In this portion of the spectrum, often termed millimeter–wave (mmWave), extremely wide 2.16 GHz channels are available, enabling very high data rates. Three different types of PHY layers are introduced in this standard: a single carrier one (peak data rate of 4.62 Gbps), an OFDM one (peak data rate of 6.76 Gbps) and a Low–Power Single Carrier (LPSC) one (peak data rate of 2.5 Gbps).

The downside of using higher frequencies is that the path loss is much stronger and make communications over long distances (more than 100 m) practically unfeasible (Daniels and Heath Jr, 2007). Moreover, transmitter and receiver should be perfectly aligned to achieve a good transmission quality and, hence, beamforming strategies must be devised. For this reason, IEEE 802.11ad also defines a custom MAC layer, based on a TDMA that alternates training phases (during which directional beams are formed), contention–based phases and contention–free phases.

Finally, the realization of IEEE 802.11ad devices also presents some technical difficulties that have slowed down the establishment of this amendment so far. They include non–efficient operations of power amplifiers at high frequencies, increased phase noise from local oscillators and difficulties in handling high multi–Gbps data rates by ADCs and DACs (Rappaport et al., 2011).

The IEEE 802.11ax and IEEE 802.11ay amendments

IEEE 802.11ax and IEEE 802.11ay are the two next major amendments to the standard, with a release date scheduled towards the end of this decade. They are conceived as

upgrades of IEEE 802.11ac and IEEE 802.11ad, respectively.

IEEE 802.11ax will be focused in improving the efficiency of WLAN deployments in the 2.4 and 5 GHz frequency bands. Specifically, this means that the focus will not be on maximizing the throughput of a single link, but rather on increasing area throughput in very dense scenarios where a lot of APs and STAs are employed (e.g., 1 user per m^2) and can interfere with each other (Bellalta, 2016). To this aim, both PHY and MAC layer will be upgraded. The former will be improved with the use of more efficient modulations and higher order Fast Fourier Transform (FFT), likely increasing the achievable data rate over 10 Gbps. The latter, instead, will be transformed with the use of an Orthogonal Frequency–Division Multiple Access (OFDMA) scheme similar to the one used in LTE instead of CSMA/CA, the extensive use of MU-MIMO and the increase of spatial reuse through adaptive Clear Channel Assessment (CCA) and Transmit Power Control (TPC). The enhancements brought by this amendment will be combined with those of other minor amendments, e.g. IEEE 802.11aq (pre–association discovery of services), IEEE 802.11ak (bridged networks) and IEEE 802.11ai (fast initial link setup time) to guarantee ever increasing WLAN efficiency (Bellalta, 2016).

The IEEE 802.11ay amendment, instead, will target the mmWave spectrum in the 57-66 GHz band (Au, 2016). Through the use of more efficient modulations, channel bonding and MIMO, the data rate will be increased significantly with respect to IEEE 802.11ad, possibly exceeding 100 Gbps (although data rates this high present severe technical challenges and realistically achievable values could be in the order of 20 Gbps). The MAC layer will also be upgraded borrowing some features from IEEE 802.11ax, such as OFDMA and MU-MIMO.

3.2 The IEEE 802.15.4 standard for LR-WPANs

The IEEE 802.15 standard was developed at the beginning of the century to define WPANs as short–distance wireless networks that could be used to connect portable and mobile computing devices belonging to an individual, as an alternative to the long–distance communications offered by WLANs. During the years, this standard went far beyond its initial scope and nowadays several different versions are present, whose data rate and range are comparable with those of IEEE 802.11.

In detail, the IEEE 802.15.4 version of the standard targets LR-WPAN, whose key requirements are low complexity, low power and low cost, at the expense of a low data rate. IEEE 802.15.4 offers a common MAC layer, defined in the 2003 version of the standard, and several different PHY layers defined over the years. More recently, different kinds

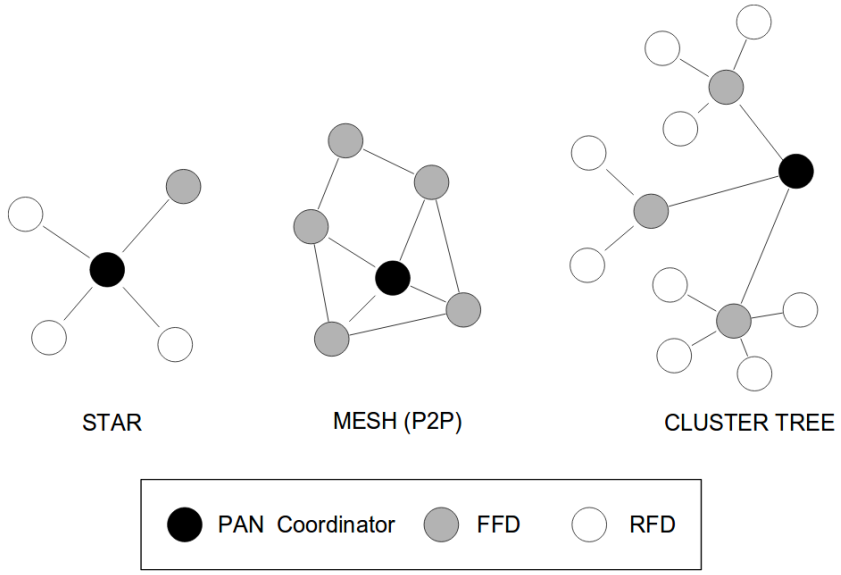


Figure 3.2: Possible topologies in an IEEE 802.15.4 network, taken from Tramarin (2012).

of MAC layers dedicated to specific applications have been developed as amendments to the standard, such as the "industrial" MAC defined in IEEE 802.15.4e, that will be covered in detail at the end of this section. The latest established version of this standard, defined in 2015, includes amendments up to IEEE 802.15.4p (IEEE 802.15.4-2015).

Two types of IEEE 802.15.4 devices are defined in the earliest version of the standard and maintained ever since: Full–Function Devices (FFDs) and Reduced–Function Devices (RFDs). The first type of device has higher capabilities and can take three roles in the network: Personal Area Network (PAN) coordinator, coordinator or a simple device. RFDs, instead, have much lower resources and can only operate as devices. Two types of network topologies are also defined: a star topology, in which all devices (either FFD or RFD) can only communicate with the PAN coordinator, and a peer–to–peer (or mesh) topology, in which RFDs can only communicate with the PAN coordinator but all FFDs can communicate with each other (Salman et al., 2010). A combination of the two topologies can also be envisioned, with a "cluster tree" in which each cluster is a small star network coordinated by a FFD and one FFD act as PAN coordinator, communicating with all the other coordinators. Fig. 3.2 reports the different possible topologies in an IEEE 802.15.4 network.

Between 2003 and 2009, several different amendments to IEEE 802.15.4 were released, defining specific PHY layers operating in different frequency bands with different modulations (Salman et al., 2010). A brief list is of the available PHY layers is given in the

following:

- *IEEE 802.15.4*: the original version of the standard, released in 2003, defined three operating bands, namely 2.4 GHz (worldwide), 868 MHz (in Europe) and 915 MHz (in USA). In the first band, an Orthogonal Quadrature Phase-Shift Keying (O-QPSK) modulation allows to reach a 250 Kbps data rate, whereas in the other two bands Binary Phase–Shift Keying (BPSK) is used, offering 20 and 40 Kbps respectively. All modulations used the DSSS technique.

- *IEEE 802.15.4b*: released in 2006, this amendment extended O-QPSK modulation to the 868/915 MHz bands, raising their data rates to 100 and 250 Kbps respectively, and introduced a Parallel–sequence Spread Spectrum (PSSS)–based BPSK and Amplitude Shift Keying (ASK) PHY layers in the same bands, both offering a 250 Kbps data rate.

- *IEEE 802.15.4a*: this amendment, released in 2007, extended IEEE 802.15.4 to the UWB band of 3.1-10.6 GHz, where different data rates are available up to 27.24 Mbps. Moreover, the fine time resolution offered by UWB allowed the development of high–precision ranging applications, often used also in industrial environments (Silva et al., 2014).

- *IEEE 802.15.4c*: in 2009, this amendment extended IEEE 802.15.4 to the 780 MHz in China, offering O-QPSK and M–Phase–Shift Keying (PSK) modulations, both capable of 250 Kbps data rates.

- *IEEE 802.15.4d*: also in 2009, this amendment extended IEEE 802.15.4 to the 950 MHz in Japan, offering BPSK and Gaussian Frequency–Shift Keying (GFSK) modulations, capable of 20 and 100 Kbps respectively.

Despite a high variety of PHY layers, the IEEE 802.15.4 MAC layer is common among them[2] and comes in two varieties: beacon–enabled and non beacon–enabled. In the former mode, periodic beacons are sent by the PAN coordinator, each signaling the beginning of a superframe of 16 slots. Some of these slots can be Guaranteed Time Slots (GTSs) assigned to specific devices, thus obtaining a contention–free access, while in other slots all devices compete for the channel using CSMA/CA. There can also be inactive slots where all devices can go in sleep mode, thus saving some energy. In non beacon–enabled mode there are no superframes and all the devices compete for the channel with an unslotted CSMA/CA. The CSMA/CA mechanism adopted in IEEE

[2]With the exception of IEEE 802.15.4a, which adopts an ALOHA channel access scheme better suited to UWB.

802.15.4 is similar to that of IEEE 802.11 with the difference that, in beacon–enabled networks, the backoff period must be aligned with the superframe slot boundary and, if the channel is found idle, a node does not transmit immediately, but performs a certain number of backoff periods, defined by the *CW* variable.

IEEE 802.15.4 only defines the two lowest layers of the protocol stack, namely PHY and MAC layers. However, this standard is often used together with the ZigBee standard, promoted by the ZigBee alliance (ZigBee), which defines all the upper layers up to the application, to the point that the two terms (IEEE 802.15.4 and ZigBee) are often used as synonyms. With the recent emergence of the IoT paradigm, new upper layer protocols and suites that adopt IEEE 802.15.4 as a low–layer solution has been proposed, such as the Thread suite for BA (Thread). Most of these solutions are based on the IPv6 over Low power WPAN (6LoWPAN) specifications, developed by the Internet Engineering Task Force (IETF), which allow to implement the Internet Protocol version 6 (IPv6) protocol also on devices characterized by low power and low capabilities such as the IEEE 802.15.4 ones (IETF-RFC 4919).

The IEEE 802.15.4e amendment

Ever since the early years of the IEEE 802.15.4 standard, this new technology started to gain the attention of the industrial communication community, with the development of IWSNs in which IEEE 802.15.4 networks are used to monitor industrial processes (Gungor and Hancke, 2009). Soon after, dedicated industrial wireless standards based on IEEE 802.15.4 were born, such as WirelessHART and ISA 100.11a. These efforts were recognized by the IEEE 802.15.4 standardization group in 2012 with the publication of the IEEE 802.15.4e amendment, specifically targeted at addressing ICNs requirements such as timeliness, reliability, scalability and energy efficiency.

This amendment borrows many ideas from WirelessHART and ISA 100.11a (De Guglielmo et al., 2016) and introduces five new MAC behavior modes: Time–Slotted Channel Hopping (TSCH), Deterministic and Synchronous Multi–channel Extension (DSME), Low–Latency Deterministic Network (LLDN), Asynchronous Multi–channel Adaptation (AMCA) and Radio Frequency Identification Blink (BLINK). The most interesting modes from an industrial perspective are the first three.

TSCH is designed for PA applications and combines the time–slotted access already defined in IEEE 802.15.4 with a multi–channel ability, that allows multiple nodes to exchange frames simultaneously in different channels, and with a channel hopping behavior, that mitigates the effects of interference and multipath fading (De Guglielmo et al., 2016). Synchronization is maintained in the network through periodic beacons

and, to avoid clock drifts, each node is associated to a time–source neighbor to which it is resynchronized each time the two nodes exchange a data or ACK frame. The TSCH superframe contains both dedicated slots and shared ones, with the access being regulated through a modified CSMA/CA procedure during the latter, but there is no default mechanism for scheduling the slots and several ones are being proposed in the scientific literature (De Guglielmo et al., 2016). Finally, the 6TiSCH working group of IETF is working on combining TSCH–based IEEE 802.15.4 networks with 6LoWPAN, effectively building a complete protocol stack for IIoT applications (Dujovne et al., 2014).

DSME is targeted at application with stringent time and reliability requirements. It is also based on time–slotted access, but a multi–superframe structure is defined, that can group cycles of repeated superframes, each composed by beacon, contention–free period with GTSs and contention–access period. Multi–channel abilities are supported in DSME also, with the availability of multiple channels being used for either channel hopping or channel adaptation (i.e., neighboring nodes agreeing on a channel schedule based on their link quality). A group ACK feature is also available, with which coordinators send a single ACK frame to acknowledge multiple transmissions. Finally, GTS scheduling is performed in a distributed way through an handshake procedure among pair of nodes that wish to communicate.

Finally, LLDN is targeted at single–hop and single–channel networks used for FA applications requiring very low latency. The topology is restricted only to a star network, with the central PAN coordinator organizing and distributing the slots schedule. Different kinds of time slots are present in a superframe (beacon, management, uplink and bidirectional) and a slot duration can be even shorter than 1 ms. During a timeslot, the access can be exclusive, contention–based (with a simplified CSMA/CA) or dedicated to the PAN coordinator. Group ACKs can be used and the synchronization is maintained through beacons.

3.3 Other relevant wireless standards

A brief overview of other wireless network standards that can be used in some industrial applications and that were mentioned in Sec. 2.4 is given in this section.

The IEEE 802.15.1 standard for WPANs (Bluetooth)

IEEE 802.15.1 has been the first standard of the IEEE 802.15 family to be published, in 2002 (IEEE 802.15.1-2005), actually embodying the original WPAN concept: short–range wireless communications between portable devices being used by an individual.

This standard was based on the Bluetooth technology, introduced by Ericsson in 1994 and handled, since 1998, by the Bluetooth Special Interest Group (SIG), which unites more than 30000 companies (Bluetooth–SIG). The work of the SIG over the Bluetooth specifications has continued over the years, with the publication of updated versions, the last of which is Bluetooth 5.0 in 2016 (Bluetooth–5.0). On the other hand, the IEEE 802.15.1 standardization group has stopped its activities in 2005 and no longer maintains the latest versions of the standard.

Bluetooth is a standard deployed in the 2.4 GHz ISM band, where 79 channels, each 1 MHz wide, are defined. The typical topology of a Bluetooth network is called piconet, where one device acts as a master and all the other devices (up to seven) act as slaves, communicating only with the master and synchronized to its clock. A scatternet can also be formed as a collection of piconets overlapping in time and space and one device can participate to multiple piconets simultaneously (Lee et al., 2007). A key feature of the Bluetooth standard has always been the use of Frequency–Hopping Spread Spectrum (FHSS): the devices of a piconet continuously change the transmission channel among the 79 available ones with a rate of 1600 hops per second, with a predefined pattern based on the master's address. This feature allows a good robustness to fading and external interference by other networks deployed in the same band (e.g., WLANs). Data exchange in Bluetooth networks is organized in slots of 625 μs duration and a packet can occupy 1, 3 or 5 slots, with master always beginning the transmission in even slots and slaves in odd ones (Bisdikian, 2001). Finally, the use of GFSK modulation in the earliest version of Bluetooth allowed to reach a 1 Mbps bandwidth.

New features have been introduced in the newer versions of Bluetooth specifications. For example, Bluetooth 2.0, released in 2004, introduced an optional Enhanced Data Rate (EDR) for enhanced data transfer, that allowed to reach 3 Mbps by combining GFSK and PSK. Bluetooth 3.0, released in 2009, provides even higher speed by collaboration with co–located IEEE 802.11 networks: Bluetooth is used for negotiation and establishment of a connection, then data are exchanged at 24 Mbps over IEEE 802.11. In 2010 Bluetooth 4.0 has been released, introducing a new protocol besides classic Bluetooth and high speed Bluetooth, that was called Bluetooth Low Energy (BLE). This protocol is based on a very light stack and on advanced power save features, aiming at achieving years–long lifetime on coin–size batteries (Gomez et al., 2012). Finally, Bluetooth 5.0 has been released in 2016, providing higher speed, longer range and even lower energy consumption, targeting the emerging IoT market (Ray and Agarwal, 2016).

The IEEE 802.15.3 standard for HR-WPANs

While IEEE 802.15.4 is in charge of defining LR-WPANs, the IEEE 802.15.3 part of the standard targets high–rate networks that can offer transmission speed comparable to IEEE 802.11, but at a lower cost and complexity and with a shorter range.

The first direction towards the realization of HR-WPANs was the use of the UWB spectrum in the 3.1-10.6 GHz band, which was standardized around the beginning of 2000s in several regions. This was the target of the IEEE 802.15.3a standard, which began its process in 2003. However, the standardization efforts failed due to the high competition and consecutive deadlock between two PHY layer proposals: Direct–sequence Ultrawide Band (DS-UWB), adopting variable–length spreading codes based on tight synchronization requirements, and Multiband Orthogonal Frequency–Division Multiplexing (MB-OFDM), that combines OFDM with time and frequency spreading. The heated competition led to the ultimate disband of this standard in 2006, with no compatible products reaching the market (Park and Rappaport, 2007).

After the failure of IEEE 802.15.3a, the HR-WPAN concept remain silent until 2009, when the first wireless standard for the mmWave spectrum in the 60 GHz band, IEEE 802.15.3c, was released. In this standard, three different PHY layers are proposed: a single carrier one, for low power and low complexity applications reaching up to 5.28 Gbps; an High–Speed Interface (HSI) one for symmetric low–latency data transfer, based on OFDM and capable of reaching 5.78 Gbps; and an Audio–Video (AV) one for uncompressed high–definition video streaming, that allowed to reach 3.8 Gbps (Emami, 2013). The IEEE 802.15.3c MAC layer is based on a centralized protocol that mixes contention–free and contention–based access and includes beamforming procedures for directional communications, very similar to the one employed in IEEE 802.11ad. Different frame aggregation modes and error–protection strategies are developed for specific kinds of traffic, such as Unequal Error Protection (UEP) to differentiate most and least significant bits in AV traffic.

A new milestone in the context of HR-WPANs has been reached with IEEE 802.15.3e, released in 2017 and also deployed in the 60 GHz spectrum. The target scenario of this amendment is high–rate and close–proximity wireless networks for multimedia applications. A simplified and optimized MAC layer is introduced to enhance throughput and several PHY layer enhancements, such as channel bonding up to 8 GHz bandwidth and MIMO up to 16×16, allow to push the transmission rate over 100 Gbps. In order to offer a good reliability, the target communication distance is restricted to 10 cm, effectively enabling only extremely close–proximity applications.

In parallel to IEEE 802.15.3e, another amendment for HR-WPANs is being developed,

although it has not been published yet. It is the IEEE 802.15.3d amendment, which will also target 100 Gbps data rate, but will not be restricted to a 10 cm range, supporting longer communication distances up to several meters. In order to achieve this ambitious goals, this amendment will be deployed in the 275 GHz spectrum and will be limited to switched point–to–point links.

Mobile network standards

Mobile communications are often called "cellular" communications due to one of their most distinctive features: network coverage over large geographical areas is ensured by the presence of cells of variable size, each characterized by the presence of at least one fixed transceiver called BS. Mobile users, also called User Equipments (UEs), that are located within a cell can hence communicate with the nearest BS, which provides them with many services, including access to the Internet. To enhance frequency reuse, neighboring cells operate in different sets of frequencies and therefore interference is avoided (Miao et al., 2016). All the adopted frequencies are proprietary ones, licensed to the cellular operators, so that interference with other networks (e.g., IEEE 802.11) is completely avoided. This architecture, that allows to increase network capacity and coverage area while decreasing the power consumption of UEs, has been the fundamental reason for the incredible success of cellular communications, that nowadays basically cover the entire planet (Dahlman et al., 2013).

At the beginning, cellular networks were analog and only supported voice communications. The second generation started with GSM, which was based on digital communications and added to voice services the possibility of sending short messages as Short Message Services (SMSs). Always belonging to the second generation, although often termed "2.5G", were GPRS and EDGE, which introduced the possibility of exchanging data packets over a cellular network, allowing mobile users to browse the Internet and changing forever the landscape of mobile communications. These technologies were based on TDMA and Frequency–Division Duplex (FDD) and allowed a data rate up to some hundreds of Kbps (Dahlman et al., 2013).

A further development in cellular networks was reached with the third generation, the first one handled as an international standard by 3GPP. The first network of this generation was UMTS, which required the deployment of new BSs and new frequency bands and defined a complete network architecture, divided in a radio access network called Universal Mobile Telecommunication System Terrestrial Radio Access Network (UTRAN), and in a core network called Mobile Application Part (MAP). In the radio access, FDD and Time Division Duplexing (TDD) could be used for uplink/downlink duplexing, as

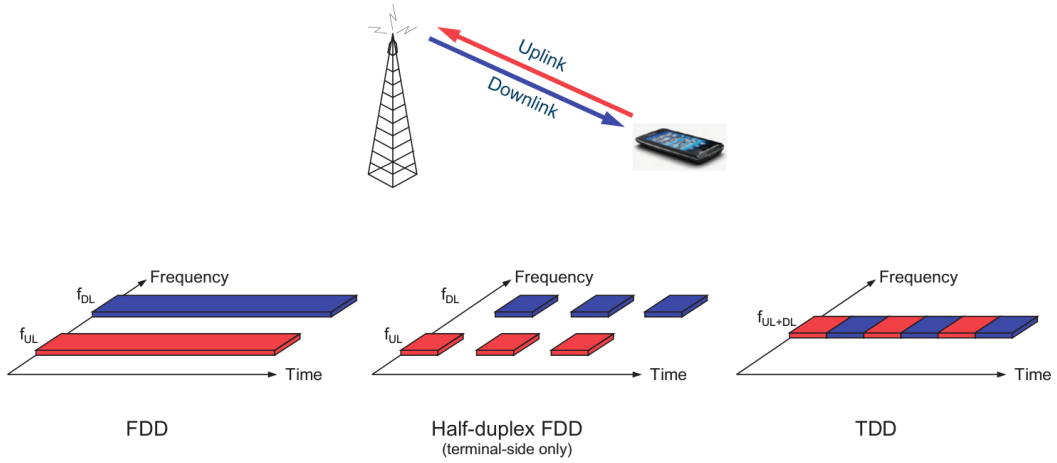


Figure 3.3: Different duplexing strategies in cellular networks, taken from Dahlman et al. (2013).

highlighted in Fig. 3.3, while Code–Division Multiple Access (CDMA) was used as a multiple access mechanism. Originally, UMTS offered a 384 Kbps data rate in both downlink and uplink, which was increased to 14.4 Mbps and 5.76 Mbps, respectively, with High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA). The last standard of the third generation, Evolved High Speed Packet Access (HSPA+), instead offered 42 Mbps in downlink and 11.5 Mbps in uplink.

The fourth generation of cellular networks began with LTE, whose most distinctive feature was the use of OFDMA as a channel access mechanism. This feature, combined with enhanced modulation and coding, higher availability of frequency spectrum, a wider bandwidth of 20 MHz and the use of MIMO, allowed to reach data rates of 300 Mbps in downlink and 75 Mbps in uplink. Data rates and coverage were further improved with Long–Term Evolution Advanced (LTE-A), thanks to enhanced MIMO, carrier aggregation (up to 100 MHz) and heterogeneous network deployment (with overlapping layers of cells), reaching 3 and 1.5 Gbps of peak data rate in downlink and uplink respectively.

The next generation of cellular networks will be the fifth one and, unlike previous generations, it will not have a specific name but will be generically termed "5G". The primary focus of 5G will no longer be only data rate; in fact, three different major scenarios will be targeted by the fifth–generation networks (Lee and Kwak, 2016): Enhanced Mobile Broadband (eMBB), Massive Machine–Type Communications (mMTC) and Ultra–Reliable and Low–Latency Communications (URLLC). The first scenario follows the classical evolution of cellular networks and will target 20 Gbps in downlink and 10 Gbps in uplink, through the use of technology enhancements such as massive MIMO

(Boccardi et al., 2014), mmWave (Rappaport et al., 2013) and new waveforms (Banelli et al., 2014) (among the others). The other two scenarios, instead, are more relevant for industrial applications and will target, on one hand, massive IoT systems with hundreds thousands of users, and on the other hand mission–critical applications that require bounded latency (lower than 1 ms) and high reliability. Lower latency will be targeted through reduced Transmission Time Interval (TTI) and some redesign of the network architecture, including Device–to–Device (D2D) communications and Network Function Virtualization (NFV) (Schulz et al., 2017; Ford et al., 2017). Thanks to these enhancements, with 5G it will be possible for the first time to use cellular networks as ICNs for some high–performance applications, such as FA (Yilmaz et al., 2015). However, it must be noted that, even taking into account the 5G–specific enhancements, cellular networks are characterized by a significant network overhead that makes them unfeasible for the most demanding industrial control applications (Luvisotto et al., 2017a).

4

Real–time WLANs

The IEEE 802.11 standard for WLANs defines wireless networks with performance comparable to those of Ethernet, especially in terms of data rate. However, in their default configuration, these networks do not meet the standards required by the majority of industrial control applications, especially concerning deterministic data exchange.

Nevertheless, the real–time performance of WLANs can be improved significantly, making them suitable for some industrial applications, simply through an opportune configuration of their parameters, especially at the MAC layer. In this chapter some possibilities in this regard are explored, with particular attention to the configuration of IEEE 802.11n WLANs and to the development of custom rate adaptation algorithms.

This chapter is mainly based on the works in Tramarin et al. (2016b), Tramarin et al. (2015), Tramarin et al. (2017), Tramarin et al. (2016a) and Luvisotto et al. (2017d).

4.1 IEEE 802.11n for industrial communications

Most of the currently deployed industrial IEEE 802.11 systems are based on dated physical layers and do not take advantage of the enhancements provided by the most recent amendments. As a consequence, issues like poor reliability and quite low transmission rates (compared to the wired RTE counterparts) still undermine the adoption of the IEEE 802.11 standard in factory communication systems.

In this direction, the IEEE 802.11n amendment provides several improvements to the previous versions. In particular, such an amendment supports multi–antenna operations so that MIMO systems capable of increased reliability, longer communication distances and higher transmission rates can be implemented. Such innovations required the design of a new PHY layer as well as the introduction of several enhancements to the MAC layer. Nowadays, IEEE 802.11n networks are widely deployed in general purpose communication systems. Indeed, several off–the–shelf available devices (e.g. personal computers and tablets) are equipped with IEEE 802.11n interfaces. Conversely, this is not the case for the industrial scenario, where these networks are still rarely deployed. However, the new features made available by IEEE 802.11n are expected to significantly improve the performance of wireless ICNs, so that undertaking a thorough investigation in this direction looks an appropriate choice.

The introduction of IEEE 802.11n in the industrial communication scenario is still at a very initial stage. However, some meaningful contributions are worth to be mentioned. Both papers Santonja-Climent et al. (2010) and Silvestre-Blanes et al. (2015) deal with the adoption of IEEE 802.11n for real–time industrial multimedia traffic, such as that generated by the transmission of video streams. In particular, Santonja-Climent et al. (2010) describes a case study in which this network is deployed to connect audio and video devices in a urban context, while Silvestre-Blanes et al. (2015) provides a performance analysis of an industrial control system that makes use of two IEEE 802.11n networks to exchange real–time multimedia data between an industrial Personal Computer (PC) and some supervisory devices. In Charfi et al. (2014) and Maqhat et al. (2012), the design of scheduling algorithms able to handle QoS–aware traffic on IEEE 802.11n is addressed. Both papers investigate the adoption of the available frame aggregation mechanism to achieve real–time performance. Interestingly, in both papers the authors come to the conclusion that such mechanism is not suitable for real–time communication and, consequently, recommend to use alternative strategies. Paper Rentschler and Laukemann (2012) analyzes the use of IEEE 802.11n to implement an application of the Parallel Redundancy Protocol (PRP). In particular, it describes the deployment of two different WLANs to duplicate the transmission of safety critical messages between two Programmable Logic Controllers (PLCs). The last contribution which is worth mentioning is Jin and Dai (2012). This paper provides a comprehensive assessment of the impact of PHY transceiver impairments on the performance of IEEE 802.11n. The authors focus on the behavior of the Bit Error Rate (BER) versus Signal–to–Noise Ratio (SNR), considering the effects of different impairments, such as signal coupling, phase noise and gain imbalance. The obtained results provide useful insights for the

design of hardware devices to be specifically used in industrial communication systems.

The above considerations allow to conclude that, although some useful indications are currently available, the actual and effective introduction of IEEE 802.11n in industrial communication systems still needs accurate investigations.

Configuring a IEEE 802.11n industrial WLAN

A brief overview of the enhancements introduced by the IEEE 802.11n amendment is first provided, then some recommendations for its configuration in industrial applications are drawn.

Overview of IEEE 802.11n enhancements

IEEE 802.11n PHY The set of available modulations has been modified with respect to IEEE 802.11a/g: a 64–Quadrature Amplitude Modulation (QAM) modulation with a higher code rate of 5/6 has been added (sacrificing one of the two BPSK modulations), thus allowing an 11% gain in raw transmission rate, even if the increased spectral efficiency of this higher modulation scheme negatively impacts on transmission robustness. As far as coding is concerned, the amendment allows to replace the legacy convolutional channel codes with the more robust LDPC ones. It has to be considered, however, that the adoption of LDPC codes requires a minimum payload length considerably greater than that needed by convolutional codes. This may represent a limitation to the use of this coding strategy for industrial traffic, which is often characterized by small payload sizes.

The number of subcarriers reserved to data within the OFDM modulation has been increased from 48 to 52, yielding a further 8% rate improvement. Combining such a feature with the new 64–QAM modulation, the maximum transmission rate increased from 54 Mbit/s up to 65 Mbit/s. More importantly, the new PHY defined the availability of wider 40 MHz transmission channels, by exploiting channel bonding between two adjacent 20 MHz channels, roughly doubling the transmission rate. Indeed, in such channels the total number of subcarriers is doubled to the value of 128, of which 108 are dedicated to data symbols, raising the theoretical transmission speed to 130 Mbit/s. This feature is available in both the 2.4 GHz and the 5 GHz frequency bands, despite the bandwidth of the lower 2.4 GHz band is quite narrow and may limit the adoption of 40 MHz channels. Moreover, a possible issue in the use of these channels is an expected 3 dB degradation in receiver sensitivity, that may eventually result in a lower coverage range. A detailed discussion about this topic is provided in Perahia and Stacey (2013).

A further enhancement allows to halve the GI between two consecutive OFDM symbols from 800 ns (which is the default value in a legacy IEEE 802.11a/g system) to 400 ns, to further raise the transmission rate, up to 150 Mbit/s. Anyway, some drawbacks of reducing the GI are found in an increased need for accurate time synchronization, and even more in a stronger sensitivity to inter–symbol interference, which may be particularly significant in industrial environments possibly characterized by high delay spread.

MIMO capabilities The most significant new feature introduced in IEEE 802.11n has been the possibility to use MIMO devices. The baseline scheme for the exploitation of a MIMO system is represented by Spatial Division Multiplexing (SDM), in which the frame payload is subdivided in independent streams of data, each one assigned to one of the available transmitting antennas, with the aim of maximizing the throughput. In this case the raw transmission rate of the system at PHY layer increases linearly with the number of independent data streams. The amendment supports at most 4×4 systems (4 transmitting and 4 receiving antennas), hence allowing up to 4 independent streams to reach the raw transmission speed of 600 Mbit/s.

MIMO can be alternatively exploited to improve network reliability. To this purpose, the standard allows to adopt the Space–Time Block Coding (STBC) technique, according to which the signals transmitted over the different antennas are suitably coded exploiting both temporal and spatial diversity. This strategy significantly increases the transmission success probability, hence enhancing the communication reliability with respect to both SDM and the previous versions of the IEEE 802.11 standard. SDM and STBC are mutually exclusive, since the latter does not allow to send multiple independent streams. Anyway, for the transmission of frames with limited–size payloads, this does not affect significantly the transmission time, while providing a substantial gain in terms of reliability.

A final opportunity is represented by TxBF, where different weights are assigned to transmitted signals to better adapt to the channel status. Although this approach looks promising, for the time being its complexity and large number of options have discouraged its adoption. Consequently, it is implemented in few IEEE 802.11n devices (Perahia and Stacey, 2013) and its use will not be further addressed here.

PHY terminology Given the high number of possible PHY configurations made available by IEEE 802.11n, the amendment exploits a numerical notation to quickly identify the various options, referred to as Modulation and Coding Schemes (MCSs), that was formerly introduced by 3GPP in its specifications (Dahlman et al., 2013).

Using such a notation, this study will restrict to MCS 0 – MCS 15, that are the most

commonly implemented in practice. The first 8 schemes (from MCS 0 to MCS 7) refer to configurations where a single spatial stream is transmitted and reflect a basic 1×1 architecture as well as a 2×2 MIMO STBC system. To a certain extent, they can be mapped on the basic modulation set of IEEE 802.11g. Conversely, schemes from MCS 8 to MCS 15 are used for the case of two spatial streams, i.e. when a 2×2 MIMO SDM system is adopted.

IEEE 802.11n MAC layer The new MAC layer strongly builds on the IEEE 802.11e foundations, which introduced QoS concepts and defined the possibility for a station to obtain a TXOP period during which it can send multiple consecutive frames avoiding contention and backoff procedures. IEEE 802.11n enhanced this feature allowing to aggregate more frames into a single one to be transmitted during a TXOP, thus reducing the overhead due to interframe spaces and headers. Analogously, the IEEE 802.11e Block Acknowledgement (BACK) mechanism, which allowed the receiver to acknowledge the transmission of multiple data units with a single frame, has been improved to account for aggregated frames. Hence, to further improve channel utilization, a BACK is now implicitly sent in response to an aggregated frame.

Two new channel access techniques have been introduced by IEEE 802.11n, namely Reverse Direction Protocol (RDP) and Power–Save Multi–Pol (PSMP). The former is conceived for highly asymmetric traffic patterns, which could result in underutilized TXOP periods, whereas RDP allows a station owning a TXOP to sublease a portion of it to a partner, increasing in such a way channel utilization.

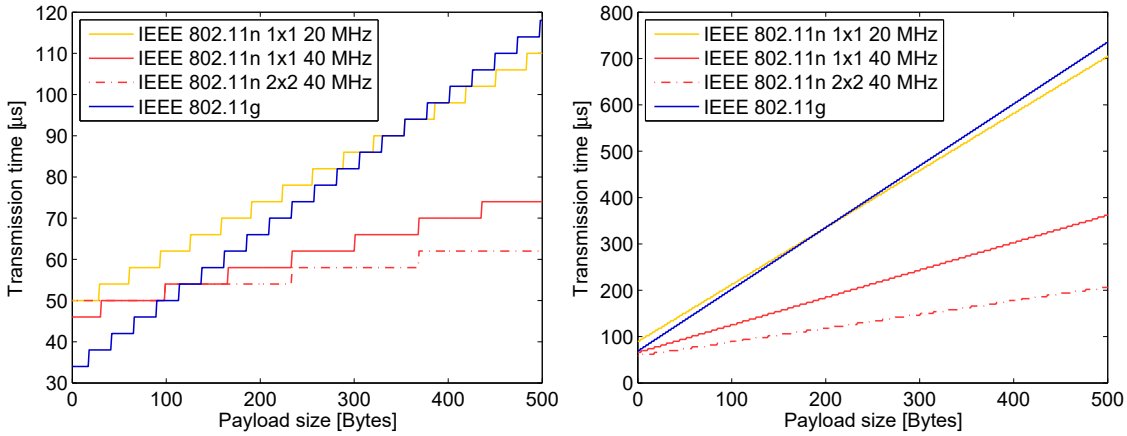
Differently, PSMP is based on a scheduling technique designed for systems in which many nodes periodically transmit small data amounts. Indeed, a central scheduler (generally an access point) periodically sends a PSMP frame which contains a schedule for subsequent time slots that are reserved for downlink or uplink transmissions. The scheduler is also responsible for performing a recovery procedure in case a station misses its reserved slot due to an error in communication.

Considerations on the use of IEEE 802.11n in industrial applications

IEEE 802.11n packet transmission time According to the protocol specification, the transmission time of an IEEE 802.11n packet can be expressed as

$$T_{TX}(l) = T_{preamble} + T_{SIG} + T_{SYM} \cdot \left\lceil \frac{l \cdot 8 + L_{MH} + L_{PH}}{N_{DBPS}} \right\rceil + T_{SE} \tag{4.1}$$

where $T_{preamble}$ and T_{SIG} are the durations of two specific frame fields, namely PHY



(a) Highest order modulation **(b)** Lowest order modulation

Figure 4.1: IEEE 802.11g/n frame transmission time for different PHY configurations.

preamble and SIGNAL, whereas T_{SYM} is the duration of an OFDM symbol. The ceiling function defines the number of transmitted symbols. The term l is the number of bytes contained in the upper levels payload, whereas the terms L_{MH} and L_{PH} account for the MAC and PHY layer headers, which have lengths of 272 and 22 bits, respectively. N_{DBPS} is the number of data bits contained in an OFDM symbol, and is the only term that depends on the adopted MCS. Finally, T_{SE} accounts for a 6 μs Signal Extension (SE) term, introduced only in 2.4 GHz systems for compatibility with legacy IEEE 802.11a devices.

The time $T_{preamble}$ deserves particular attention since, for backward compatibility purposes, a frame begins with an IEEE 802.11a/g legacy preamble, followed by a new High Throughput (HT) preamble containing fields related to the new options. This structure is called Mixed Format (MF) and evidently introduces a considerable overhead. An alternative exists, called Greenfield (GF) preamble, in which the legacy fields are removed, leading to a fixed 12 μs reduction of the packet transmission time, at the cost of losing any compatibility with IEEE 802.11a/g devices.

The packet transmission time has been evaluated for different payload sizes and PHY configurations, to provide a clear assessment of the impact of the enhancements brought by IEEE 802.11n on this metric. This analysis reveals helpful to determine the set of features that may actually yield a performance gain with respect to the previous version of the standard. A comparative analysis of the outcomes from Eq. (4.1) for both IEEE 802.11n and IEEE 802.11g is hence reported in Fig. 4.1. In the figure, three possible configurations of IEEE 802.11n have been included, namely 1×1 with 20 MHz channels, 1×1 with 40 MHz channels and, finally, a 40 MHz 2×2 system configured with SDM.

Specifically, Fig. 4.1a shows the outcomes for the highest order modulation, namely 54 Mbit/s for IEEE 802.11g, MCS 7 for 1×1 and MCS 15 for 2×2 SDM IEEE 802.11n. Similarly, Fig. 4.1b refers to the comparison among 6 Mbit/s, MCS 0 and MCS 8.

The calculations that led to Fig. 4.1 did not take into account the impact of short GI. Also, they did not consider the adoption of LDPC codes, since their use would require payload sizes larger than the ones typically used in industrial applications. Moreover, configurations with more than two antennas have not been considered, since they are less common in practice.

The following important observations can be made with reference to Fig. 4.1.

- The IEEE 802.11n related trends show that the use of 40 MHz channels, in general, significantly decreases the transmission time even for short payloads. The benefits of using wider channels obviously increase for greater payloads.
- The 2×2 SDM configuration generally allows for a reduction of the transmission time with respect to 1×1 configurations. However, while for the lowest order modulation such a reduction takes place for any payload size, this is not the case for highest one. Indeed, Fig. 4.1a shows that a 2×2 SDM configuration performs better than a 1×1 system only for payloads greater than approximately 170 bytes.
- Considering 20 MHz channels, IEEE 802.11g has lower transmission times for small payloads thanks to a lower preamble overhead. For example, with the lowest order modulation, IEEE 802.11n is slightly advantageous only for payload greater than 224 bytes. This difference may be partially mitigated enabling the GF preamble (not considered in the figure).
- Analogously, Fig. 4.1a shows that for payloads shorter than 140 bytes, IEEE 802.11g results faster than IEEE 802.11n with 40 MHz channels. In this case, the use of the GF preamble would completely remove this difference.

Improving reliability As it is well known, the possibly high PER typical of the wireless medium represents one of the most critical aspects for industrial wireless networks, since several retransmissions may be necessary to eventually achieve successful packet delivery, with the consequent introduction of random delays due to the CSMA/CA mechanism. As a consequence, the possibility offered by STBC of exploiting multiple antennas to increase the reliability of data transmission represents an interesting and beneficial option.

STBC requires that consecutive OFDM symbols are encoded in time and sent over different antennas in order to improve successful decoding probability at the receiver.

Table 4.1: Overview of IEEE 802.11n MCSs for 40 MHz channels

	Modulation Parameters					Packet Transmission Time [μs]		
2×2 STBC	MCS	Data Rate [Mbit/s]	Modulation	Code rate	N_{DBPS}	$T_{TX}(50)$	$T_{TX}(200)$	$T_{TX}(500)$
	0	13.5	BPSK	1/2	54	86	170	354
	...	...	...	...	...	...	...	...
	3	54	16–QAM	1/2	216	46	66	114
	...	...	...	...	...	...	...	...
	7	135	64–QAM	5/6	540	38	46	62
2×2 SDM	MCS	Data Rate [Mbit/s]	Modulation	Code rate	N_{DBPS}	$T_{TX}(50)$	$T_{TX}(200)$	$T_{TX}(500)$
	8	27	BPSK	1/2	108	62	106	198
	...	...	...	...	...	...	...	...
	11	108	16–QAM	1/2	432	42	54	78
	...	...	...	...	...	...	...	...
	15	270	64–QAM	5/6	1080	38	42	50

To achieve this result, several coding schemes have been proposed, and in this section the implementation of STBC with the Alamouti method, which is the most popular and adopted one (Alamouti, 1998), was selected. In the case of a 2×2 system, the Alamouti scheme requires two symbol times for the transmission of two OFDM symbols over two antennas, yielding the same transmission speed of an equivalent system with a single data stream, and actually preventing the contemporaneous use of multiple independent schemes.

Tab. 4.1 shows the packet transmission times relevant to 40 MHz channels with standard GI for different MCSs and payload sizes (specifically, the last three columns refer to payloads of 50, 200 and 500 bytes). The values are directly computed from Eq. (4.1) and reported here to better analyze the impact of different PHY configurations. It can be observed that, especially for small packets, the gain in terms of packet transmission time obtained by using a 2×2 SDM system, with respect to a 2×2 STBC one, is quite limited. Hence, given the significantly increased transmission reliability provided by STBC, such a technique reveals convenient for most industrial communication applications.

MAC layer configuration While the QoS–aware features of IEEE 802.11e (which are included in the IEEE 802.11n MAC layer) have been proven to be helpful in an industrial context (Cena et al., 2010), the effectiveness of both frame aggregation and BACK is questionable. Indeed, these techniques perform at their best when a station sends big chunks of consecutive data, so that they can be adopted to reduce the overall channel occupation time. However, in most industrial applications packets are transmitted individually, either on a periodic basis or triggered by a specific event. In such a scenario,

hence, frame aggregation and BACK would lead to large overheads (Saif et al., 2010), thus increasing the delays of the whole data transmission because the aggregated data units need to be completely received before being forwarded to upper layers. Furthermore, in order to employ BACK techniques, a session has to be initiated with the exchange of specific frames between two nodes, causing an additional overhead. On the contrary, in ICNs, efforts are generally made to avoid any frame exchange that is not strictly related to data transfer, such as RTS/CTS, since they reduce the transmission efficiency (Willig et al., 2005). For the above reasons, the systematic adoption of both frame aggregation and BACK can not be considered as an option for industrial communication systems, even if they might reveal advantageous for some specific applications.

As far as new channel access techniques are concerned, RDP is only useful when the traffic pattern is highly asymmetrical. A typical application is the transfer of huge amounts of data adopting TCP at transport layer. However, this situation is only rarely encountered in industrial networks. On the other hand, PSMP looks very promising, in that it is conceived for typical master/slave architectures, which are very common in industrial applications. Actually, PSMP extends and improves the functionalities offered by the traditional Point Coordination Function (PCF) access method. Channel access efficiency is significantly increased through the avoidance of backoff procedures and acknowledgments. Unfortunately, this innovative channel access technique is supported by only very few commercial devices and, consequently, it is difficult to assess it in practice.

Recommendations From the analysis carried out so far, the following recommendations can be suggested for the appropriate deployment of IEEE 802.11n WLANs in industrial communication systems.

- If operational conditions are adequate (e.g. enough bandwidth is available), 40 MHz channels should be preferred to 20 MHz ones, since transmission rate is doubled without compromising reliability for close range communications, which are the most common scenario in industrial networks.

- If backward compatibility is not an issue, then GF preamble should always be adopted, since this choice saves 12 μs in transmission time.

- The use of short GI is discouraged. Indeed, numerical simulations (not shown here due to space limitations) revealed it does not reduce significantly the packet transmission time. On the other hand, short GI increases vulnerability to ISI.

- The use of LDPC codes may be advisable in that they provide a 2 dB SNR gain to achieve the same error probability as convolutional encoding (Perahia and Stacey, 2013). However, as previously discussed, such an option has to be carefully evaluated since LDPC implies more complex encoding/decoding processes and requires longer minimum payload sizes.
- If a multi–antenna system is available, the use of STBC is recommended since it allows to achieve greater reliability without significantly affecting packet transmission times for the traffic profiles of industrial interest.

Experimental assessment

In order to validate the considerations done so far and to precisely assess the performance gains brought by the adoption of IEEE 802.11n, a thorough experimental campaign has been carried out in a typical research laboratory. Unfortunately, it was not possible to completely isolate the test set–up from the surrounding electromagnetic environment. Nonetheless, with the aim of providing an adequate repeatability of experiments as well as to ensure their independence from unknown factors, the environment was carefully monitored through a real–time spectrum analyzer. This allowed to locate the network under test on a suitable portion of the 2.4 GHz ISM band, not continuously used by other WLANs. Hence, only some sporadic interference by mobile devices was present, as confirmed by the measurement outcomes collected.

The test bench was composed by two desktop PCs, namely Dell Optiplex model 745 and 755, running Ubuntu 14.10, kernel Linux 3.16.4 and equipped with two Wireless Network Interface Cards (WNICs) by TP–LINK, namely TL–WN851ND and TL–WN881ND, each one allowing operations with two antennas. The cards are based on the Atheros AR9287 chip, fully compliant with IEEE 802.11n. Both of them are "SoftMAC" devices, i.e. cards that allow a fine control of the transmission path by executing most part of the MAC layer in software. Specifically, they are managed by the open source "ath9k" Linux driver. The WNICs were placed at a distance of roughly 2 meters from each other, with a Line–of–Sight (LOS) path always available and dominant. The network setup for both the MAC and PHY layers followed the recommendations provided in this section, even if, unfortunately, the adopted WNICs did not allow to adopt neither LDPC codes nor the GF preamble. The prototype network under test just described is schematically represented in Fig. 4.2.

It has to be noticed that industrial environments are often characterized by harsh conditions, typically due to longer communication distances, stronger multi–path and external interference than the ones that could be experienced in the laboratory. However,

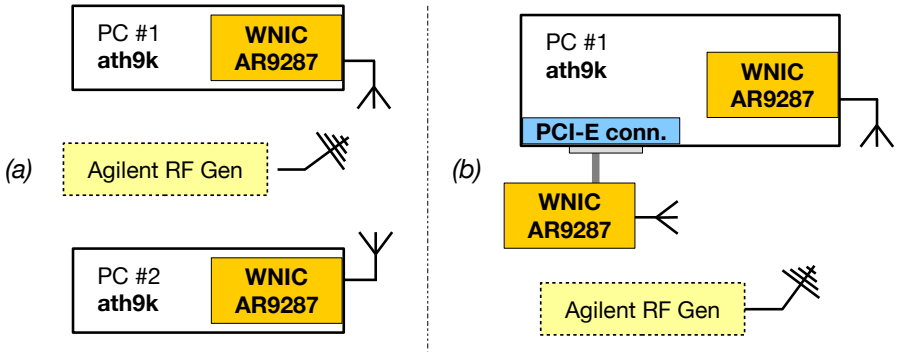


Figure 4.2: Sketch of the adopted measurement setups.

in the attempt to emulate realistic industrial conditions, in terms of both path loss and external interference effects, the transmission power was adjusted and the receiver was suitably disturbed with an additional noise injected through a Radio Frequency (RF) generator. To this regard, an Agilent E4433B generator was used to introduce an accurate and controlled source of channel impairments. Specifically, in the experiments the generator injected wide–band Additive White Gaussian Noise (AWGN)–like noise centered on the carrier frequency of the selected channel by means of a directional antenna. The noise power was then modulated to reproduce the channel conditions suitable for a particular analysis. Unfortunately, due to the configuration of the experimental setup, a similar accurate control on multi–path fading could not be achieved, even if such a phenomenon was surely present during the experiments.

PER assessment

In order to provide an accurate evaluation of the PER, an application was developed for the experimental setup of Fig. 4.2(a), that performs a periodic transmission of short payload UDP packets (50 Bytes) from a PC to the other one. The directional antenna of the RF generator was oriented in such a way to interfere only with the receiving wireless card, in order to avoid triggering the carrier sense mechanism at the transmitter. The interference power was varied to scan a range of 35 different SNR values, at 1 dB steps. For each SNR value, 1000 packets were sent by the application.

To obtain an accurate relationship between PER and SNR, a two–step approach has been followed. Firstly, correspondence was established between the interference noise power injected by the RF generator, and the perceived SNR at the receive side (WNIC of PC#1). To this aim, this value was accurately through a real–time spectrum analyzer able to demodulate the IEEE 802.11OFDM symbols. Then, the relationship between SNR and interference noise power allowed a realistic estimation of the PER for several

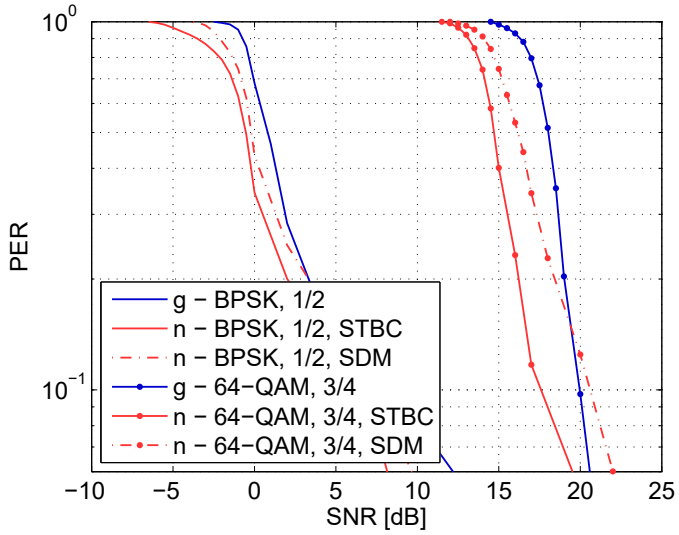


Figure 4.3: PER for different system configurations (50 bytes payload).

values of SNR and different PHY configurations by controlling the noise power of the RF generator. During the tests, the MAC layer source code of the transmitting node was modified to completely avoid MAC–layer retransmissions, thus eliminating any kind of error recovery for packets. Consequently, to calculate the PER it has been sufficient to count the number of packets correctly arrived at the receiving node, since the number of total performed transmissions was known.

The obtained results are reported in Fig. 4.3 for three different configurations, namely 40 MHz IEEE 802.11n 2×2 SDM (red dashed lines), 40 MHz IEEE 802.11n 2×2 STBC (red solid lines), and legacy 20 MHz 1×1 IEEE 802.11g (blue lines). It is worth observing that the obtained trend of the PER is actually in good agreement with both the theoretical analysis carried out by the IEEE TGn working group that defined the amendment (Aldana et al., 2006), and the typical PER curves found in the literature about similar system configurations (Perahia and Stacey, 2013). Nevertheless, under the conditions stated above, the measurements could be considered representative only within a certain threshold of PER, as found in the log–linear representation of Fig. 4.3.

As can be seen, the leftmost curves, relevant to the lowest transmission rates, show a gain in the order of 1-2 dB of IEEE 802.11n with SDM over IEEE 802.11g. A motivation for this fact is that the adoption of wider 40 MHz channels provides a frequency diversity gain with respect to the 20 MHz channels used in IEEE 802.11g (Perahia and Stacey, 2013). Notably, the adoption of STBC as an alternative to SDM allows an even more significant increase in reliability, yielding a further 2-3 dB SNR gain.

The rightmost trends of Fig. 4.3 provide a comparison between the highest available

transmission rate of IEEE 802.11g (54 Mbit/s) and the equivalent 2×2 IEEE 802.11n configurations, namely MCS 6 (2×2 STBC) and MCS 14 (2×2 SDM). These two latter MCSs were selected to carry out a fair comparison, since they use the same modulation of IEEE 802.11g at 54 Mbit/s (64–QAM with code rate 3/4). Also in this case, the beneficial impact of STBC is evident. Indeed, the solid red line highlights an improvement with respect to both IEEE 802.11n with SDM and IEEE 802.11g.

For the sake of completeness, analogous measurements have been performed also for the case of 2×2 MIMO configurations at the highest MCSs, namely MCS 7 (2×2 STBC) and MCS 15 (2×2 SDM), though the relevant outcomes have not been included in the figure to avoid clutter. The measurements for MCS 15 highlighted the occurrence of unsuccessful packet deliveries in nearly half of the transmission attempts, even for SNR values higher than the range in Fig. 4.3. Such a low reliability can be ascribed to the use of 40 MHz channels in the crowded 2.4 GHz band as well as to the adoption of high order (and hence less robust) modulations. Nonetheless, even in this case the use of STBC dramatically decreased the number of failed transmissions, bearing a gain of 1-2 dB with respect to the highest 54 Mbit/s of IEEE 802.11g.

Service time evaluation

In an industrial communication context a very meaningful performance index, that often serves as a basis for several other metrics, is represented by the service time (IEC 61784-2 - 2007). This is defined as the time that elapses from the instant in which a frame is sent (at the MAC layer) to the instant in which the transmitter receives an acknowledgement of its successful delivery. In the case of an ideal IEEE 802.11g/n system without elaboration delays and not affected by transmission errors, the service time can be expressed as

$$T_S = T_{DIFS} + T_{TX} + T_{SIFS} + T_{ACK} \tag{4.2}$$

where $T_{DIFS} = 28$ μs and $T_{SIFS} = 10$ μs are defined by the standard. T_{TX} can be computed from Eq. (4.1). Finally, T_{ACK} is the transmission time of the ACK frame.

A real system, however, is likely to experience significant latencies due to elaboration overheads, as well as to suffer from channel impairments or interference. In these cases, the occurrence of frame losses will trigger retransmissions that, due to the CSMA/CA mechanism, will be interleaved by random backoff times. Thus, to provide a meaningful characterization of the service time behavior in a real environment, an extensive measurement campaign was carried out.

The measurement setup was that shown in Fig. 4.2(b). Two independent WNICs

were installed on the same PC, with the second card connected to the motherboard through a specific PCI–express extension cable to ensure a suitable separation with the other WNIC. Such an arrangement allowed to increase the accuracy of the measurements, since the timestamps for the various events are inherently synchronized between both WNICs.[1]

In these experiments, one of the two stations periodically sent a unicast data frame by means of a purposely developed software, which continuously gathered timestamps both at the transmitting and receiving sides. In each test 10000 frames were delivered, with a payload of 50 bytes and a period of 5 ms.

To provide an exhaustive assessment of the service time, measurements were carried out both with and without interference on the system. The injection of interference at the receiver side was achieved following the same approach adopted in the PER assessment. In particular, the interference noise power was set such that the SNR level resulted equal to 2 dB. As a further step, in this new measurement setup the noise power was modulated with a stochastic process, that already revealed effective for the industrial wireless communication scenario, characterized by ON periods (in which the interference is present) alternated with OFF periods in which interference is not present, and consequently the SNR assumes a high value (Willig et al., 2002). The durations of ON periods were drawn from a uniform random variable ranging from 100 μs to 200 μs, while durations of OFF periods had an exponential distribution with 200 μs mean. This pattern emulates the presence of an external interfering network that uses a completely different transmission scheme with respect to IEEE 802.11, and as such can be regarded as white noise for the prototype system.

It is worth pointing out that the interferer is not influenced by the network under test, since the RF generator does not perform any carrier sensing on the channel. Moreover, the chosen values for the uniform random periods are representative of the transmission of short–size packets, whereas the exponential distribution has been selected to generically resemble a stochastic transmission process in agreement with formerly validated practical experiments (Gamba et al., 2010). Moreover, the whole interference generation procedure allows for high accuracy, reproducibility and controllability of the experiments.

Two system configurations have been compared, namely a legacy IEEE 802.11g and a 40 MHz IEEE 802.11n 2×2 STBC, which represents the best configuration identified in this section. A first set of results is reported in Tab. 4.2.

Comparing the outcomes of the experimental measurements of the service time for

[1] To achieve the best accuracy in these measurements, the timestamps from the internal Time Stamp Counter (TSC) processor register were retrieved. This register is able, under suitable conditions, to provide readings in the order of some tens of nanoseconds.

Table 4.2: Assessment of service time with 50 bytes packets (no rate control)

Configuration	No Interference		Interference	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
IEEE 802.11g 6 Mbit/s	434.0 μs	5.8 μs	919.3 μs	961.8 μs
IEEE 802.11n MCS 0	391.1 μs	13.4 μs	489.5 μs	212.6 μs
IEEE 802.11g 54 Mbit/s	332.1 μs	3.9 μs	556.4 μs	460.1 μs
IEEE 802.11n MCS 7	344.1 μs	8.9 μs	428.4 μs	184.9 μs

the case without interference, with the theoretical values expected from Eq. (4.2), a considerable difference can be found. This discrepancy can be mainly accounted to two factors, namely the processing times introduced by both the operating system and the WNICs, and the time needed for the transmission of the ACK frame in this setup, which resulted to be significantly higher than expected.

Indeed, the ACK transmission rate has to be automatically selected (by the receiver), following the IEEE 802.11 specifications, as a function of the corresponding data frame rate. As an example, if the data frame is transmitted at 54 Mbit/s, then the ACK has to be sent at 24 Mbit/s, which results in a transmission time $T_{ACK} = 34$ μs, leading to a theoretical service time (from Eq. (4.2)) $T_S = 118$ μs, evidently much lower than the values shown in Tab. 4.2. Conversely, measurements collected on the wireless channel showed that the adopted WNICs always sent the ACK in IEEE 802.11b mode at the lowest rate (1 Mbit/s), regardless of the rate used for data transmission, yielding the much higher value of $T_{ACK} = 208$ μs.

If this constant overhead, caused by a unexpected choice of the ACK transmission rate, is removed from the values reported in the first column of Tab. 4.2, still a small discrepancy with the theoretical values provided by Eq. (4.2) can be found. This can be accounted to elaboration overheads, that have experimentally assessed to be a random variable that equally affects all the measurements, whose mean value is about 50 μs with a very low standard deviation.

Tab. 4.2 also allows to observe and further confirm several conclusions made throughout this section. As an example, from the comparison between MCS 0 (i.e. 13.5 Mbit/s) for IEEE 802.11n and 6 Mbit/s for IEEE 802.11g, in the case of no interference, a significant reduction of the service time is obtained, as the average value decreases from 434 to 391 μs. Conversely, when the highest order modulations are employed, namely MCS 7 (i.e. 135 Mbit/s) for IEEE 802.11n and 54 Mbit/s for IEEE 802.11g, the "g" network reveals faster, confirming the theoretical analysis carried out in this section. Indeed, the impact of the new HT preamble on highest modulations is significant, since in this case

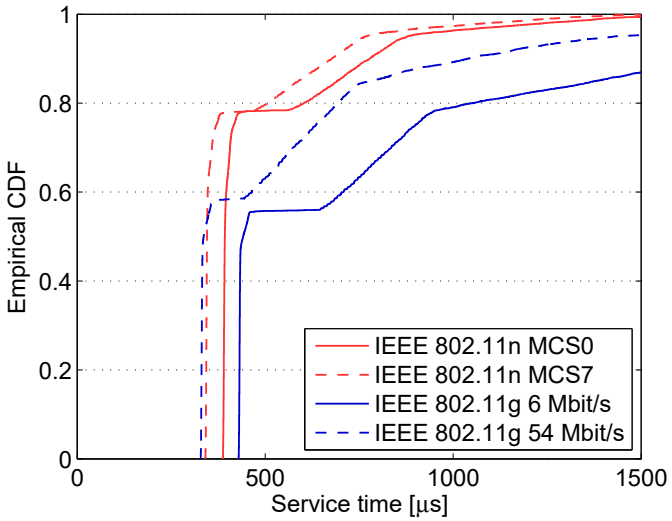


Figure 4.4: ECDF of service time with interference for 50 bytes packets.

the transmission of the considered low–size payload (50 bytes) fits in, most of the times, with a single OFDM symbol, with the result that the packet transmission time of IEEE 802.11g is actually lower than that obtained by IEEE 802.11n.

Considering now the interfering scenario, it can observed a strong increase of standard deviations in all the considered configurations, due to the impact of the retransmission procedure. Nevertheless, it can also be highlighted that the standard deviation for lower transmission rates (both in the case of IEEE 802.11g and n) is higher than that of the highest order modulations. This is due to the fact that, for a specific retransmission attempt, at the lower rates, the higher packet transmission time impacts directly on the variability of the service time.

The most significant result, that confirms the theoretical analysis provided in this section, is the definitely better performance that IEEE 802.11n STBC configurations are able to ensure with respect to IEEE 802.11g in a realistic interfering scenario. This is evident also looking at the Empirical Cumulative Distribution Function (ECDF) plots shown in Fig. 4.4, where it is clearly noticeable the effect of the random backoff times introduced by the CSMA/CA algorithm.

Indeed, a significant increase of the success probability at the first transmission attempt (from 60% to nearly 80%) is recognizable when IEEE 802.11n is adopted. This is explained by considering that, for a given SNR value, the introduction of STBC allows a significant reduction of PER, as observed in Fig. 4.3, and the consequent increase of reliability, also confirmed by the strong reduction of the service time standard deviation highlighted in Tab. 4.2. Such performance improvements are evident even in terms of

Table 4.3: Assessment of service time with 50 bytes packets (rate control enabled)

Configuration	No Interference		Interference	
	Mean	Std. Dev.	Mean	Std. Dev.
IEEE 802.11g	333.6 μs	19.1 μs	576.0 μs	493.1 μs
IEEE 802.11n (2×2 STBC)	344.8 μs	15.1 μs	434.3 μs	203.7 μs

average service time, that results lower for both IEEE 802.11n configurations.

As a final remark, the ECDFs show that, in the given scenario all packets are transmitted within 1500 μs for IEEE 802.11n. Conversely, with an IEEE 802.11g system, in the same environmental conditions, a non negligible percentage of frames (almost 10%) require a significantly longer time to be successfully delivered.

Impact of multi–rate support

A further set of tests has been conducted with the MRS enabled on the WNICs of Fig. 4.2(b). It is worth pointing out that the activation of the rate control algorithm does not generally require modifications neither to the IEEE 802.11MAC layer, nor to the firmware of the adopted WNICs. Indeed, RA can be enabled/disabled typically by a suitable configuration of the operating system, without significant implementation efforts.

In this context, the recent and widely adopted Minstrel rate control scheme was addressed, which is the default algorithm implemented by the Linux kernel (Minstrel), to investigate its impact on the service time. During its operations, Minstrel collects statistical data about the error rate of the previous transmission attempts for each of the allowed transmission rates, building a table based on the "recent past behavior" of the network, from which the next best transmission rate is chosen. The algorithm is tailored for best effort traffic, having the goal of maximizing long–term throughput, while the case of time–critical transmission, typical of most industrial applications, is not considered at all. The statistics of the service time, obtained from the experimental measurements with the rate control enabled, are reported in Tab. 4.3.

Comparing Tab. 4.3 with Tab. 4.2, for the case of the highest modulations, it is evident that the performance figures of both IEEE 802.11g and n are degraded when Minstrel is used. Specifically, while the mean values of the service time show only minimal differences, the standard deviations result, instead, considerably higher. This is due to the specific behavior of the Minstrel algorithm. Indeed, the analysis carried out revealed that the main cause of the increase of the service time jitter is represented by a feature of Minstrel which periodically selects, on a random basis, a transmission rate, ignoring the

rate that should be used (the best choice) for the purpose of gathering statistics about the channel status.

Conclusions

The "n" amendment to the IEEE 802.11 standard can represent an interesting opportunity for ICNs. In particular, the introduction of MIMO capabilities can be profitably exploited to enhance the reliability of data delivery, through an STBC scheme, rather than using the presence of multiple antennas to increase the throughput. Under this configuration, IEEE 802.11n links allow to achieve the same reliability level of comparable IEEE 802.11g installations with 2-3 dB less of SNR, as proved by extensive experimental campaigns. The increased reliability, combined with the faster data exchange brought by wider 40 MHz channels, allows to achieve a faster and more deterministic service time, ultimately enhancing the real–time performance of the network. Finally, when the default rate adaptation mechanism is enabled, the performance get worse instead of improving, specifically in terms of standard deviation of the service time. This result pushes towards the development of new rate control algorithms for IEEE 802.11n, specifically conceived for industrial applications, as it is be addressed in the following of this chapter.

4.2 Industrial rate adaptation algorithms

In the context of IEEE 802.11 WLANs, an important feature is represented by the MRS functions defined by the standard for any compliant device provided with a suitable set of available transmission rates which, in principle, allows a station to select the most suitable rate with the objective of improving performance. To this aim, RA algorithms have been made available since quite a very long time for different WLAN versions (Kulkarni and Quadri, 2009), and the selection of the transmission rate has been commonly based on an estimation of the transmission channel status. Unfortunately most of these strategies, such Automatic Rate Fallback (ARF), revealed unsuitable for industrial communication due to their inherent randomness that introduces additional uncertainty on packet delivery. Thus, in Vitturi et al. (2013) the authors introduced two new RA techniques specifically conceived for the industrial scenario, namely Static retransmission rate ARF (SARF) and Fast reduction rate ARF (FARF), that provide better performance in an industrial communication scenario.

In this section, industrial RA techniques as well as general–purpose ones, such as ARF and Minstrel, are described and evaluated. Moreover, an original algorithm for optimal RA in industrial WLANs is presented.

Legacy RA algorithms

Description

ARF The dynamic selection of the transmission rate carried out by ARF is based on the number of consecutive failed or, respectively, successful transmission attempts. In particular, given a transmission rate set, a station decreases its rate to the immediately lower one after K consecutive failed attempts. Conversely, after N consecutive successful attempts, the rate is increased to the immediately higher value. The default values for parameters K and N are, respectively, 2 and 10. To enhance the effectiveness of ARF, two additional features have been introduced. The first one specifies that, if the first transmission attempt after a rate increase fails, the rate is immediately restored at the previous value. The second feature is meant to avoid a station remains at a low rate for long time, and is achieved with the use of a timer, started when the rate is decreased, whose expiration triggers a rate increase, regardless of the number of successful transmissions collected yet.

The analysis carried out in Vitturi et al. (2013) for IEEE 802.11g showed that ARF, particularly in the presence of fast varying channel conditions, may introduce considerable randomness in packet delivery as well as increase the PER. These problems were solved, at least partially, with the definition of the two new ARF–based techniques, specifically conceived for industrial communication.

SARF SARF specifies that each retransmission attempt that takes place after a failure is carried out at the lowest transmission rate in the set supported by a station. Since the success probability at this rate is very high, the number of retransmissions (and hence the randomness introduced by the backoff procedures) will be limited. However, the dynamic rate selection of SARF is the same as ARF. In particular, successful retransmission attempts at the lowest rate are not taken into consideration to increase the rate. In practice, K consecutive failures at a specific rate (of different packets) trigger a rate decrease.

FARF FARF is based on a modified rate selection mechanism with respect to ARF. Specifically, after a failed transmission, the new rate is selected as the lowest one and then it will be incremented at the immediately higher value in the transmission set after N consecutive successful transmissions.

Minstrel The Minstrel rate control strategy Minstrel was firstly proposed in 2005 as part of the MadWifi driver, developed for Atheros chipsets in Linux–based system. Since

then, its popularity has increased thanks to the good performance that it offers in general purpose wireless networks subjected to best effort traffic. Consequently, Minstrel has been included as the default rate control algorithm in the Linux kernel, and in many popular wireless drivers currently employed by off–the–shelf devices, such as *ath5k* and *ath9k*.

With respect to other RA strategies, Minstrel can be classified as a random sampling algorithm, which tries to select, within the set of available transmission rates, the optimal one(s), based on the statistics collected for each rate during the previous communication history.

In practice, a station keeps a table (called retry chain) with four elements of type (R_i, C_i), $i = 1, \ldots, 4$, where each entry indicates a specific rate R_i and a number of attempts C_i. When a frame has to be transmitted, the station performs the first C_1 transmission attempts at rate R_1, then C_2 attempts at rate R_2, and so on. The retry chain is built using a specific procedure: R_1 and R_2 are chosen as the rates that guarantee the highest throughput (defined as the best transmission speed weighted by success probability), R_3 is the rate that yields the highest success probability and R_4 is the lowest available rate. The number of attempts C_i is computed as the maximum number of transmission attempts at rate R_i that can be performed in a window of length T_{max} for a generic frame with a payload length L_{ref}, taking into account the delays due to exponential backoff mechanism. The default values for those parameters are $L_{ref} = 1200$ bytes and $T_{max} = 6$ ms respectively.

The retry chain is updated with a period $T_u = 100$ ms adopting an Exponential Weighted Moving Average (EWMA) approach. Specifically, if $p_i(k-1)$ is the success probability for rate R_i before the update, and p_i^u is the success probability empirically computed on all frames transmitted at rate R_i during the current window T_u, the new value of the success probability for rate R_i is updated as

$$p_i(k) = p_i(k-1) \cdot \alpha + p_i^u \cdot (1-\alpha) \tag{4.3}$$

where α is the weight given to the previous history, whose default value is 0.75. After the transmission rates update, the four new entries of the retry chain can be selected according to the aforementioned rules.

Finally, Minstrel is a random sampling RA strategy in the sense that, with a probability P_s, it performs a frame transmission at a random rate R_s to "sample" the channel performance for that rate. Indeed, this random sampling procedure is used by Minstrel to gather statistics for all the available rates, in order to be always able to select the one better suiting the current channel conditions. From a practical point of view, this

sampling procedure is achieved by substituting the first two entries of the retry chain with the new one R_s and R_1, where the faster rate will be placed at the first position. In this way, the algorithm calculates a number of attempts C_s that have to be carried out with the temporary rate R_s, and hence is able to keep updated its table even for those rates that are less utilized.

From the performance point of view, Minstrel has been evaluated so far in general purpose networks (Xia et al., 2013), where it showed better performance than previous algorithms, for example ARF–like ones. Conversely, an evaluation of this strategy in an industrial communication scenario has never been addressed.

Performance analysis

The RA algorithms described so far, namely ARF, SARF, FARF and Minstrel, have been compared through numerical simulations based on Matlab. These simulations not only allowed to assess the performance gains brought by industrial–specific algorithms, but also suggested some possible improvements that enhanced the performance of these algorithms in an industrial context.

Simulations setup The network setup considered in this work is composed by one central station (the controller) which is connected to a set of M sensors/actuators (nodes), as described in Fig. 4.5.

All stations are compliant with the IEEE 802.11n standard and the configuration of their PHY and MAC layers is derived from the discussion provided in Sec. 4.1. Communications in the proposed network are based on a polling scheme, in which the controller node acts as a master, whereas the connected nodes represent the slaves. Given this protocol, the master issues a *polling–request* frame (with payload size L_{req}) toward a single node, which in turn responds with its *polling–response* frame (with payload size L_{data}). Once the polling procedure of node i is completed, the controller continues with the following node $i+1$ in the set, until the whole sequence of M nodes is completed. A polling trial fails if either a polling–request or a polling–response frame is lost, within a maximum number of possible transmission retries, which has been set to n_{max}. The time to complete the polling of the entire sequence of nodes is hence stored as a measure of the current cycle time, that represents a significant performance indicator for this study.

The simulations do not take into account external interference effects, and hence the source for erroneous receptions is only found in the bad performance of the wireless channel. At each received frame, the simulator calculates the perceived SNR level sampling the channel status at the frame start. The SNR level is then used in the

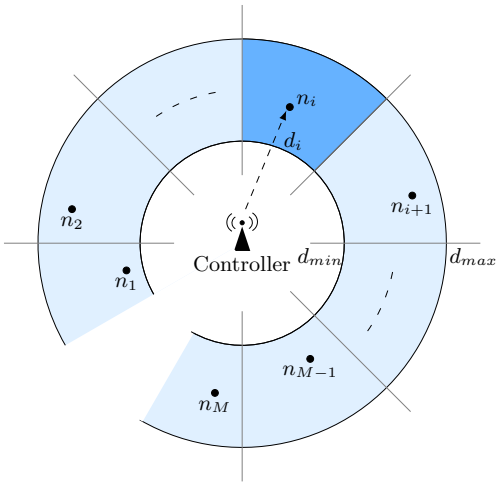


Figure 4.5: Sketch of the adopted simulation setup.

decision for the correctness of the reception, based on the experimentally measured PER curves of Fig. 4.3.

The developed simulator takes into account several effects relevant to an indoor wireless medium in the ISM band, basically a variable transmit power, the log-distance path-loss model, and a freely adjustable noise power at the receiver. Moreover, since the considered communication scheme is a multi–antenna system, a suitable model for small–scale fading is also provided. To model accurately the fading effects on a MIMO channel, the analysis originally carried out by the IEEE 802.11TGn group, in charge of the development of the standard, was adopted. In Erceg et al. (2004), six different MIMO channel models are defined, identified with the capital letters from A to F. In the perspective of a wireless communication system applied in a typical industrial environment, the network of Fig. 4.5 is deployed to manage a single production island. Hence, in this scenario, the TGn model F represents the best choice to model fading effects on the wireless medium since, even if nodes are placed in a limited area around the controller, reflections and multi–path effects could result from objects far away from the considered stations, in agreement with the assumption of that channel model (Erceg et al., 2004). Consequently, all the simulation outcomes have been obtained assuming the aforementioned model.

Finally, singularities due to particular positions of the stations as well as to particular network configurations have been reduced by exploiting a random node placement strategy. Specifically, with reference to Fig. 4.5, the circular area is split in M contiguous sections. Within each section, during the network setup phase each node n_i is placed in a random position identified by (d_i, θ_i), which are two values chosen randomly from a uniform

Table 4.4: Main simulation parameters

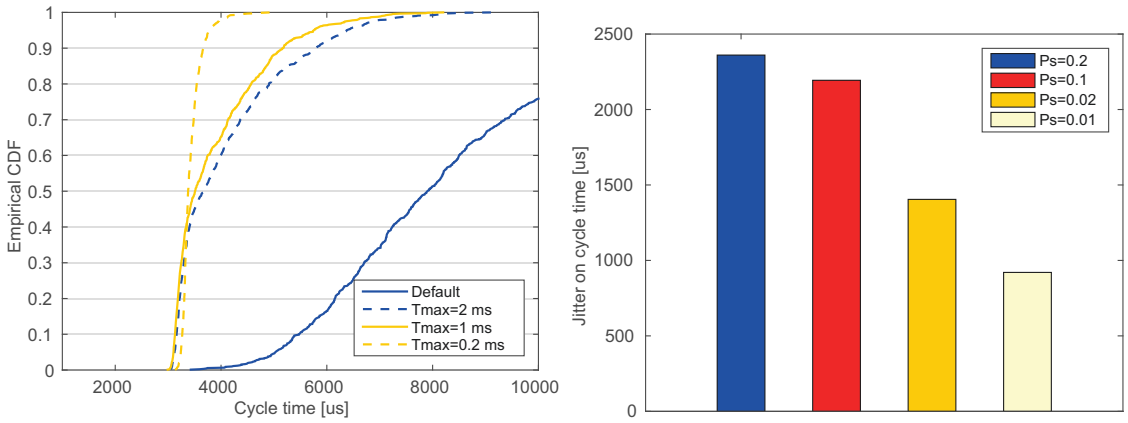
Parameter	Description	Value
M	Number of slave nodes	10
N_s	Number of simulations	100
N_c	Network cycles for each simulations	10000
d_{min}	Minimum distance between controller and node	3 m
d_{max}	Maximum distance between controller and node	6 m
L_{req}	Payload size of polling–request frame	50 bytes
L_{data}	Payload size of polling–response frame	50 bytes
P_{tx}	Transmit power	100 mW
n_{max}	Maximum number of transmission attempts	7
CW_{min}	Initial size of the contention window	15

distribution, and d_i ranges between $d_{min} = 3$ meters and $d_{max} = 6$ meters, the minimum and maximum allowed distances, respectively, chosen in such a way that even the farthest node will be able to communicate with the controller. Furthermore, each simulation is repeated a number N_s of times, and the placement of nodes is randomly generated at each repetition. The full set of simulation parameters is reported in Tab. 4.4.

Tuning of Minstrel algorithm The Minstrel algorithm was designed for office networks, where the traffic patterns and desired behavior are typically very different from those of industrial communication scenario. As a consequence, its parameters configuration can be optimized to better meet the requirements and the expected performance of this new context.

A first step toward an enhanced version of the Minstrel algorithm is represented by the revision of the reference payload size L_{ref}. Indeed, the default value of 1200 bytes is actually much greater than that typically adopted by industrial networks. Therefore, in accordance with the parameters reported in Table 4.4, the reference payload size has been set to $L_{ref} = L_{data} = 50$ bytes, since this value is representative of a large set of industrial applications, and is also able to capture for frame coming from a wired Ethernet segment and encapsulated within a wireless frame.

Such a lowering of the reference payload size has a direct influence on the computation of the throughput associated with each rate, and also strongly impacts on the number of possible attempts C_i that the algorithm is able to perform for each rate, thus reflecting on network performance. Indeed, with a smaller reference payload size, a higher number of transmission attempts can be accommodated in a window of length $T_{max} = 6$ ms. In particular, it was observed that within the allowed transmission window the value of C_i calculated by the algorithm is always equal to the maximum number of transmission



(a) Cycle time versus T_{max} **(b)** Jitter on cycle time versus P_s

Figure 4.6: Tuning of various parameters of the Minstrel algorithm.

attempts for a frame, n_{max}, given the fact that with a very small reference payload a frame takes much less time to be transmitted. As a result, only the first rate in the retry chain happens to be used, which is clearly an undesired behavior.

It is clear that such an issue is raised up from the choice for T_{max}, whose default value of 6 ms was empirically derived from measurements on a TCP–based office network, as the authors of Minstrel reported (Minstrel). However, in industrial applications with generally much lower service times, that value revealed not suitable and also not well balanced with the previous tuning of the value of L_{ref}. Indeed Fig. 4.6a reports the ECDF of the obtained cycle time for varying values of T_{max}.

Here the default configuration is that with $L_{ref} = 1200$ bytes and $T_{max} = 6$ ms, while the other configurations adopt the new reference payload size $L_{ref} = 50$ btes. It can be clearly observed that with a reference payload size of 50 bytes, the cycle time behavior is much more deterministic for lower values of T_{max}, and the best results are obtained with a period of $T_{max} = 200$ μs, which will be hence considered in the following as the optimal value for this parameter.

Another issue identified in the Minstrel algorithm is the sampling probability P_s, which is set by default to 0.1, meaning that one packet out of ten is sent at a random rate to gather statistics. This can be regarded as a quite high value in the considered context, since it introduces a very high degree of randomness in the packet delivery time. An increased determinism in rate selection can be actually achieved by decreasing the sampling probability P_s, as highlighted by Fig. 4.6b, which reports the jitter on cycle time for different values of P_s, keeping fixed all the other parameters, including L_{ref} and T_{max}.

Table 4.5: Cycle time for Minstrel algorithm with different values of P_s and T_u

Metric	$\mathbf{T_u = 100\ ms}$		$\mathbf{T_u = 1\ s}$	
	$\mathbf{P_s = 0.1}$	$\mathbf{P_s = 0.02}$	$\mathbf{P_s = 0.1}$	$\mathbf{P_s = 0.02}$
Average cycle time	8.37 ms	6.73 ms	6.41 ms	5.57 ms
Jitter on cycle time	3 ms	1.57 ms	2.23 ms	1.58 ms

From the figure it could be inferred that P_s must be as low as possible to improve real–time performance of Minstrel. However, it has to be noted that a very low value of P_s along with a short update window T_u may result in an almost static behavior of Minstrel algorithm. Indeed, if the window is short and the sampling probability is low, very few information about rates different from the current ones will be gathered, and the algorithm will be strongly dependent on the initial rates set and poorly reactive to changes in the environment. As a consequence, very low sampling probabilities, such as $P_s = 0.01$, should be avoided.

Tab. 4.5 shows the average and standard deviation of the cycle time for different values of P_s and T_u (all other parameters are set to default values). It is evident that a decrease in P_s should be balanced by an increase of T_u in order to reduce both mean value and jitter of the cycle time under the same environmental conditions.

Finally, based on the above considerations, the following optimized parameters configuration has been chosen for the Minstrel algorithm in the proposed industrial communication scenario: $L_{ref} = 50$ bytes, $T_{max} = 200$ μs, $T_u = 1$ s, $P_s = 0.02$. The weight α has been kept equal to 0.75 since further simulations showed that it does not have any defined impact on the cycle time behavior.

Improving rate control strategies All the rate control algorithms discussed so far rely on the previous communication history to select the optimal transmission rate. The assumption behind this behavior is that consecutive samples of the communication channel are correlated. However, one should consider that the communication links between each possible node pair in the network form a set of different realizations of the shared wireless medium, all of them characterized by a different behavior. Whereas, when a generic station A has to send a packet, on the basis of the discussed rate control scheme it selects the next transmission rate by looking at the past channel history regardless of the intended receiver. This means that the rate used to send a packet to a destination node B is likely influenced by the rate(s) previously selected to send packets to another destination node, say node C, despite the fact that the communication channel between A and B is generally uncorrelated with that between A and C. As a consequence of this

fact, the current RA algorithm implementations can lead to poor system performance, especially for stations that have packets for multiple receivers.

In the simulated industrial network, this issue is never encountered by the slaves, since they only exchange data with the controller. Conversely, it occurs for the controller and, hence, it can significantly affect the performance. To overcome this problem, smarter rate control strategies to be implemented in the controller are provided, that are able to differentiate the previous communication history according to the suitable destination node, and hence only exploit the statistics relevant to that destination node when choosing the transmission rate.

From an implementation point of view, in the case of SARF and FARF strategies a node needs only to keep track of the previous rate and the number of consecutive failed/successful transmissions at this rate, whereas with the proposed modification, the controller needs to store those information for each of the M slave nodes in a suitable vector. Analogously, in the case of the Minstrel algorithm, the controller needs to store a different retry chain and differentiated rate statistics for each of the M slave nodes. In both cases, the previous communication history is accounted by looking only at the result of the packet exchange with the single node of interest. Clearly, this modification requires a higher amount of memory in the controller to be used by the RA algorithm, especially for the case of the Minstrel algorithm. However, this does not seem to be an issue, since the controller is generally implemented by devices with adequate resource availability.

To prove the effectiveness of the adoption of a smarter controller, in Fig. 4.7 the ECDF of the cycle time for the two different implementations of the FARF rate control strategy are provided, indicating with FARF the standard strategy and with FARF 2 the one with a smarter controller. The other RA schemes show a similar behavior, and have been omitted in the figure to avoid clutter. As can be seen, the benefits in terms of system timeliness are evident.

Comparison of different RA strategies The next Fig. 4.8 reports the ECDF of the cycle time when the four RA techniques considered in this assessment, namely ARF, SARF, FARF and Minstrel, are adopted in their standard (i.e., without any tuning) implementation. It is given here to provide a clear picture of the expected performance of these techniques when introduced in an industrial applications context.

This simulation also provide a confirmation that the Minstrel algorithm in its standard implementation is totally unsuitable for industrial communication, mainly because of the high sampling probability P_s and the fact that all retransmissions are performed at the initial rate because of the values assigned to L_{ref} and T_{max}. Conversely, FARF, SARF

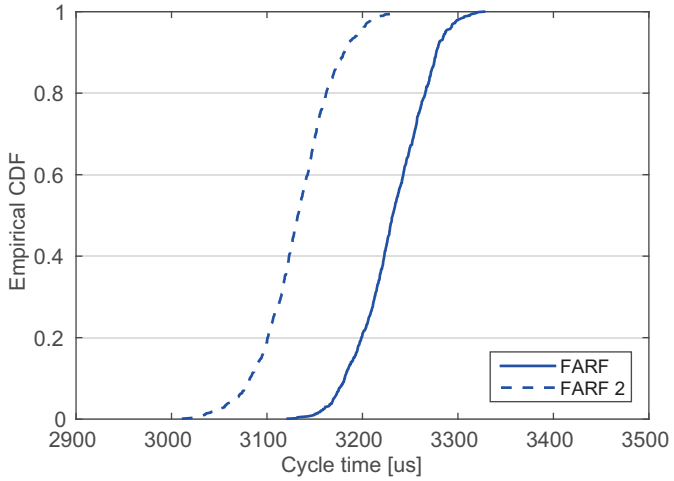


Figure 4.7: Cycle time for standard FARF, and FARF with a smarter controller (FARF2).

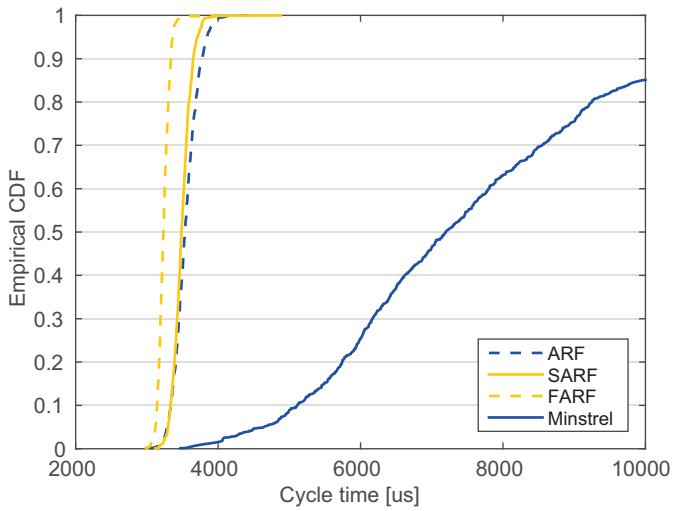
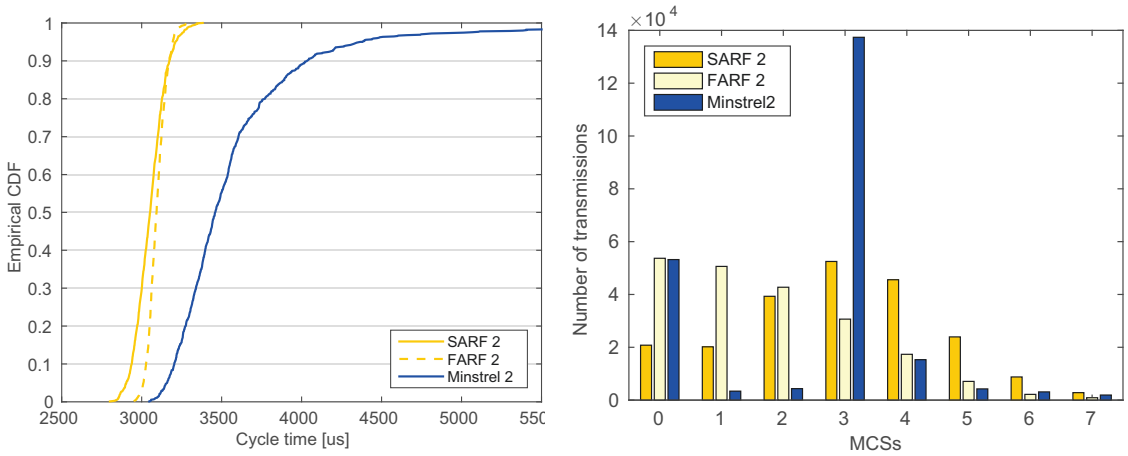


Figure 4.8: Cycle time with different RA algorithms in their standard implementation.

and even ARF exhibit a much more deterministic behavior, with FARF outperforming the other two techniques.

A subsequent evaluation has been carried out by considering the proposed improved versions of the rate control algorithms. Specifically, FARF 2 consider the use of the discussed smart controller, while in SARF 2 both the smart controller is introduced and, also, K has been reduced from 2 to 1 to further improve robustness. Finally, Minstrel 2 has the parameter configuration previously determined, namely $L_{ref} = 50$ bytes, $T_{max} = 200\mu$s, $T_u = 1$ s and $P_s = 0.02$, along with the smart controller modification. The ECDF of the cycle time is reported in Fig. 4.9a.

As can be seen, the behavior of Minstrel is much improved with respect to that of



(a) ECDF of cycle time **(b)** Histogram of selected MCSs

Figure 4.9: Performance comparison of improved RA schemes.

Fig. 4.8, but the enhanced versions of both SARF and FARF behave still rather better, providing both a lower average cycle time and a lower bound on the maximum cycle time (almost 3.5 ms versus more than 5.5 ms). The improved SARF 2, with a smart controller and K=1, is slightly faster than FARF 2, while the latter is more conservative, yielding a reduced jitter.

To better understand the behavior of the different RA strategies, Fig. 4.9b reports the histogram of the MCSs adopted during the whole simulation by all nodes and the controller, showing the results for each of the three proposed rate control strategies. The different approach of each strategy is clearly visible. SARF 2 tends to explore all the rates with a prevalence of MCS 3 and, in general, lower rates are preferred to higher ones due to the conservative rate selection scheme and the channel conditions. The same holds true for FARF 2 which, however, adopts more often MCS 0, then MCS 1 and so on in a descending order. This is due to the fact that, each time there is an error, FARF restarts from the lowest MCS, and hence lower MCSs are explored more often. Finally, with Minstrel 2 two rates are mostly adopted, namely MCS 0, which gives the best transmission probability, and MCS 3, which guarantees the highest throughput weighted for success probability (in the simulated channel conditions).

The Rate Selection in Industrial Networks (RSIN) algorithm

The analysis carried out so far (that, it is worth remembering, is derived from theoretical models as well as from simulations), represents a valuable step in the context of RA strategies for IEEE 802.11real–time industrial applications. However, further investi-

gations can be envisaged toward more exhaustive achievements. Particularly, new RA algorithms need to be designed to ensure a prompt reaction to channel status variations without unnecessarily penalizing the transmission rate. Also, the practical feasibility of the different RA techniques has to be addressed and a comparative performance analysis needs to be carried out.

In this section, a new technique, named RSIN, is proposed, characterized by the following main features:

- dynamic identification of the channel status, exploiting device–measured SNR levels;
- rate selection based on a constrained minimization of the packet error rate;
- knowledge of the deadline on packet delivery time.

It is worth observing that some general purpose RA techniques based on the SNR knowledge were already proposed, as for example Received Based Autorate (RBAR) (Holland et al., 2001). However, with such strategies, the SNR value was obtained via the exchange of RTS/CTS frames, a procedure that increases the transmission overhead and, as such, negatively impacts on the behavior of real–time applications.

Formal description

RSIN has been conceived to target real–time industrial communications, and hence configurations and protocols typical of such a scenario (Sauter, 2010) will be addressed. In this context, it is quite customary that a central controller is in charge of managing a set of sensors/actuators. During operations, the controller exchanges process data with each sensor/actuator, either on a cyclic basis or triggered by specific events, and the generated traffic is characterized by a prevalence of scheduled transmissions that have to be completed within tight deadlines. The most common wireless configuration that reflects such a scenario is an infrastructure WLAN, where the central controller is connected to the AP, and sensors/actuators are located on some wireless STAs.

From the design perspective, RSIN is based on two main assumptions. The first one specifies that in any data exchange between two stations, each packet has to contain an additional field in which the transmitting node inserts the perceived SNR relevant to the last received packet from the other node. This is feasible, since the SNR value can be evaluated by the WNICs adopted by the stations and then included in the payload of the exchanged frames, as briefly sketched in Fig. 4.10. Indeed, the SNR evaluation can be carried out on a per–packet basis by extracting the Received Signal Strength

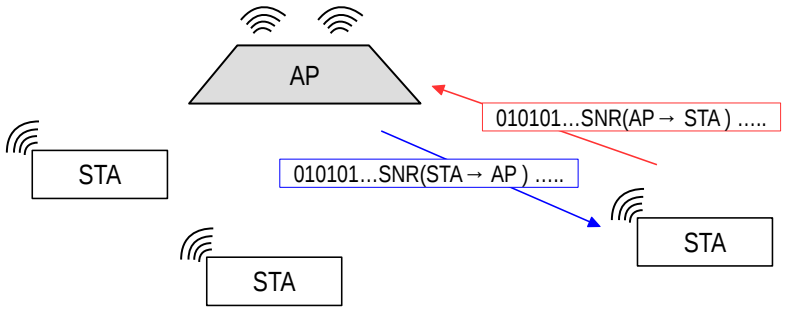


Figure 4.10: Inclusion of the SNR value in the payload of exchanged frames.

Indicator (RSSI) from the incoming frame, and then subtracting from this value the noise floor power. To this regard, it is worth recalling that both the RSSI and the noise floor level have to be measured by the WNIC to correctly perform frame decoding. Hence, the availability of the SNR value to RSIN does not depend on the physical features of the WNICs but, rather, by their device drivers that may or may not provide such an information.

It is also worth observing that the measured SNR value can be typically stored in one byte. As such, in the considered scenario, where usually small amounts of data are exchanged,adding a field with this information to the frame payload has a negligible impact on the overall frame size, as well as on its delivery time.

RSIN further assumes that any transmitting node is aware of the relationship between the PER and SNR for any possible transmission rate. Such an information can be actually derived from theoretical analyses (Perahia and Stacey, 2013), or through extensive experimental measurements campaigns, as the ones presented in Tramarin et al. (2016b).

This assumption allows to state that each node is provided with the map:

$$\mathcal{F} : S \times \mathcal{R} \rightarrow \mathcal{P} \tag{4.4}$$

where S represents the set of possible SNR levels, $\mathcal{R}$ is the set of the available transmission rates[2] and $\mathcal{P}$ is the set of probability values. Indeed, the outcome of the map is the probability P_e with which the next frame transmission fails, given a particular combination of R and S. Clearly, P_e is a real number which belongs to the range $[0, 1]$. It is worth noticing that the PER also depends on the transmitted frame size L, and hence the provided map $\mathcal{F}$ should scale accordingly.

The RSIN technique is defined as an optimization problem. Given a packet to be transmitted with a deadline D, and a specific transmitter–receiver pair, the problem can

[2]For instance, $\mathcal{R}$ can be constituted by the 4 different IEEE 802.11b rates, the 8 IEEE 802.11a/g ones, or the various MCSs available for IEEE 802.11n.

be formulated as to find the number of attempts and the relevant sequence of rates to be used for the transmission of that packet, with the twofold goal of minimizing the residual transmission error probability, while ensuring the packet is delivered within its deadline. This reflects in the solution of the following minimization problem

$$\min_{N \leq N_{max},\, r^{(i)} \in \mathcal{R}} \mathcal{L}\left(L, S, N, r^{(1)}, r^{(2)}, \ldots, r^{(N)}\right) \tag{4.5}$$

subject to the constraint

$$\max_{N \leq N_{max},\, r^{(i)} \in \mathcal{R}} \mathcal{D}\left(L, S, N, r^{(1)}, r^{(2)}, \ldots, r^{(N)}\right) \leq D \tag{4.6}$$

In Eq. (4.5), $\mathcal{L}(\cdot)$ is a function that calculates the residual packet error probability for a packet with a payload of L bytes, transmitted to a receiver which perceives a SNR level of S dB, after N consecutive transmission attempts have been carried out at the rates $r^{(1)}, r^{(2)}, \ldots, r^{(N)}$ (where $r^{(i)}$ is the rate selected for the i–th attempt). Moreover, in both Eq. (4.5) and Eq. (4.6) the condition $N \leq N_{max}$ has to hold, where N_{max} is the default maximum number of attempts specified by the IEEE 802.11 standard, typically set to $N_{max} = 7$.

The constraint imposed to the minimization function is relevant to the frame delivery time, $\mathcal{D}$. This is defined as the time elapsed from the instant in which a packet starts to be transmitted to the instant in which the transmitter receives the correspondent ACK frame. In the considered real–time communication scenario, $\mathcal{D}$ has to be lower or equal to the deadline D.

RSIN is invoked to obtain, within all the possible combinations of N (number of transmission attempts) and the corresponding rates $r^{(1)}, \ldots, r^{(N)}$, the sequence that represents the optimal solution to the problem in Eq. (4.5), by considering the most updated level of SNR S perceived between the transmitter–receiver pair, the map $\mathcal{F}$ and the constraint of Eq. (4.6). This solution is constituted by the sequence of rates at which any single attempt of transmitting the packet has to be carried out.

Considering that the minimization problem through the map $\mathcal{F}$ assumes a probabilistic behavior, the maximum value of $\mathcal{D}$ to be used in Eq. (4.6) is determined under the worst–case assumption that the first $N - 1$ consecutive attempts are failed, whereas the N–th

one is successful.Therefore, the expected maximum delivery time can be expressed as

$$\max \mathcal{D}\left(L, S, N, r^{(1)}, \ldots, r^{(N)}\right) = N \cdot t_{DIFS} + t_{data}(L, r^{(1)}) + \sum_{i=1}^{N-1} \left[t_{ACK\_TO}(r^{(i)}) + t_{slot} \cdot \max[I_{bo}(i)] + t_{data}(L, r^{(i+1)})\right] + t_{SIFS} + t_{ack}(r^{(N)}) \tag{4.7}$$

where t_{DIFS} and t_{SIFS} are the duration of the DIFS and SIFS periods, respectively, while the term $t_{data}(L, r^{(i)})$ represents the actual transmission time, at the i–th attempt, of a frame with payload of L bytes at rate $r^{(i)}$. Moreover, $t_{ACK\_TO}(r^{(i)})$ is the ACK timeout, i.e., the maximum time a node waits for the reception of an ACK frame before considering its transmission as failed. Then, $t_{ack}(r^{(i)})$ is the time to transmit the ACK frame given that the rate $r^{(i)}$ is used to transmit the originating data frame. Finally, the IEEE 802.11 CSMA/CA procedure introduces after each failed attempt a random backoff time, denoted with $I_{bo}(n)$, drawn from a uniform distribution. The maximum duration of such a time at the i-th transmission attempt is a multiple of the slot time t_{slot}, depends on the CW length and can be expressed as $2^{i-1} \cdot (CW_{min} + 1) - 1$. The aforementioned values can all be retrieved from the IEEE 802.11 specifications, and clearly depend on the selected physical layer.

Solution of the optimization problem The solution of the problem formulated by Eq. (4.5) and Eq. (4.6) requires to deal with some issues that impact on the practical implementation of RSIN.

The first issue is concerned with the possible existence of more than one valid solution. Indeed, it is likely that a set of optimal rate sequences along with the corresponding number of attempts exist, each solution satisfying both Eq. (4.5) and Eq. (4.6). For instance, if the deadline is long enough, it may happen that several combinations, all constituted by a sequence of attempts at the minimum transmission rate, would easily solve the optimization problem.

However, since in such cases RSIN has to perform some further selection steps, some suitable selection rules can be drawn to further optimize the final solution. First, among the above set of solutions, RSIN selects the sequences with the minimum number of transmission attempts. Indeed, the lower the number N, the lower the jitter on frame delivery, since less backoff procedures are necessary to successfully complete the transmission. Actually, even after this step it is not ensured to obtain a single solution. Therefore, considering that the remaining solutions already ensure the minimization of

the transmission error probability (this is a priori, by design of RSIN) and the minimum number of transmission attempts, RSIN will eventually choose the rate sequence that minimizes the delivery time $\mathcal{D}$, without affecting any of the previous constraints. This will also ensure to increase the real–time throughput, a further meaningful performance indicator, since unnecessarily low–rate combinations are avoided.

Another important issue is concerned with the solution of Eq. (4.5) and Eq. (4.6). Indeed, such a solution actually represents a non–linear problem, that depends on a high number of variables. For this reason, the search for the exact optimal solution can be carried out through a "brute force" approach, which explores the whole set of possible combinations. This means that, trivially, this brute force algorithm has to iterate on each one of the $(N+1)$-tuple of type $N, r^{(1)}, \ldots, r^{(N)}$, check if the constraint in Eq. (4.6) is met, and calculate the residual error probability. Unfortunately, such an approach may lead to a considerable number of iterations and, possibly, to high processing times. Indeed, if R is the number of available MCSs and N that of the available transmission attempts, the resulting number of iterations is in the order of $\propto R^{N+1}$. Thus, for example, in the common case of R=8 and N=7, RSIN would have to complete ∼19 million iterations.

It is hence necessary that adequate strategies to limit the number of iterations are undertaken. In this direction, the most immediate option is the reduction of the solution space size. Thus, a limit on the number of subsequent transmission attempts N_{max} may be imposed (this has the further benefit of limiting the jitter on the frame delivery time). However, the most important assumption introduced for the solution of the problem in Eq. (4.5) is the constraint that the resulting sequence of transmission rates has to contain only monotonically decreasing values:

$$r^{(1)} \geq r^{(2)} \geq \ldots r^{(N-1)} \geq r^{(N)} \tag{4.8}$$

This is highly reasonable, since progressively reducing the transmission rate of the subsequent attempts in an IEEE 802.11 network, as this assumption does, allows in general to decrease the residual packet transmission error probability, which represents the goal of the RSIN technique. With the above assumptions a "heuristic" solution of the problem formulated by Eq. (4.5) can be obtained, that allows to considerably decrease the number of iterations, with respect to the "brute force" approach, as can be observed in Fig. 4.11, while providing the same results to the constrained minimization task.

Experimental setup

To assess the performance of RSIN, it is required its actual implementation on real devices, as well as the deployment of an adequate prototype network that emulates a

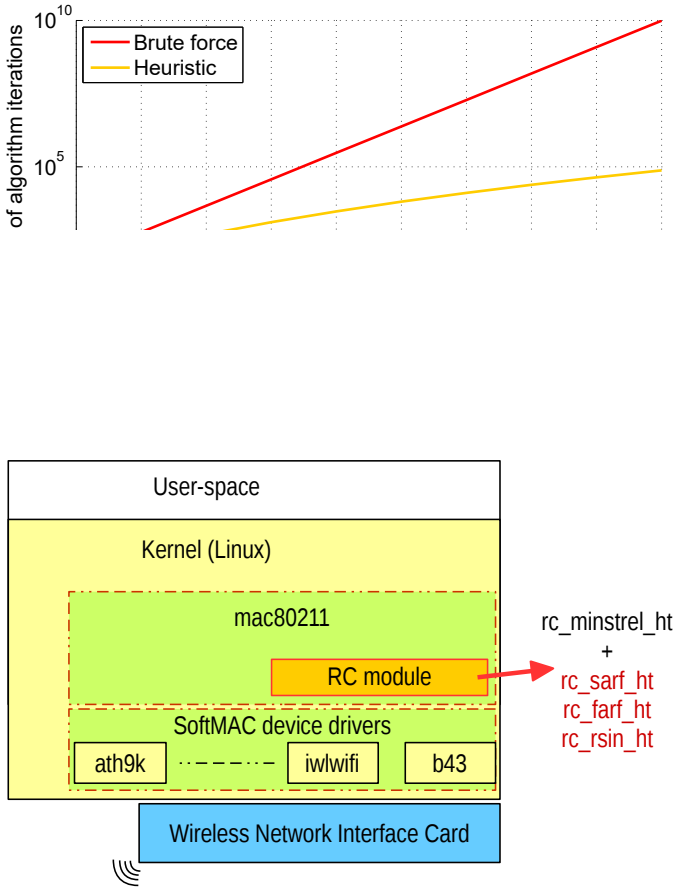


Figure 4.12: Schematic representation of the internal Linux kernel structure relevant to the IEEE 802.11 stack.

typical industrial communication scenario.

Implementation of RA strategies The RSIN technique has been implemented on some devices based on commercial WNICs, along with both SARF, FARF and Minstrel (all tuned according to the considerations carried out at the beginning of this section).

To provide an effective implementation, the IEEE 802.11 networking architecture provided by the Linux kernel, briefly depicted in Fig. 4.12, was exploited. Within this framework, any implementation of RA techniques has to reside within the *mac80211* kernel module, as highlighted in the rightmost part of the figure. At the beginning of a packet transmission procedure, any RA algorithm has to provide the WNIC driver with the list of rates to be used for each subsequent transmission attempt. In the case of RSIN, such a list is that obtained from the solution of the optimization problem expressed by Eq. (4.5). Since the computational burden of the RSIN algorithm may impact on the

performance of the stations that use it, a specific assessment was carried out to estimate the time it takes to complete the solution of the problem formulated by Eq. (4.5) and Eq. (4.6) on the PCs used in the prototype network considered in this section. Under the assumption that the set of MCSs contains eight different transmission rates, and that a frame may eventually undergo up to four retransmissions, this processing delay is actually bounded to 50 μs.[3] Clearly, the implementation of RSIN on different devices could require different times.

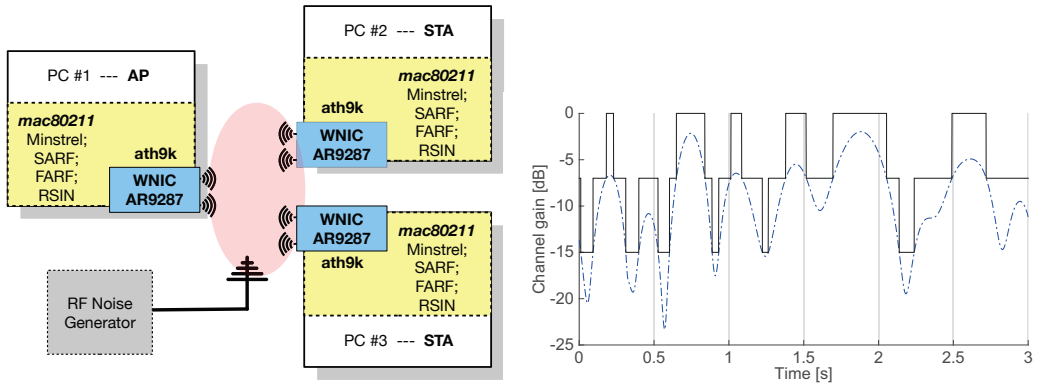
In the proposed implementation, the RSIN algorithm is called by the WNIC immediately after the reception of a frame (that carries the measured SNR value) from the partner, and hence its execution has to be concluded before the next transmission. This always happened in all the tests carried out. However, in case the execution of the algorithm is not completed before the next scheduled transmission, the choice of transmitting at the last selected rate was adopted. From a practical point of view, this reflects in a slightly reduced responsiveness of RSIN, that actually would employ more time to adapt to channel status variations.

The above described evaluation of the RSIN execution time clearly refers to the most general operational context in which the frames to be transmitted have different payloads and deadlines, so that each new transmission requires a new complete execution of the algorithm. However, for the case in which all frames share the same length and are subjected to the same deadline, an alternative, more convenient, solution could be devised. Indeed, given the map $\mathcal{F}$, the WNIC may initially (off–line) execute the RSIN algorithm to build a look–up table where the final rate sequence is stored for each possible value of the SNR. This is actually feasible, since in this case the unique variable left is the SNR value. Consequently, the selection of the suitable rate sequence to be used for the transmission of a packet simply reduces to a search procedure within the look–up table, with a considerable reduction of the computational burden.

As a concluding remark, in the proposed implementation the map $\mathcal{F}$ exploited by RSIN has been retrieved through the extensive measurement campaign described in Sec. 4.1.

Prototype network The software modules that implement the aforementioned RA techniques have been introduced in some desktop workstations (Dell Optiplex PCs, models 745, 755 and 960), all running the Ubuntu 14.10 Operating System based on the Linux kernel version 3.16.4. The adopted workstations were equipped with WNICs by TP-LINK (models TL-WN851ND and TL-WN881ND), each one compliant with the

[3]This value has been measured on a Dell Optiplex PC model 755, equipped with an Intel Core 2 Quad Q6600 processor and 4 GB of RAM.



(a) Sketch of the setup

(b) Synthesized communication channel

Figure 4.13: Prototype network for the experimental assessment of RSIN.

IEEE 802.11n standard, and allowing 2×2 MIMO operations. The cited WNICs exploit an Atheros AR9287 chip, so that they leverage the "SoftMAC" device principle, allowing a fine–grained control of the transmission path from the kernel–space device drivers. In this specific case, they are handled by the open–source *ath9k* module.

The above workstations have been deployed in a prototype network that comprises three nodes, as schematically represented in Fig. 4.13a. Such a network allows an effective control of devices and wireless medium, without sacrificing the generality of the obtained results. The network is configured in infrastructure mode, where one workstation behaves like an AP, while the other ones, placed at 2 meters from the AP, act as IEEE 802.11 STAs. A software that implements the desired exchange of packets between network nodes at the MAC layer was implemented. Such an application is clearly based on a master–slave architecture, where the master generates polling requests on a periodic basis, to which each slave responds immediately.

All the experiments have been carried out on an IEEE 802.11n network. The configuration parameters for both the PHY and MAC layers were set in agreement with the analysis provided in Sec. 4.1. The main network parameters adopted in these tests are summarized in Table 4.6.

The experimental measurements have all been carried out in a research laboratory where, unfortunately, a complete electromagnetic isolation was not achievable. However, a channel not steadily used by other WLANs was selected by monitoring the surrounding environment with a real–time spectrum analyzer.

As far as the physical channel is concerned, it is worth remembering that an industrial environment is typically affected by quite relevant fading effects, reflections from metallic surfaces, possibly long communication distances and multi–path interference. Most of

Table 4.6: IEEE 802.11n parameters (2.4 GHz band)

Description	Value
MIMO configuration	2×2 STBC
Channel Bandwidth	40 MHz @ 2.4 GHz
MCSs	0-7
Transmission rates	13.5, 27, 40.5, 54, 81, 108, 121.5, 135 Mbit/s
Slot Time	9 μs
DIFS	28 μs
SIFS	10 μs
Max number of MAC–layer retries, N_{max}	7
Payload size	50 bytes

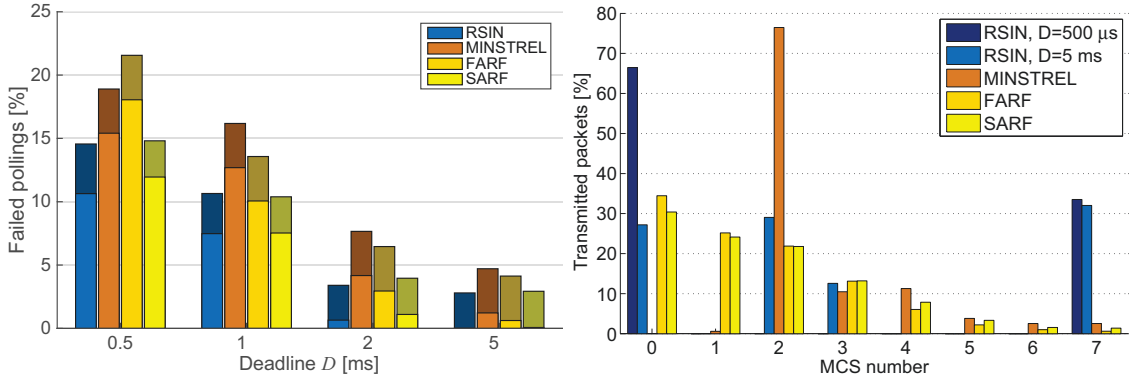
these aspects have been taken into consideration by the IEEE 802.11 TGn that developed adequate channel models, including model "F" specifically conceived to describe the industrial scenario (Erceg et al., 2004). A realization of such a channel is provided in Fig. 4.13b (dashed blue lines), in terms of channel gain behavior over time.

To emulate in practice such a channel, an RF signal generator (model Agilent E4433B) was used as a controlled artificial source of impairments. Specifically, the instrument was set to yield a wide–band AWGN–like noise, whose power level was modulated to mimic the fluctuations of the channel gain as described by the considered model "F". However, for feasibility reasons, the channel gain behavior was quantized using three levels and the resulting pattern, described by the solid black lines in Fig. 4.13b, was adopted to modulate the RF generator power. This artificial disturbance, centered on the carrier frequency of the selected channel, was then injected on the medium through a directional antenna with the main lobe directed toward all the WNICs.

As can be inferred from Fig. 4.13b, the synthesized channel will result in three significantly different SNR levels at the receivers. The lowest one has been calibrated to block any transmission but those at the lowest rate, which will be anyway impaired as well. The intermediate level will seriously impair only the highest MCSs. Finally, in the absence of injected noise, all transmission rates can be exploited without significant errors, even if occasionally some interference from co–located mobile devices may arise.

Performance evaluation

The outcomes of a comprehensive performance assessment carried out on the proposed RSIN as well as on the other RA techniques, are presented here.



(a) Percentage of missed deadlines **(b)** Histogram of employed MCSs

Figure 4.14: Experimental comparison among the different RA techniques.

Experimental assessment The first experiments presented are concerned with the ability of a station to successfully deliver a packet within a specific deadline.

In these experiments, the AP in Fig. 4.13a continuously polls the two STAs. For each query, the AP sends a request frame to the addressed STA which, consequently, answers with a response one. If a STA does not answer within a specific timeout (set to a value much higher than the deadline), the polling is considered as failed and the AP moves to the subsequent STA. Several experimental sessions have been carried out for different deadline values, each one comprising 10000 network cycles. Then, by looking at the received packets, the number of missed deadlines for each RA scheme are analyzed and the obtained results are reported in Fig. 4.14a.

It is worth considering that in the experiments a failure may also arise, so that a transmission results completely unsuccessful and the packet lost. These occurrences, have been considered separately from those relevant to a successful transmission but with a delivery time exceeding the deadline. To this regard, Fig. 4.14a summarizes the aforementioned results for some significant deadline values. Each bar is subdivided in two parts, with the lower one relevant to the percentage of missed deadlines on the total number of delivery attempts, whereas the upper part reports the percentage of transmission failures. Focusing only on the percentage of missed deadlines, the figure highlights that the proposed RSIN algorithm greatly outperforms Minstrel and FARF and provides performance on par with SARF.

A further performance indicator of interest in this assessment is represented by the statistics of the delivery time of the frames involved in polling operations, that are presented in Table 4.7. The results are relevant to all the considered RA techniques. In particular, to provide a more exhaustive assessment, the behavior of RSIN has been

Table 4.7: Delivery time statistics

RA technique	Mean	Standard Deviation
RSIN, D=0.5 ms	502.7 μs	325.9 μs
RSIN, D=5 ms	477.2 μs	304.6 μs
Minstrel	635.2 μs	883.1 μs
FARF	630.0 μs	947.8 μs
SARF	516.7 μs	458.7 μs

analyzed for two different deadline values, namely 0.5 ms and 5 ms. Moreover, since in the experimental setup the channel behavior was the same for each packet transmission, the values provided in the table have been obtained taking into account all the data packets exchanged in the network cycles.

As can be seen, RSIN shows a better behavior than all other techniques, since it allows to achieve lower values of both mean and standard deviation of the delivery time. Particularly, with respect to SARF, which showed a comparable performance in terms of missed deadlines and lost packets, RSIN is able to provide a significantly lower standard deviation. This represents a meaningful result, since such a metric is closely related to the jitter on packet delivery.

As a final analysis, Fig. 4.14b shows the distribution of the IEEE 802.11n MCSs as selected by the different RA techniques during network operations (the correspondent transmission rates can be inferred from Table 4.6). As can be seen, the Minstrel technique typically settles around MCS 2-4, whereas both FARF and SARF tend to prefer lower MCSs to higher ones. Conversely, RSIN is able to select the lowest as well as the highest MCSs (in agreement with the channel status) confirming in this way its effectiveness.

Simulative assessment The experimental evaluation discussed so far could not deal effectively with the case of more complex networks, for practical reasons, since their implementation would have required a significantly higher number of nodes, as well as larger and different test environments.

As a consequence, the second part of the performance assessment has been carried out through a simulation analysis. Therefore, an effective simulation framework was devised, purposely designed and implemented in the Matlab environment. In the simulator, all the relevant modules of the IEEE 802.11 protocol were implemented, with a special attention to model the behavior of the CSMA/CA procedure, the characteristics and the parameters of the IEEE 802.11n amendment. Also, in the simulation environment the full IEEE 802.11 TGn channel model "F", as depicted by the blue dashed lines in Fig. 4.13b,

has been exploited. Finally, all the RA algorithms considered here were implemented.

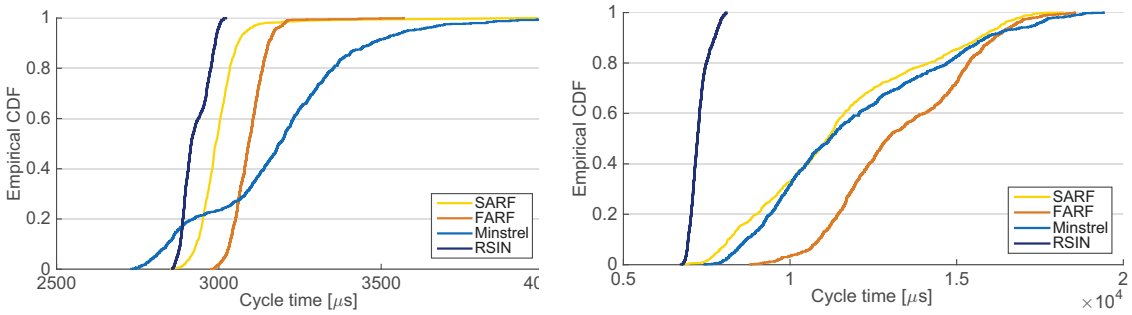
As far as the simulation setup is concerned, the network topology is actually identical to that considered in the experimental validation. Nonetheless, the network can be now composed by n wireless sensors/actuators (specifically, n=10), attached to the AP through IEEE 802.11n links. The distance between any node and the AP is randomly selected at each new simulation.

In this context, among several meaningful performance indicators, the attention was mainly focused on the network cycle time, defined as the time required at the controller to complete the polling procedure on all the attached slaves. Indeed, the communication protocol is the same one described in the previous subsection, and hence the controller sequentially polls all the slaves.

The simulation environment was configured to analyze two representative traffic profiles, the first one characterized by typical industrial small–sized packets (the same 50 bytes payload adopted in the experimental measurements), while the second profile targeted at emulating multimedia real–time traffic, with payloads in the order of hundreds of bytes. Given a set of simulation parameters, 10000 network cycles are performed, and the analysis is repeated ten times with different nodes placement within the environment.

A first set of outcomes is reported in Fig. 4.15a, which shows the ECDF of the cycle time. The deadline imposed for the delivery of each packet, set in the RSIN algorithm, is equal to 500 μs. Comparing the trend obtained for RSIN with the other ones, it is evident that, under the same channel and network conditions, this strategy is able to provide a cycle time considerably lower and more stable. SARF and FARF share a quite similar trend, even if with steadily higher values, and are also characterized by a long tail of cycles needing more than 3.2 ms to complete. Minstrel is sometimes able to reach lower cycle times, but at the expense of an increased jitter, and of a non–negligible percentage of cycles needing more than 3.5 ms to conclude.

In a second set of simulations, the configuration was modified so that the response packets from slaves have a payload length of 500 bytes, to which a deadline of 1.5 ms is associated. The obtained outcomes are presented in Fig. 4.15b, which highlights even more evidently the performance gain obtained by RSIN. Indeed, while the increased payload size clearly led to longer packet transmission times and also to an increased chance of delivery failure, the cycle time variability is again very limited when RSIN is adopted. Conversely, the other three considered RA schemes are clearly unable to provide satisfactory results, since the cycle time values are distributed between 7 ms and 20 ms. This behavior can be explained by the fact that RSIN performs an optimization step based on a more complete set of constraints with respect to the other schemes. In



(a) Response packets of 50 bytes **(b)** Response packets of 500 bytes

Figure 4.15: Simulated ECDF of the cycle time with n=10 nodes.

particular, it relies on a better estimation of the channel status and it is explicitly aware of the application level deadline. Moreover, Fig. 4.15b clearly puts in evidence that the choice of always retransmitting at the lowest available rate, as done by both SARF and FARF, leads to longer cycle times, especially for increased payloads.

A summary of the simulation outcomes is reported in Table 4.8, which allows to draw some further considerations. RSIN, besides providing a lower average cycle time, is also always able to achieve a far lower standard deviation. This is observed both for the case of a payload of 50 bytes, and when the payload is increased to 500 bytes, where the standard deviation is almost an order of magnitude lower with respect to the other techniques. Another aspect that deserves attention is the achievable real–time throughput, that is, the net transfer speed of data bytes in the unit of time, relevant only to real–time data flows. Table 4.8 shows that RSIN provides always a higher real–time throughput than the other RA schemes. However, while in the first scenario this index is inherently limited by the high network overhead compared to the small payloads, in the second scenario RSIN is much more able than the other algorithms to exploit the network resources, and as such delivers a higher quantity of real–time data. It is worth observing that the results of the simulations are not directly comparable with those of the practical experiments. Indeed, the values reported in Table 4.7 are considerably higher than those of Table 4.8 (normalized by the number of network nodes) since the former are necessarily affected by the internal delays of components.

As a final experiment, an assessment of the dynamic rate selection carried out by the different RA techniques was performed. To this regard the occurrence, on the network, of a short deep fade on the SNR value was emulated. Fig. 4.16 reports the behavior of the transmission attempts as they were performed by a specific station on the network. The figure shows, on the x–axis, the attempt number, that is, the progressive number

Table 4.8: Cycle time and Real–Time Throughput for all RA algorithms

	RSIN	SARF	FARF	Minstrel
Metric	**L_{data}=50 bytes**			
Average cycle time [ms]	2.93	3.00	3.09	3.19
Standard deviation [μs]	45	130	77	333
Real–time Throughput [Mbit/s]	2.46	2.41	2.33	2.29
	L_{data}=500 bytes			
Average cycle time [ms]	7.28	11.47	13.34	11.87
Standard deviation [μs]	364	2883	2440	3001
Real–time Throughput [Mbit/s]	5.46	3.68	3.07	3.55

of transmitted frames taking into account also all the retransmissions. In particular, empty circles represent successfully delivered frames, while filled circles represent failed transmissions. The figure considers a small snapshot kept during network operations and highlights, in the solid black line the behavior of the SNR level. This situation, though quite severe to deal with, instantaneously modifies the probability that frames can be delivered successfully, and hence stresses the ability of a RA scheme to promptly address this change. As it can be observed from the lowest figure, RSIN is the only RA algorithm able to track and correspondingly adapt to the channel behavior, without the need of retransmissions. All the other techniques required a (variable) number of retransmissions in order to adapt.

Assessment of RSIN robustness

The behavior of RSIN strongly relies on the assumption that a station willing to transmit a packet is aware of the channel status at the receiver. In order to achieve that, any node that receives a packet first stores the SNR during the reception process, then inserts this value in the next packet it transmits to the sending node. This process, however, may suffer from some intrinsic inaccuracies.

In order to evaluate the SNR value, a station has to extract the RSSI from an incoming packet and then measure the noise floor power. The difference (in dB) between these two values actually returns the SNR value. Unfortunately, this procedure may be affected by errors since the RSSI is measured in a very short time at packet arrival and on a short number of bits (those included in the frame preamble). Both these aspects contribute to introduce inaccuracy in the measurement. Moreover, the noise floor power level is determined by a specific procedure which is typically executed periodically, and often it

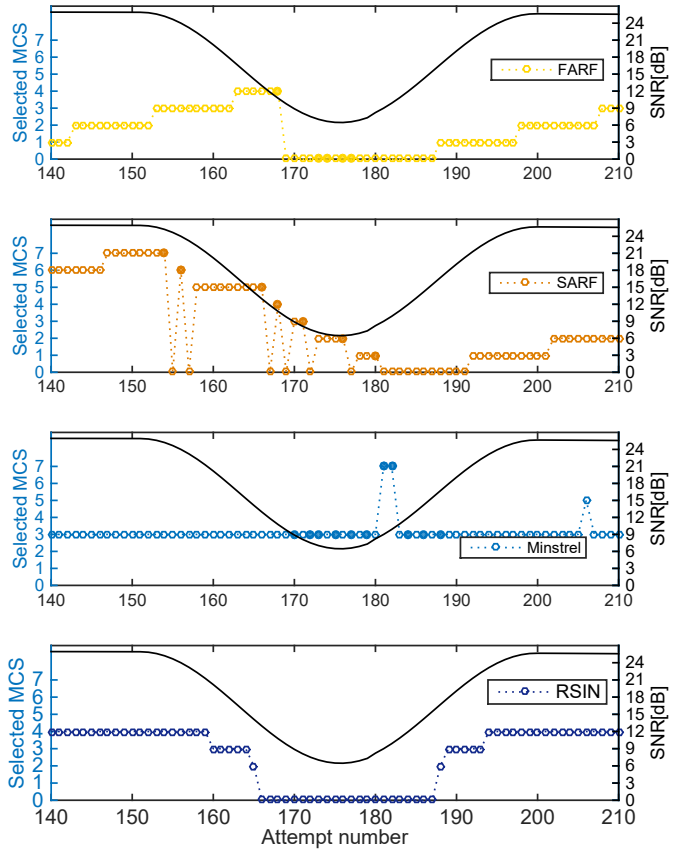


Figure 4.16: Rate selection behavior for all the considered RA algorithms.

is not triggered by the arrival of a packet. Then, it may happen that at the time of the SNR evaluation, the value of the noise floor power is not updated, and hence possibly inaccurate. It may be concluded that the value of the SNR on which RSIN optimization is based might not be always very accurate and, hence, it may negatively influence the behavior of the algorithm.

Another potential source of uncertainty for RSIN is represented by the relationship $\mathcal{F}$ between the PER and the SNR value. For example, the map can be experimentally derived through measurements in a real system under some well–defined conditions, as shown in Fig. 4.3. As an alternative, one can use a theoretical approach to derive the relation between PER and SNR for a given modulation and coding strategy. However, both approaches will only return approximate relations that may not reflect the exact packet loss probability for a given channel status.

To figure out the combined effect of both the aforementioned sources of uncertainty for RSIN, Fig. 4.17 provides an example where it is highlighted that the real working point may actually be located in a different position than the ideal one. Therefore,

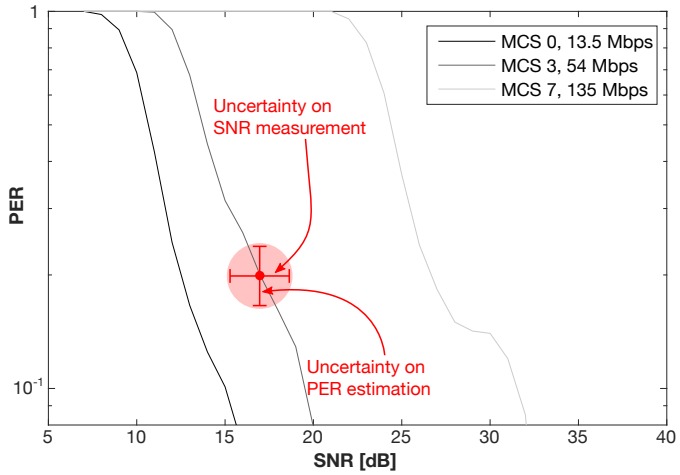


Figure 4.17: Uncertainty on SNR measurements and PER estimation

an assessment of the robustness of RSIN with respect to uncertainties affecting the aforementioned fundamental parameters is needed.

Simulations setup The assessment is carried out via numerical simulations based on the popular ns3 platform (ns3). This open–source simulation tool is widely adopted to emulate different kinds of communication scenarios, and presents several useful features. In particular, it provides a complete implementation of the IEEE 802.11 standard, whose general consistency with real–world IEEE 802.11 performance has been proved through several experimental campaigns (Baldo et al., 2010).

Despite its wide adoption, the default ns3 configuration does not contain any channel model or traffic source specifically conceived for the industrial communication scenario. Consequently, a significant work has been carried out to implement both the needed traffic and channel models, to be integrated in the ns3 platform in order to perform the desired assessment. In order to provide an effective model of small–scale fading in industrial environments, the channel model "F" among those proposed by the IEEE 802.11 TGn was adopted (Erceg et al., 2004). As far as traffic profiles are concerned, this analysis is restricted to a point–to–point configuration, where only two nodes are considered and a polling application is implemented, where the first node (master) periodically sends request packets to the other node (slave), which immediately replies with a response packet. The distance between the two nodes is fixed to 5 meters and the polling period is set to T=2.5 ms, with a total of 10000 polling attempts simulated. The PHY and MAC layers of IEEE 802.11n were configured as recommended in Sec. 4.1. Finally, it is worth to mention that all packets exchanged in the presented simulations have a fixed payload

of 50 bytes, as industrial communications are typically characterized by a small payload size (Willig et al., 2005).

Performance evaluation In the performance evaluation, RSIN has been configured with a deadline on the delivery time of each packet equal to D=1.2 ms, so that the two packets (poll request and response) can be delivered within the polling period T.

In order to assess the robustness of RSIN two types of errors have been artificially added to the simulations. The first error source, referred to as SNR errors, models uncertainty on SNR measurement and consists in the addition of a random noise σ each time the SNR is read at packet reception. It is worth to remember that this value represents the basis on which the optimization procedure of RSIN is carried out, hence an error in the measurements may lead to non–optimal performance. In the simulations, σ was modeled as a uniform random variable lying in the range $[-2 \text{ dB}, 2 \text{ dB}]$. This choice comes from the observation that, typically, both the RSSI reading and the measured noise floor have a granularity of 1 dB. Since the SNR value is obtained as the difference between these two values, the resulting uncertainty lies in the specified range.

The second source of errors, referred to as PER errors, models possible uncertainties that could affect the map $\mathcal{F}$. To this extent, RSIN was tested with a map obtained from the superposition of the measured $\mathcal{F}$ with an error term ρ. Specifically, ρ has been modeled as a uniform random variable in the range $[-0.2, 0.2]$ for each SNR value and each available rate. Clearly, this represents a large range of uncertainty for the PER that accounts for even significant deviations of the map from its expected behavior.

It is worth remarking that both the aforementioned kind of errors may cause the optimization procedure to yield inexact results, hence worsening the performance of RSIN. To get an assessment of the impact of these errors, the percentage of failed pollings was measured, which represents a meaningful performance indicator. Moreover, the statistics of the polling time t_p were collected, to provide further insights. Several different simulations were run, considering separately the cases with only SNR errors, only PER errors and both types of errors simultaneously.

Fig. 4.18 reports the obtained results for the different scenarios, including also the performance of ideal RSIN and Minstrel from the previous simulation as a baseline.

It can be seen that the RSIN algorithm is robust towards uncertainties in SNR errors, in that the percentage of failed pollings does not increase significantly with respect to the ideal case. The introduction of uncertainties on PER estimation, conversely, has an higher impact, causing an increase of almost 10% in failed polling attempts. However, in every case, even when the two errors are considered together, the RSIN algorithm is still much more reliable than Minstrel, which loses a significantly higher number of packets in

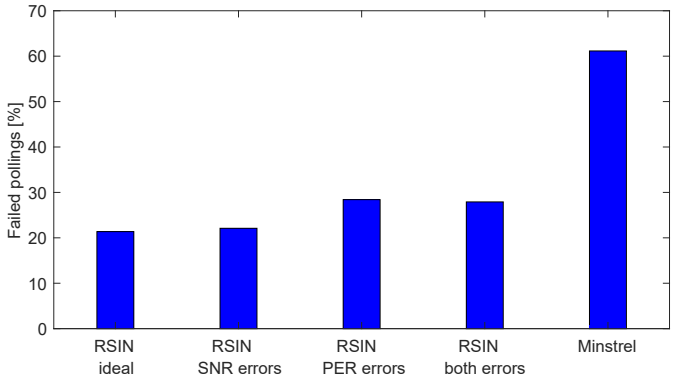


Figure 4.18: Failed polling attempts for different RA schemes and with the presence of different types of errors in RSIN.

Table 4.9: Statistics of polling time for different RA schemes and with the presence of different types of errors

Algorithm	**Mean t_p [us]**	**Std. dev. of t_p [us]**
RSIN - ideal	224.73	40.63
RSIN - SNR errors	223.16	40.90
RSIN - PER errors	211.36	29.38
RSIN - both errors	217.30	51.98
Minstrel	362.60	389.86

the same external conditions.

Tab. 4.9 reports the statistics of the polling time for the different cases, and confirms that the introduction of uncertainties in both SNR and PER only slightly affect the performance of RSIN. Surprisingly, when only PER errors are introduced, the mean and standard deviation of t_p are lower than the ideal case. This is because the random realization of PER errors in the considered simulation leads the optimization procedure to yield slightly higher rates than the optimal ones. Consequently, the algorithm is able to deliver data faster, at the expense of losing more packets when the channel gain is low (as can be argued by the increase of failed pollings shown in Fig. 4.18). This also explains why the mean values of t_p in the presence of errors are lower than in the ideal case. As a final consideration, it can be observed that both the average and standard deviation of the polling time obtained with RSIN are significantly lower than those of Minstrel.

RSIN with estimation of the SNR

The knowledge of the SNR value is clearly of prominent importance for the appropriate operation of RSIN. In the assessments carried out so far, the SNR was actually measured

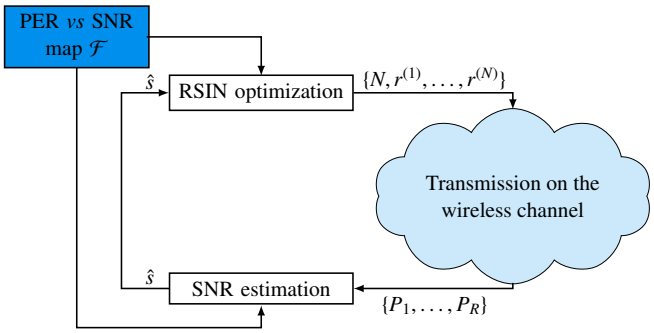


Figure 4.19: Schematic representation of the RSIN–E algorithm.

by each slave upon receiving the poll–request frame and then encapsulated in the payload of the poll–response frame sent back to the master.

However, it has to be noted that this behavior limits the range of applications in which RSIN can be profitably exploited, since in several cases the SNR value might not be available to a station willing to deliver a frame. For example, unidirectional communication applications do not require response frames and, hence, a sender can not rely on this strategy to know the SNR perceived by the receiver. Also, it may happen that the SNR measurement is impaired by unpredictable factors, leading to suboptimal rate selection. To extend the applicability of RSIN to a more widespread range of applications, a new, enhanced, version of RSIN called Enhanced Rate Selection in Industrial Networks (RSIN–E), is introduced. This new algorithm leverages a learning approach to estimate the SNR. In this framework, the previous transmission history is exploited by a station as an input to the learning algorithm to provide an accurate estimation of the channel status.

SNR estimation procedure

A mathematical framework is provided that allows the transmitter to achieve an accurate estimation of the SNR based on the analysis of the previous transmissions. Consequently, the performance of RSIN–E does not rely on the ability of measuring the SNR anymore.

Fig. 4.19 proposes a schematic representation of RSIN–E. In the lower part, it clearly highlights the new SNR estimation block, while the legacy RSIN optimization phase is found in the upper part of the figure. The latter block outputs the transmission rates $r^{(1)}, \ldots, r^{(N)}$ for the subsequent frame delivery, based on the input map $\mathcal{F}$ and the SNR level. Such a map has to be made available in advance.

The estimation block leverages on the fact that the final correct reception (or discarding) of a data frame is acknowledged by an ACK (or by its missed reception).

Consequently, the transmitter gains a precise knowledge of the performance of the previous transmissions and can compute the error probabilities $P_1, \ldots, P_R$ for any of the adopted rates. Combining this information with the PER vs. SNR map $\mathcal{F}$, the needed SNR estimate $\hat{s}$ can be obtained, through the procedure detailed in the following.

Formal description A first and significant assumption is that the given estimation procedure is implemented as a periodic process. That is, a new SNR estimate is provided every *update period*, whose duration is T_u. Consequently, at the end of the k-th update period, based on the observed performance, a transmitting station has to refresh its estimation of the SNR for the subsequent $(k+1)$-th period, namely $\hat{s}\,(k+1)$.

Clearly, the update period must be selected carefully: if it is too short, there will not be enough data to use for the estimation; conversely, if it is too long, the wireless channel status might have changed significantly during the observation period so that the obtained estimation reveals not adequate for the forthcoming transmissions. Nevertheless, it is possible to provide reasonable bounds for T_u. Indeed, since the typical real–time industrial traffic is to a large extent cyclic, even if with different periods, the lowest transmission period T_p represents a lower bound to T_u, which ensures that at least one packet transmission attempt has occurred within the update period. On the other hand, the channel coherence time T_c constitutes an upper bound for T_u, ensuring that the channel remains stationary in that fraction of time (Jung et al., 2011). A general rule of thumb for the choice of the update period is hence

$$T_p \leq T_u \leq T_c \tag{4.9}$$

After the value of T_u has been selected, for each update period k, let $S_i\,(k)$ be the number of observed successful transmission attempts when the rate R_i was employed, and let $A_i\,(k)$ be the total number of transmission attempts performed in k with the same rate R_i, with $S_i\,(k) \leq A_i\,(k)\,, i = 1, \ldots, R$. The observed error probabilities during the k–th period for each different rate can be derived as:

$$P_i\,(k) = \frac{A_i\,(k) - S_i\,(k)}{A_i\,(k)}, \quad i = 1, \ldots, R \tag{4.10}$$

Furthermore, the total number of successes and attempts during the k–th period can also be computed as:

$$S\,(k) = \sum_{i=1}^{R} S_i\,(k)\,, \quad A\,(k) = \sum_{i=1}^{R} A_i\,(k) \tag{4.11}$$

The following weights may also be defined:

$$w_i(k) = \frac{A_i(k)}{A(k)}, \quad i = 1, \ldots, R \tag{4.12}$$

which indicate the frequency with which the different rates were selected during the k–th observation period.

With these premises, the task of the SNR estimation algorithm is to find the SNR value which better explains the observed error probabilities defined in Eq. (4.10), based on the *a priori* knowledge represented by the PER vs. SNR map $\mathcal{F}$.

To solve the aforementioned problem, the following cost function is defined

$$\forall s \in \mathcal{S}, \quad \mathcal{E}(s,k) = \sum_{i=1}^{R} w_i(k) \left[\mathcal{F}(s, R_i) - P_i(k)\right]^2 \tag{4.13}$$

which represents, for each SNR value s, the square difference between the expected error probabilities according to the map $\mathcal{F}$ and the observed error probabilities during the k–th period. For each rate R_i, the weighting factor w_i, defined in Eq. (4.12), is used to give more importance to the rates which have been selected more frequently during the observation period. The SNR estimation for the k–th period is hence defined as

$$\hat{s}(k+1) = \arg\min_{s \in \mathcal{S}} \mathcal{E}(s,k) \tag{4.14}$$

that is, the SNR value which minimizes Eq. (4.13) for a given k.

Although Eq. (4.14) is actually the most obvious solution to the presented SNR estimation problem, the typical behavior of industrial real–time traffic and wireless channel non–idealities demand for further refinements to obtain adequate performance. Indeed, during an update period of length T_u only few channel observations may be available, often relevant to a single frame transmission, i.e. $T_u \simeq T_p$. Clearly, this avoids an effective tracking of wireless channel variations, impairing the accuracy of the SNR estimation, which in turn results not stable over time. For this reason, the SNR estimation procedure has also to take into account the past history of transmission attempts, not limiting to the last observation period.

Consequently, the estimation problem of Eq. (4.14) is modified and set up as a regularized optimization problem (Friedman et al., 2001). Specifically, a penalty function $\mathcal{H}$ is defined, designed in such a way that the estimated SNR values leading to high error probabilities are penalized. To this aim, let $\tilde{A}(s,k)$ be the total number of MAC layer transmission attempts up to the k-th period, where s indicates the (estimated) SNR

value at which those attempts have been carried out. Analogously, let $\tilde{S}(s,k)$ be the corresponding number of successful transmissions. These two functions, differently from the ones included in Eq. (4.10), are not related to the adopted transmission rates but only depend on the values of the estimated SNR adopted in the previous update periods. Therefore, given that in the k-th update period the estimated SNR value is $\hat{s}(k)$, $\tilde{A}(s,k)$ is updated as follows

$$\forall s \in \mathcal{S}, \quad \tilde{A}(s,k) = \begin{cases} \tilde{A}(s,k-1) + A(k) & \text{if } s = \hat{s}(k) \\ \tilde{A}(s,k-1) & \text{otherwise} \end{cases} \tag{4.15}$$

whereas the update of $\tilde{S}(s,k)$ is performed analogously.

Exploiting these quantities, the penalty function for the update period k can be defined as:

$$\mathcal{H}(s,k) = \begin{cases} 0 & \text{if } \tilde{A}(s,k) = 0 \text{ or } \frac{\tilde{A}(s,k)-\tilde{S}(s,k)}{\tilde{A}(s,k)} < P_{th} \\ \frac{\tilde{A}(s,k)-\tilde{S}(s,k)}{\tilde{A}(s,k)} & \text{otherwise} \end{cases} \tag{4.16}$$

In particular, Eq. (4.16) shows that $\mathcal{H}(s,k)$ generally corresponds to the error probability observed in the whole transmission history up to period k, when s was selected as the estimated SNR. However, if either a particular SNR s was never estimated in the past, i.e. $\tilde{A}(s,k) = 0$, or the observed error probability is smaller than a threshold P_{th}, whose significance will be discussed in detail in the next subsection, then no penalization takes place.

Introducing now a penalty coefficient $\lambda \in [0,1]$, it is possible to weigh the contributions of the cost function $\mathcal{E}(s,k)$ and the penalty function $\mathcal{H}(s,k)$. If λ is large, then the penalty function (and hence the history of the network) is weighted more; conversely, if λ is small, the cost function (and hence the results of the last observation) period takes more importance. To this aim, the following new objective function is defined

$$E(s,k) = (1-\lambda)\,\mathcal{E}(s,k) + \lambda \mathcal{H}(s,k) \tag{4.17}$$

Consequently, the final SNR estimate for each update period T_u can be obtained as the solution of the following regularized estimation problem

$$\hat{s}(k+1) = \arg\min_{s \in \mathcal{S}} E(s,k) \tag{4.18}$$

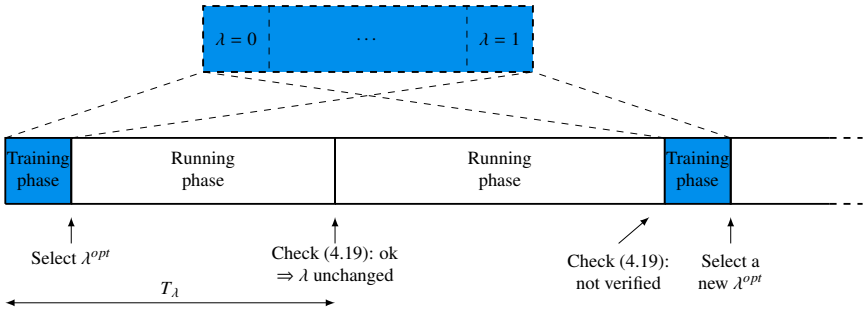


Figure 4.20: Tuning penalty coefficient in RSIN–E: initial training phase and subsequent running phase.

Tuning of the parameters The first meaningful parameter to be tuned is the update period T_u, regulated in general by Eq. (4.9). In industrial applications, the update period can be chosen quite low (i.e., $T_u \simeq T_p$), as a prompt response to channel variations is needed. Moreover, the estimation of the channel coherence time T_c is generally a tricky task, highly dependent on the surrounding environment. Hence, a clear upper bound for T_u might not be available.

The selection of the probability threshold P_{th} also plays an important role in the value of the penalty function defined in Eq. (4.16). P_{th} is introduced to avoid penalizing those SNR values that, when estimated, have led to a relatively low error probability. If P_{th} is high, this would happen for too many SNR values, thus decreasing the weight of the penalty function $\mathcal{H}(s,k)$. A small value for this parameter is hence generally selected, for instance $P_{th} = 0.1$.

Furthermore, the penalty coefficient λ strongly influences the performance of the SNR estimation and, consequently, of the RSIN–E algorithm. The role of λ is to balance the minimization of the cost function in Eq. (4.13) with that of the penalty function in Eq. (4.16). An optimal value for λ cannot be derived a priori and depends on several factors, such as packet size, transmission period, nodes positions, channel impairments, etc. Nonetheless, the typical application scenarios for real–time networks are often characterized by well–defined traffic profiles and mobility patterns. This is the case, for instance, of industrial real–time communication applications, where traffic follows a cyclic pattern and nodes usually have a very limited mobility compared to general purpose wireless networks (e.g., cellular networks).

Consequently, in the context of industrial networks, a strategy for the experimental tuning of λ can be devised, as roughly represented in Fig. 4.20 and, more detailed, in Alg. 1. After the network is deployed, a training phase is initiated, during which the nodes placement and traffic flows reproduce exactly those of the subsequent running phase.

Algorithm 1 Selection of penalty coefficient λ

procedure PENALTYCOEFFSELECTION($\Lambda, A_\lambda, \Delta_\lambda, T_\lambda$)
  $\lambda^{opt} \leftarrow 0$ ▷ Initializations
  $J^{opt} \leftarrow 0$
  $J^{curr} \leftarrow \infty$
  while true do ▷ Network running forever
    $t \leftarrow 0$ ▷ Reset time
    if $\frac{J^{curr} - J^{opt}}{J^{opt}} > \Delta_\lambda$ **then**
      for each $l \in \Lambda$ **do** ▷ Training phase
        $\lambda \leftarrow l$
        Perform A_λ transmission attempts
        $A_l \leftarrow A_\lambda$
        $S_l \leftarrow$ number of successful attempts
        $J(l) \leftarrow \frac{A_l - S_l}{A_l}$
      end for
      $\lambda^{opt} \leftarrow \arg\min_{l \in \Lambda} J(l)$
      $\lambda \leftarrow \lambda^{opt}$
      $J^{opt} \leftarrow J(\lambda^{opt})$
    end if
    Run the network until $t = T_\lambda$ ▷ Running phase
    $A_R \leftarrow$ number of total attempts during the running phase
    $S_R \leftarrow$ number of total successes during the running phase
    $J^{curr} \leftarrow \frac{A_R - S_R}{A_R}$
  end while
end procedure

During this training phase the penalty coefficient is varied among different values picked from a set Λ and, for each value, relevant information on the transmission outcomes are collected. At the end of this phase, a metric of specific interest, called J, is computed based on the collected information. For example, in Alg. 1, the collected information are the number of transmission attempts A_l and successes S_l for each value $l \in \Lambda$ and the metric J is computed as the total percentage of failed transmissions. Finally, the optimal value of the penalty coefficient λ^{opt} is selected as the one that minimizes the metric J and the corresponding value of the metric is stored in J^{opt}. The network then goes on with a running phase, where the value of the penalty coefficient is fixed to λ^{opt}.

Clearly, this value represents the best choice for the current network configuration and wireless channel status, provided that the training phase is long enough to collect a meaningful amount of data. Nevertheless, in the long term, the operating conditions can change (for example due to the sudden disturbance by an external interferer). As a result, an adaptive tuning of λ is hence needed to ensure that the selected value always yields the best performance under the actual operating conditions. To this aim, during the running phase, the nodes keep collecting information on the transmission outcomes and, after a time T_λ, the metric J is computed again and stored in J^{curr}. At this point,

as reported in both Fig. 4.20 and Alg. 1, each node checks if the relative error between the current value of the metric J^{curr} and the previously computed optimal value exceeds a threshold Δ_λ

$$\frac{J^{curr} - J^{opt}}{J^{opt}} > \Delta_\lambda \tag{4.19}$$

If the threshold is exceeded, a new training phase is started, which leads to the definition of new values for λ^{opt} and J^{opt}. Conversely, the network continues the running phase, maintaining the penalty coefficient fixed. Clearly, the sensitivity and delay with which RSIN–E detects and reacts to changes in the operating conditions depend on the choice of the parameters T_λ and Δ_λ.

Computational complexity Since the SNR estimation phase has to be performed online, it is crucial that its computational burden is limited.

To this regard, the following considerations can be made. First, at each period T_u, the update of both $A_i(k)$ and $S_i(k)$ requires only R operations each. The computation of probabilities $P_i(k)$ and weighting factors $w_i(k)$ also requires R operations each. In addition, the update of $\tilde{A}(s,k)$, $\tilde{S}(s,k)$ and the penalty function $\mathcal{H}(s,k)$ requires Σ operations each, one for each SNR value in $\mathcal{S}$. Finally, the computation of the cost function $\mathcal{E}(s,k)$ requires $R \cdot \Sigma$ total operations. Therefore, the complexity of the proposed estimation algorithm is $O(R \cdot \Sigma)$, being $R(\Sigma + 4) + 3\Sigma$ the total number of operations that have to be carried out at each update period. For instance, in a system where $R = 8$ transmission rates are available, and considering 30 dB as the range of variability for the SNR values (with a granularity of 1 dB), then a total of 362 elementary operations are required.

Considering the whole procedure involved in RSIN–E with reference to Fig. 4.19, its complexity is actually dominated by both the discussed estimation algorithm and the optimization step carried out by the legacy RSIN. As already pointed out, the computational burden associated to RSIN may actually result demanding, but the algorithm can be reduced, under some specific hypotheses, to a simple search within a look–up table. Indeed, if the frame length L is constant (as it often happens in real–time industrial networks) and the measured PER vs. SNR map $\mathcal{F}$ does not change significantly over time, the optimization phase of RSIN that leads to the generation of the sequence of rates can be carried out offline for each value of the SNR $s \in \mathcal{S}$ and stored in the memory of the station. Then, at run time, after the SNR value has been estimated, it is sufficient to search for the corresponding solution. Stemming from the above considerations, it may be concluded that RSIN–E can be effectively adopted even by simple industrial devices such as sensors/actuators.

Performance assessment in a prototype network

The performance figures of RSIN–E have been experimentally evaluated in different scenarios and compared with those of both the legacy RSIN and Minstrel.

Experimental setup The RSIN–E algorithm has been implemented in a prototype network based on commercial off–the–shelf WNICs. The implementation is based on the IEEE 802.11 stack provided by the Linux kernel and, specifically, on the mac80211 module, as it was represented in Fig. 4.12. In detail, the optimal rate selection of RSIN–E is executed before any packet transmission procedure, in order to provide the WNIC driver with the optimal rates and number of attempts, whereas the SNR estimation is performed periodically with period T_u.

The modified mac80211 modules have been loaded in a set of desktop workstations (Dell Optiplex models 745, 755 and 960) with Ubuntu 14.10 operating system based on Linux kernel version 3.16.4. Each workstation is equipped with a TP-LINK WNIC (models TL-WN851ND and TL-WN881ND), based on the AR9287 chip, which is handled by the open–source ath9k driver. Both WNICs are compliant with the IEEE 802.11n standard and have been configured as recommended in Sec. 4.1.

The preliminary experimental campaign allowed to observe that the meaningful range of SNR values goes from 7 to 36 dB. Indeed, for SNRs lower than 7 dB a PER of 1 is observed for any MCS, while an SNR of 36 dB is high enough to guarantee successful transmissions for all MCSs. Consequently, the set $\mathcal{S}$ used for the SNR estimation contained $\Sigma = 30$ SNR values, ranging from 7 to 36 dB with 1 dB spacing. This quantization choice allowed a fair comparison with the legacy RSIN algorithm based on measured SNR, since both the RSSI reading and the noise floor in the adopted WNICs have a granularity of 1 dB.

The desktop workstations have been arranged in a prototype IEEE 802.11n network composed of three nodes, as shown in Fig. 4.21. The network is configured in infrastructure mode and emulates a typical industrial application: a central controller, which acts as AP, periodically sends request packets to distributed sensors/actuators, represented by WLAN STAs, that send response packets when polled. This exchange of packets is realized through a purposely developed software application, installed in all nodes. The period with which the controller sends a request packet to each one of the nodes is fixed to $T_p = 25$ ms. To mimic typical real–time industrial applications, packets are exchanged at the data–link layer, avoiding network and transport protocols and hence, besides the MAC header and trailer, they only contain an application layer payload of length L. Two example values have been considered for the payload length: $L = 50$ bytes,

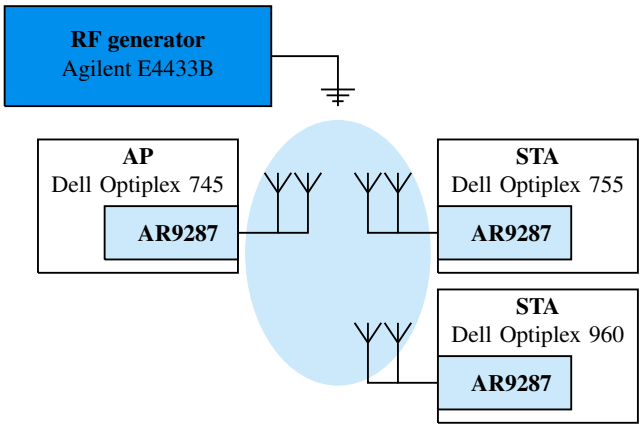


Figure 4.21: Prototype network used for the experimental validation.

which corresponds to a traditional industrial application involved with the exchange of very short commands and sensor readings, and $L = 500$ bytes, which may be instead representative of more advanced industrial multimedia applications (Silvestre-Blanes et al., 2015).

The prototype IEEE 802.11n network has been tested in a research laboratory, where a complete electromagnetic isolation was not feasible. However, the carrier frequency used by the WNICs has been carefully selected after monitoring the environment with a real–time spectrum analyzer, so that to avoid channels where other wireless networks were operating. Moreover, particular attention was dedicated to emulating typical wireless channel impairments found in industrial environments. To this aim, the channel models proposed by the IEEE 802.11 TGn (Erceg et al., 2004) have been taken into consideration and, specifically, model "F" was chosen. In order to reproduce this channel behavior in the prototype network, an Agilent E4433B RF signal generator has been introduced in the setup, and configured to generate AWGN–like noise centered on the operating frequency of the WNICs. The output power was varied to reproduce a quantized version of the channel gain behavior found in model "F". The controlled noise produced by the RF generator has been injected on the channel through a directional antenna pointed towards all the employed WNICs, as depicted in Fig. 4.21. In the quantization of the channel model, that resembles the one depicted in Fig. 4.13b, three levels have been used: a low one, corresponding to an SNR level at which all available MCSs fail except the lowest one (which is however impaired); an intermediate one, which impairs only the highest MCSs; a high one, which corresponds to a high SNR level at which all MCSs can be used without producing any transmission error.

The tuning of both the probability threshold for the SNR estimation and the update

Description	Value
PHY/MAC standard	IEEE 802.11n
MIMO configuration	2×2 STBC
Channel bandwidth	40 MHz @ 2.4 GHz
MCSs	0-7
Transmission rates	13.5, 27, 40.5, 54, 81, 108, 121.5, 135 Mbit/s
Channel model	IEEE 802.11TGn model "F"
Payload size (L)	$\{50, 500\}$ bytes
Polling period (T_p)	25 ms
Packet deadline (D)	$\{0.5, 1, 2, 5\}$ ms
Polling attempts during the running phase	10000
Max. number of MAC–layer attempts (N_{max})	10
RSIN–E SNR range for estimation ($\mathcal{S}$)	$\{7, 8, \ldots, 36\}$ dB
RSIN–E update period (T_u)	25 ms
RSIN–E probability threshold (P_{th})	0.1
RSIN–E set of penalty coefficients (Λ)	$\{0, 0.3, 0.6, 0.9\}$
RSIN–E training phase attempts (A_λ)	1000
RSIN–E penalty coefficient period (T_λ)	250 s
RSIN–E penalty coefficient threshold (Δ_λ)	0.1
RSIN–E optimal penalty coefficient (λ^{opt})	0.6 ($L = 50$ bytes), 0.3 ($L = 500$ bytes)

Table 4.10: Parameters of the experimental setup

period has been carried out in agreement with the considerations made so far, thus selecting $P_{th} = 0.1$ and $T_u = 25$ ms, respectively. Hence, the estimated SNR is updated at each polling cycle and the cost function in Eq. (4.13) is computed by taking into account only the outcome of the very last packet transmission. The choice of the penalty coefficient λ, has been carried out emulating the network training phase and the metric to minimize, J, is represented by the percentage of failed pollings, as in Alg. 1. Specifically, the set of candidate values for λ contained four values, $\Lambda = \{0, 0.3, 0.6, 0.9\}$ and, during the training phase the outcomes of $A_\lambda = 1000$ polling cycles have been recorded for each value in Λ. The update parameters were set to $T_\lambda = 250\ s$ and $\Delta_\lambda = 0.1$. It is worth observing that, since the laboratory environment is static and controlled, the condition in Eq. (4.19) was always observed and, hence, only one initial training phase was needed in the experiments. Specifically, the outcomes of this training phase were $\lambda^{opt} = 0.6$ for the case of short packets ($L = 50$ bytes) and $\lambda^{opt} = 0.3$ for longer ones ($L = 500$ bytes). This difference confirms the observation that the optimal value of λ depends on the traffic features, e.g. the payload size, and hence has to be carefully selected through a preliminary training phase.

The meaning and value of all the parameters adopted in the experimental evaluation are reported in Tab. 4.10.

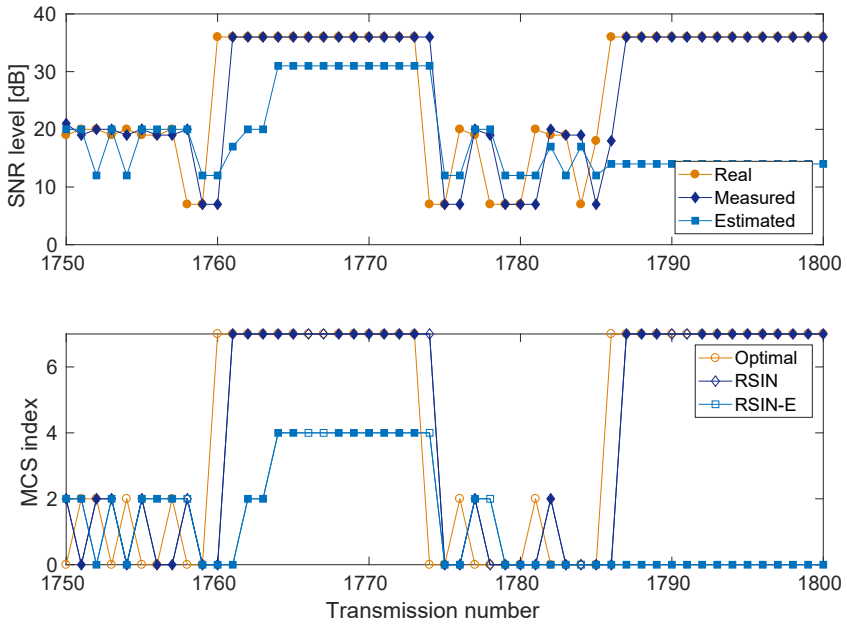


Figure 4.22: Qualitative assessment of the SNR estimation performance for $L = 50$ bytes and $D = 0.5$ ms: real, measured and estimated SNR (top) and corresponding MCS chosen for the first transmission attempt (bottom). In the MCS plot, for RSIN and RSIN–E a filled marker indicates a successful transmission, while an empty marker indicates a failed one.

Qualitative assessment of the estimation performance A first assessment of the RSIN–E performance is shown in Fig. 4.22 in a qualitative way. The figure reports the estimated SNR and the corresponding chosen rates by the AP for 50 consecutive polling cycles during the network running phase. In this assessment, short packets of $L = 50$ bytes have been exchanged (hence λ has been set to 0.6) and the deadline has been set to $D = 0.5$ ms.

In detail, the top figure shows three different SNR patterns: the orange line with circular markers reports the real SNR value at the receiving STA, the dark blue line with diamond markers indicates the measured SNR sent to the AP (basing on which the legacy RSIN algorithm selects the optimal rates) and the light blue line with square markers reports the SNR estimated by RSIN–E. First, it is noticeable that the real SNR levels are approximately distributed around three main values, corresponding to the three channel quantization levels: 7 dB (bad channel), 20 dB (average channel) and 36 dB (good channel). Moreover, it is evident that the measured SNR follows exactly the real value with a delay of 1 transmission. Indeed, the measured SNR is retrieved from the response packet and, hence, corresponds to the value measured by the receiver during

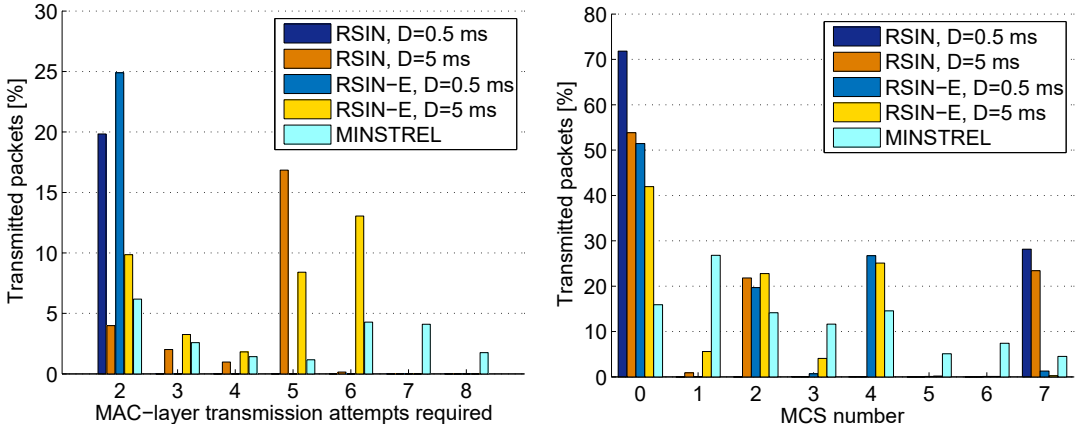
the previous packet transmission. As far as the SNR estimation is concerned, it may be observed that, although there is a certain difference between measured and estimated values, the general trends of the two curves are the same: when the measured SNR goes up (due to a high channel gain), also the estimated one increases, and the same happens when the SNR suddenly drops.

The lower part of the figure shows the MCS selected by both versions of RSIN for the first packet transmission attempt, based on the SNR value. Specifically, the orange line reports the MCS choice for the real SNR, the dark blue line reports the MCS chosen by the legacy RSIN algorithm, and the light blue line reports the MCS chosen by RSIN–E. For RSIN and RSIN–E filled markers indicate a successful transmission, whereas empty markers indicate a failed one. Again, it can be observed that the trend of the RSIN–E algorithm follows that of the legacy RSIN. Moreover, it can be noticed that the former is generally more conservative, in the sense that most of times it chooses a rather lower MCS than the one chosen by RSIN. In particular, looking at the transmission attempts after #1785, it can be observed that RSIN–E evidently estimates a low SNR (and hence selects MCS 0), even if the real SNR is high. This is due to the fact that some previous transmission attempts (e.g., #1766, #1767 and #1774) failed despite the estimated SNR was good. This specific aspect of RSIN–E driven by the presence of the penalty function in Eq. (4.16), may increase the packet transmission times even if the channel state is good but, on the other hand, ensures higher success probabilities when the channel state is average or bad, which may reveal a good strategy especially if a high reliability is mandatory.

System performance with short packets In this section the performance figures of RSIN–E are compared with those of the legacy RSIN as well as with those of the widespread Minstrel algorithm tuned for real–time communications (Minstrel).

A first insight is provided by Fig. 4.23a, which shows the histogram of the MAC–layer transmission attempts[4] required for the delivery of a packet for the three different algorithms and two values of the RSIN deadline, $D = 0.5$ ms and $D = 5$ ms. Ideally, a good rate adaptation algorithm for real–time communication must perform the lowest possible number of transmission attempts, since each attempt adds significant delay and jitter due to the retransmission and backoff mechanisms of IEEE 802.11. It can be first observed that setting a lower deadline results, intuitively, in a lower number of transmission attempts: for both RSIN algorithms, when $D = 0.5$ ms no more than 2 attempts are ever required. Moreover, the distribution of transmission attempts for

[4]The transmissions that required only one attempt are not reported here for the sake of clarity. The figure, hence, only reports the cases when MAC–layer retransmissions have been necessary.



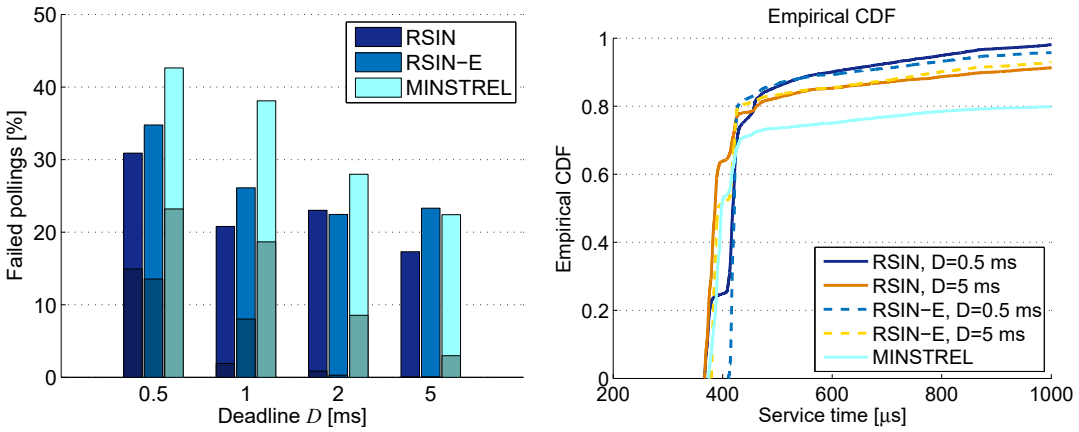
(a) Histogram of TX attempts **(b)** Selected MCSs

Figure 4.23: Insights on the performance of different RA algorithms for different deadline values ($L = 50$ bytes).

RSIN and RSIN–E is quite similar for a given deadline value, with only small percentage differences. Finally, when the Minstrel algorithm is employed, a non–negligible percentage of transmissions require a high number of attempts (6 or more), as this algorithm keeps on retransmitting a packet until it is received or the maximum number of attempts is reached.

Fig. 4.23b shows the histogram of the MCSs that have been selected by the rate adaptation algorithm considering all transmission attempts at the MAC layer and, again, for $D = 0.5$ ms and $D = 5$ ms. Several observations can be drawn. First, when the deadline is short, the RSIN algorithm only selects two values: the lowest one (MCS 0), when the channel is bad, and the highest one (MCS 7), when the channel is good. The behavior of RSIN–E, instead, is different: MCS 0 is also the most frequently selected MCS, whereas MCS 7 is almost never selected, with MCSs 2 and 4 being adopted instead. This is a reflection of the conservative behavior already spotted in Fig. 4.22. When the deadline is higher, the outcomes are slightly different but follow the same trend: RSIN selects either MCS 0, 2 or 7, whereas RSIN–E chooses mostly the first five MCSs. Finally, the Minstrel algorithm selects most frequently MCS 1 (which yields the best "average" performance), even if all the available MCSs are adopted a non–negligible amount of times, due to the sampling behavior of this technique (Minstrel).

The most important performance indicator for industrial RA algorithms is arguably the percentage of failed pollings, reported in Fig. 4.24a for different deadline values, from $D = 0.5$ ms to $D = 5$ ms. The figure allows to distinguish between pollings failed because the packet (either request or response) was lost (lighter part of the bar) or



(a) Failed pollings vs. deadline **(b)** Service time ECDF

Figure 4.24: Performance of different RA algorithms for different deadline values (L = 50 bytes).

Algorithm	Mean	Std. deviation
RSIN $D = 0.5$ ms	528.9 μs	416.8 μs
RSIN–E $D = 0.5$ ms	529.6 μs	309.1 μs
RSIN $D = 5$ ms	566.6 μs	443.9 μs
RSIN–E $D = 5$ ms	558.3 μs	429.1 μs
Minstrel	1061.1 μs	1770.4 μs

Table 4.11: Service time statistics with $L = 50$ bytes long request packets, for different RA algorithms and deadline values.

because it arrived after the corresponding deadline (darker part of the bar). As a general trend, it can be observed that, for all algorithms, the higher the deadline, the better the performance, as it is intuitive. Moreover, in almost all cases, the performance figures of RSIN–E are slightly worse than those of RSIN but significantly better than those of Minstrel, the only exception being when the deadline is very high ($D = 5$ ms), since in this case Minstrel has enough time to perform many retransmissions. Specifically, looking at the percentage of packets arrived after the deadline, it emerges that Minstrel violates significantly more deadlines than the RSIN and RSIN–E algorithms. This is linked to the fact that the latter are aware of the deadline and consider it in the rate selection, whereas the former does not.

Another important performance indicator is provided by Tab. 4.11, which shows the statistics of the service time (mean and standard deviation) for request packets (i.e., packets sent by the AP to the STAs), considering all rate adaptation algorithms and deadlines of $D = 0.5$ ms and $D = 5$ ms. As a first important observation, it is evident that the performance of Minstrel are extremely poor, both in mean and standard deviation,

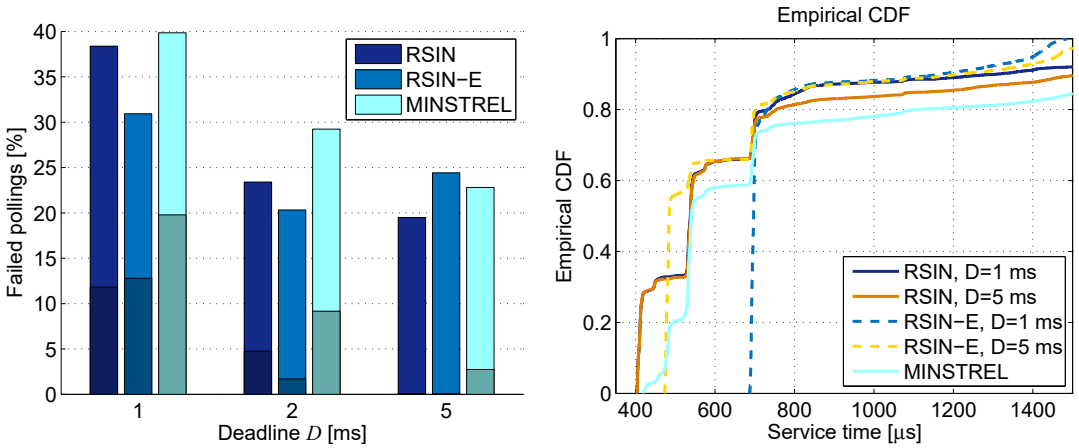
stemming from the fact that it is not designed for real–time communications. The high standard deviation, in particular, makes it very badly suited for real–time industrial applications. The performance of RSIN and RSIN–E, instead, are very similar in terms of mean and standard deviation. Furthermore, as can be seen, the deviation of the service time with RSIN–E results always slightly lower than that obtained with RSIN, due to the conservativeness of the former procedure as previously observed.

The results of Tab. 4.11 are confirmed by Fig. 4.24b, which shows the ECDF of the service time for the different rate adaptation algorithms. The figure, however, allows obtaining some further important insights. Indeed, as can be seen, the ECDFs of RSIN and RSIN–E when the deadline is high ($D = 5$ ms) are practically overlapping and significantly outperform Minstrel in terms of packets delivered within 1 ms (more than 90%). Also, looking at the ECDFs when the deadline is short ($D = 0.5$ ms), it can be observed that RSIN delivers a notable percentage of packets (more than 20%) in less than 400 μs, whereas RSIN–E never does it, since, due to its conservativeness, it almost never adopts the highest MCS. On the other hand, the two versions of RSIN are able to deliver a very high percentage of packets (more than 95%) within 1 ms.

System performance with long packets The growing complexity of NCSs calls for different kinds of applications, where bigger amount of data must be exchanged with real–time constraints. For example, real–time industrial multimedia applications are concerned with the exchange of images and/or video frames that allow performing video–surveillance or real–time tracking of objects in an industrial setup (Silvestre-Blanes et al., 2015). In these cases, the typical payload length is much higher than that of the classical applications and may reach several hundreds of bytes. Thus, in a further experimental session, similar tests have been performed with a payload length of $L = 500$ bytes for both request and response packets. In this case, the penalty coefficient of the SNR estimation algorithm has been set accordingly to $\lambda = 0.3$. The minimum deadline employed for the RSIN optimization has also been increased from 0.5 to 1 ms.

The percentage of failed pollings for the different RA algorithms is presented in Fig. 4.25a. The biggest difference with the case of shorter packets in Fig. 4.24a is that the RSIN–E algorithm now outperforms the version of RSIN that relies on measured SNR for the shortest deadline values. Indeed, the higher conservativeness of RSIN–E is more effective when packets are longer, since the impact of retransmissions may lead to considerably longer transmission times, and likely results in missing the deadline,especially if such deadline is short. Moreover, the performance gap between RSIN and Minstrel is lower in this case, since Minstrel works better when the packet size is high.

The statistics of the service time are reported in Tab. 4.12. As can be seen, both RSIN



(a) Failed pollings vs. deadline

(b) Service time ECDF

Figure 4.25: Performance of different RA algorithms for different deadline values (L = 500 bytes).

Algorithm	Mean	Std. deviation
RSIN $D = 1$ ms	737.8 μs	551.7 μs
RSIN–E $D = 1$ ms	844.0 μs	465.4 μs
RSIN $D = 5$ ms	804.7 μs	694.7 μs
RSIN–E $D = 5$ ms	736.2 μs	558.9 μs
Minstrel	1191.7 μs	1777.1 μs

Table 4.12: Service time statistics with $L = 500$ bytes long request packets, for different RA algorithms and deadline values.

and RSIN–E have similar performance and are able to deliver a substantial reduction in mean and standard deviation of the service time with respect to Minstrel.

A final insight is given by Fig. 4.25b, which reports the ECDF of the service time for request packets and presents a quite different situation with respect to Fig. 4.24b. The Minstrel algorithm is again capable of providing very low service time values (around 500 μs), but a significant percentage of packets (more than 10%) takes more than 1.5 ms to be delivered. The legacy RSIN algorithm is also able to reach very low service time values (around 400 μs) with a good probability, whereas for RSIN–E the minimum values of the service time are around 500 μs (with $D = 5$ ms) and 700 μs (with $D = 1$ ms). However, RSIN–E guarantees an upper bound of the service time, since almost all packets are delivered within 1.5 ms (for both deadline values), contrarily to the performance of the legacy RSIN where the percentage of packets that takes more time is around 10%.

Simulations in a larger industrial network

In typical real–world applications, more complex networks than the one represented in Fig. 4.21 may be deployed, where several stations are distributed in the environment and all communicating with a central controller. Unfortunately, reproducing such a kind of setups with desktop PCs in a research laboratory is challenging. Thus, a simulative assessment is presented, aimed at evaluating the scalability of RSIN–E in larger networks.

The performance assessment has been carried out with the ns3 network simulator ns3. To enhance the dependability of the simulations even further, the IEEE 802.11 PHY layer and the wireless channel of the legacy ns3 implementation have been modified according to the results obtained during the experimental campaigns. To this regard, as a first relevant upgrade, the experimentally measured PER vs. SNR curves have been inserted to determine whether a packet is successfully received or discarded at the PHY layer, basing on the SNR, the packet size and the transmission rate. Moreover, a quantized realization of TGn channel model "F" has been inserted in the simulations, mimicking the artificial noise introduced in the experimental evaluation through the RF generator.

The ns3 platform upgraded with the aforementioned features has been used to simulate an IEEE 802.11n infrastructure network composed of one central controller (that acts as AP) and $M = 10$ attached nodes. A cyclic communication schedule is established, where the controller sequentially polls each node sending a request packet and receiving a response packet, both of size L bytes. The cycle period is set to $T_{cycle} = 50$ ms, with a slot assigned for the polling of each node set to 5 ms. A polling is considered as failed if the response packet does not arrive within the assigned slot. The nodes are randomly deployed on a circular area centered on the controller, with a minimum distance of 1 meter and a maximum one of 3.5 meters. It is worth highlighting that the results presented in this section have been averaged over 10 different random dispositions, to avoid dependence on a particular disposition of nodes.

The simulative assessment has focused only on the comparison between the two versions of the RSIN algorithm, which have been both implemented in ns3. Different deadlines for the service time have been considered, with the maximum one limited to $D = 2$ ms, to allow for an exchange of two packets within the 5 ms slot. For what concerns the SNR estimation parameters, the update period T_u has been set equal to the period with which a single node is polled, i.e., the cycle time $T_{cycle} = 50$ ms, whereas the other parameters have been kept to the values used during the experimental evaluation, i.e., $\lambda^{opt} = 0.6$ and $P_{th} = 0.1$.

A first set of results, concerned with the exchange of $L = 50$ byte long packets, is reported in Tab. 4.13, which shows the cumulative percentage of failed pollings. As

Deadline	Failed pollings [%]	
	RSIN	RSIN–E
$D = 0.5$ ms	19.32	13.65
$D = 1$ ms	17.79	12.41
$D = 2$ ms	15.43	7.64

Table 4.13: Percentage of failed pollings for the two versions of the RSIN algorithm with different deadlines in a simulated industrial infrastructure network of $M = 10$ nodes ($L = 50$ bytes).

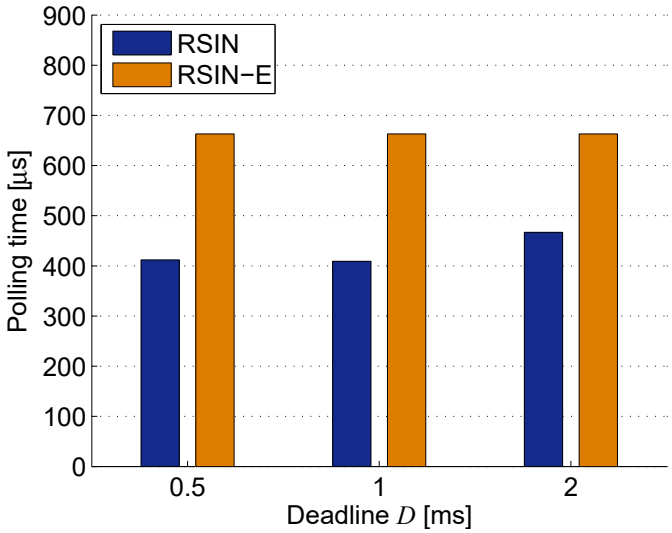


Figure 4.26: Average polling time for the two versions of the RSIN algorithm with different deadlines in a simulated industrial infrastructure network of $M = 10$ nodes ($L = 500$ bytes).

can be seen, in the simulative assessment, RSIN–E always outperforms the legacy RSIN strategy. This result contrasts with what observed during the experimental evaluation and is mostly due to the fact that the conservativeness of RSIN–E allows a higher degree of robustness in an extremely controlled environment, such as the one simulated with ns3. In practice, RSIN–E tends to estimate a low SNR and hence selects the lowest rates (which correspond to the most robust modulation and coding schemes) more often than the legacy RSIN, as already noted in the experimental evaluation (see Fig. 4.23b).

While the conservativeness of RSIN–E reveals effective in terms of reliability, it may reveal detrimental for the timeliness of the polling procedure.Indeed, the legacy RSIN is able to immediately recognize good channel conditions, so that it reacts by selecting the fastest transmission rates, while the estimation–based approach is more conservative and almost always prefers the slowest ones. To provide an example, Fig. 4.26 reports the average polling time of the two algorithms for different values of the deadline D for the

case of $L = 500$ bytes packets. It can be observed that the legacy RSIN strategy is in general able to achieve better performance than RSIN–E. This is particularly evident when the deadline is low, that results in almost a half polling time with respect to RSIN–E.

Conclusions

The MRS feature of the IEEE 802.11 standard allows to adapt the transmission rate to the channel conditions, although no standard algorithm is defined in the standard, leaving the implementation to the users. Default RA algorithms, such as ARF and Minstrel, have been proven to be a bad choice for real–time industrial applications, although their performance can be improved with the opportune tuning of some parameters. Other rate control strategies specifically designed for industrial applications, such as SARF and FARF, show better performance but still not completely satisfactory.

An original algorithm for RA in industrial WLANs, called RSIN, have been presented and tested on commercial IEEE 802.11 devices. This algorithm is based on a constrained optimization procedure, that selects both the number of transmission attempts and the corresponding rates for each frame with the goal of minimizing the residual error probability while ensuring that the delivery time does not exceed a predefined deadline. The optimization requires an explicit feedback of the SNR measured at the receiver. A comprehensive experimental assessment showed that RSIN outperforms all the other considered RA schemes, for different packet sizes. This trend is confirmed also by numerical simulations on larger network setups.

The robustness of RSIN has been tested by considering the performance degradation in presence of errors in the SNR readings or in the PER values. Moreover, the applicability of this algorithm has been extended to the cases where an explicit feedback of the SNR is not available by applying an SNR estimation procedure. The performance of this new version, called RSIN–E are slightly worse than the original version based on explicit feedback but still considerably better than other RA algorithms.

5

Full duplex Wireless Networks

One of the key limitations suffered by wireless networks in comparison with their wired counterparts is the so–called "half–duplex constraint": a wireless terminal cannot transmit and receive simultaneously in the same frequency band. Indeed, due to the much shorter distance, the power of the signal transmitted by a terminal at its own receiver is much higher than that of any other received signal, hence preventing a successful reception whenever a transmission is taking place. This phenomenon is called SI.

In recent years, several technique to reduce SI down to the noise floor level, effectively allowing simultaneous transmission and reception, have been proposed and experimentally validated. The possibility of FD wireless communications has opened a wide range of opportunities as well as challenges, including the development of new MAC layer algorithms able to efficiently handle this new feature. In this chapter, after a brief overview on the concept of FD wireless, an original MAC protocol for FD wireless networks is presented, and some potential applications in the industrial communication framework are discussed.

This chapter is mainly based on the works in Luvisotto et al. (2016a) and Luvisotto et al. (2016b).

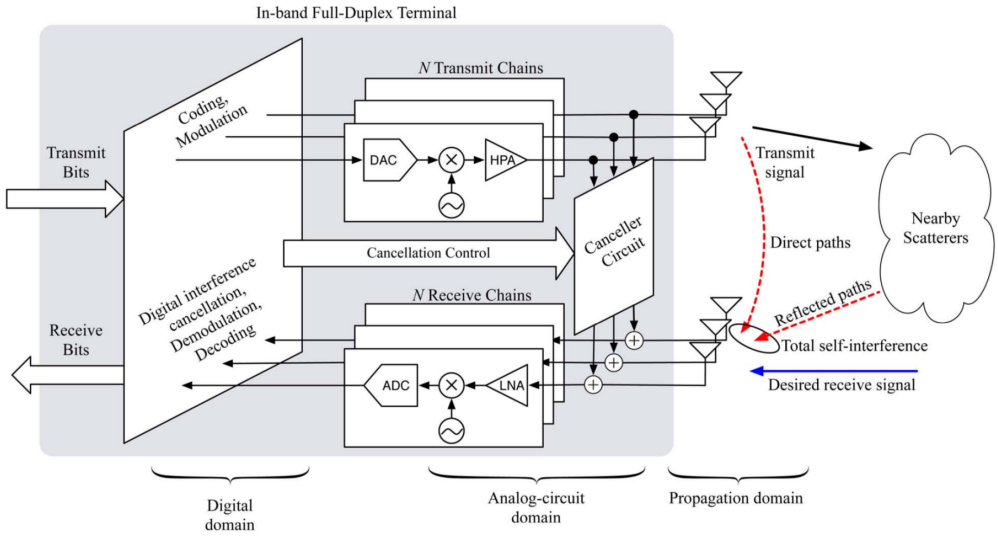


Figure 5.1: Anatomy of a wireless terminal highlighting different SIC domains, taken from Sabharwal et al. (2014).

5.1 Fundamentals of full duplex wireless

The possibility for a wireless terminal to transmit and receive simultaneously in the same band allows immediately to double its spectral efficiency, measured in terms of information bits reliably transmitted per second per Hz (Sabharwal et al., 2014). Furthermore, advantages can extend beyond the PHY layer, benefiting the performance at the access and network level. For example, the possibility to receive frames while transmitting can enable immediate collision detection and instantaneous feedback in contention–based networks (Choi et al., 2012).

In the following, the most commonly adopted techniques to cancel the SI signal are detailed. Subsequently, some proposals that exploit FD capabilities to increase the performance of the network, especially in terms of access to the channel, are presented.

Techniques for self–interference cancellation

Removing the effect of the SI signal due to an ongoing transmission from the receive path is not an easy task. To provide a numerical example, suppose that the transmit power of a terminal is $P_{TX} = 20$ dBm and that the noise floor (i.e., the sum of the power of all noise sources) is $P_N = -90$ dBm. The SI power is roughly equal to the transmit power. Through various SIC techniques, this power can be reduced to a Residual Self–Interference (RSI) power, $P_{RSI} \approx P_{TX} - P_{SIC}$. The goal is to reduce P_{RSI} to a level comparable to that of noise floor, so that a correct decoding of received signals is

allowed. In the example considered, this means that the SIC techniques must provide 110 dB of cancellation. This ambitious goal can be reached through the combination of various approaches, as detailed in the following and depicted in Fig. 5.1.

Propagation domain SIC techniques The first approach considered is to electromagnetically isolate the transmit and receive paths, thus suppressing the SI before it actually impacts on the reception process. While a small isolation factor of some dBs is always present, this value can be significantly increased by exploiting proper antenna spacing (Duarte and Sabharwal, 2010), shielding structures (Everett, 2012), cross–polarization (Everett et al., 2014) and directional antennas (Everett et al., 2011), among other techniques.

Propagation domain techniques can offer good performance on direct SI, i.e., the one caused directly by the transmitted signal, achieving more than 70 dB isolation (Everett et al., 2014). However, these approaches suffer from reflected SI, i.e., the one caused by the transmitted signal that, after reflecting off nearby scatterers, hits the receive path with a significantly high residual power. To handle this kind of SI, channel–aware strategies should be developed, such as transmit beamforming (Senaratne and Tellambura, 2011), which have the drawback of requiring preliminary channel estimation procedures.

Analog–circuit domain SIC techniques These techniques aim at suppressing SI in the analog circuitry of the receive chain, before the ADC. This can be achieved by directly tapping the analog transmit signal before the antenna and subtracting it from the received one (the "canceller circuit" in Fig. 5.1), or by applying the necessary gain/phase/delay adjustments to the digital transmitted signals and then converting it in analog form before subtracting it (the "cancellation control" in Fig. 5.1).

Similarly to propagation domain techniques, analog domain approaches can be channel–unaware or "passive", hence suppressing only direct SI (Duarte et al., 2012), or they can be channel–aware or "active", aiming at dealing also with reflected SI (Jain et al., 2011).

Digital domain SIC techniques The last possible approach to achieve SIC is to suppress the SI signal after the ADC, applying digital processing techniques to filter out the interfering signal from the received one. These techniques are based on discrete–time models of the transmit/receive paths, which are then used to develop sophisticated algorithms to cancel the effect of the transmitted signal (Day et al., 2012).

The advantage of working in the digital domain is that complicated processing becomes simpler. On the other hand, there is an intrinsic limitation to the amount of SIC that can be performed through this approach, linked to the dynamic range of the

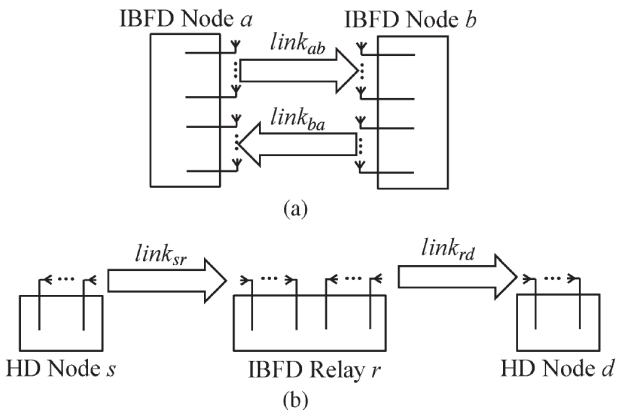


Figure 5.2: Possible configurations of a FD link: bidirectional (a) or relay (b). Figure taken from Kim et al. (2015).

ADC (Sabharwal et al., 2014). Indeed, suppose that a terminal uses a 14–bit ADC whose effective number of bits is $ENOB = 11$. In this case, its effective dynamic range is approximately $6.02\,(ENOB - 2) \approx 54$ dB. This means that, even if the digital domain SIC techniques are able to work perfectly, residual errors due to quantization and noise will remain 54 dB below the level of the initial SI. Considering the example at the beginning of this section, this means that, in order to reach the noise floor, propagation and analog domain SIC techniques must provide at least 56 dB of cancellation.

From this analysis it is evident how an effective FD system must rely on a combination of propagation, analog and digital domain SIC techniques. For example, the design in Bharadia et al. (2013) offers 62 dB of propagation/analog cancellation and 48 dB of digital cancellation, allowing to effectively reduce the SI to the noise floor level for 80 MHz bandwidth transmissions.

MAC layer protocols for FD wireless networks

The main goal of a MAC protocol is to coordinate access to the shared communication channel among the nodes of a network. This task can be centralized if a network has an *infrastructure* configuration, in which a central node (e.g., the AP in WLANs) is within the communication range of all the nodes in the network and can broadcast a schedule, regulating their access to avoid collisions. Conversely, in *ad hoc* configurations where all nodes have the same priority and multiple collision domains can exist, a distributed strategy to access the channel must be envisioned.

When the nodes in the network have FD capabilities, the problem of coordinating channel access, be it centralized or distributed, is even more complex. There can be two

possible kinds of FD links, as represented in Fig. 5.2: a *bidirectional* or *symmetric* one, in which two nodes, a and b, communicate simultaneously; and a *relay* or *asymmetric* one, in which a relay node r receives from a source node s and simultaneously transmits to a destination node d. In both cases, the additional transmission allowed by FD capabilities can cause enhanced interference to the surrounding nodes, effectively resulting in collisions if transmissions are not scheduled properly.

Several MAC protocols for FD wireless networks have been reported in the scientific literature (Kim et al., 2015), for both infrastructure and ad–hoc networks and allowing to schedule bidirectional FD links, relay FD configurations or both. Some of these solutions also consider the problem of interference and deal with typical problems such as hidden terminal and Exposed Terminal (ET) (Wang et al., 2012). A brief overview of some of the proposed solutions is reported in the following.

Considering infrastructure network configurations, some schemes have been developed for the case of asymmetric traffic, that aim at identifying bidirectional FD opportunities and solving hidden terminal problems, through either busy tones (Jain et al., 2011) or header snooping, shared backoff and virtual contention resolution (Sahai et al., 2011). These works do not consider interference among nodes, unlike the protocol proposed in Kim et al. (2013), which develops a centralized algorithm in which both bidirectional and relay FD links can be scheduled. The authors in Choi et al. (2015) presents a power–controlled MAC, where the transmit power of each node is adapted in order to maximize the Signal–to–Interference plus Noise Ratio (SINR) in bidirectional and relay FD transmissions.

More strategies are available for ad hoc networks, such as that mentioned in Singh et al. (2011), which proposes a distributed scheduling protocol aimed at enhancing efficiency while preserving fairness among the scheduled bidirectional and relay FD links. Always in ad hoc configurations, the work in Duarte et al. (2014) makes use of RTS/CTS packets to identify FD transmission opportunities, although only bidirectional ones. The work in Cheng et al. (2013) adopts a similar strategy and considers also relay configurations. The MAC scheme proposed in Goyal et al. (2013) deals with contention resolution techniques to handle inter–node interference in both bidirectional and relay FD configurations. The combined use of FD wireless and directional antennas to enhance relay configurations is proposed in Miura and Bandai (2012), whereas a protocol based on synchronous channel access for both bidirectional and relay FD configurations is proposed in Tamaki et al. (2013). Finally, a cross–layer approach based on PHY layer node signatures is exploited in Zhou et al. (2013) to schedule bidirectional and relay FD transmissions in ad hoc network configurations.

5.2 The RCFD full duplex MAC protocol

In this section an original MAC protocol for wireless networks composed by FD devices is proposed. The protocol targets the ad hoc network configuration and it is limited to bidirectional FD links, not taking into account relay configurations. The protocol is called RTS/CTS in the Frequency Domain (RCFD), since it mimics the RTS/CTS exchange often used to negotiate transmission opportunities, but the exchange is performed in frequency rather than in time, as described in the rest of this section.

Motivation

As previously underlined, the possibility for a node to receive and transmit at the same time increases the exposure to interference and considerably complicates the scheduling of transmissions. Consequently, the design of new channel access schemes to efficiently exploit the FD capabilities and produce significant performance gains compared to currently deployed Half–Duplex (HD) systems represents a very important and timely research topic. Specifically, an original channel access scheme for FD ad hoc wireless networks is proposed, where contention is performed in the frequency domain, allowing to overcome many limitations of the traditional, time–based, channel access protocols.

Limitations of traditional channel access schemes

The majority of currently available strategies for distributed channel access coordination in ad hoc wireless networks are "time domain" strategies, able to ensure high throughput and fairness, but failing to provide good enough performance in terms of delay and efficiency.

In detail, these strategies are all based on some variations of the CSMA/CA protocol, where each node that wishes to transmit does a preliminary channel sensing operation and, if no activity is detected, it waits for a random waiting time (backoff) to avoid collision, then it finally transmits. This approach has two main drawbacks. First, the randomness in waiting time allows to ensure fairness, but it causes a non–deterministic (and possibly unbounded) channel access time, thus harming the industrial control applications discussed in this thesis and, in general, any application that requires QoS guarantees. Second, the channel sensing procedure is limited by the sensing range and can cause problems such as hidden terminal and ET. Specifically, in the former two nodes wishing to transmit to the same receiver do not sense each other and then transmit simultaneously, causing a collision. In the latter, two nodes a and b wishing to transmit to nodes c and d which are out of the respective sensing ranges do not transmit

simultaneously, even if they could, resulting in an underutilization of the channel (Wang et al., 2012). The RTS/CTS strategy allows to solve at least the hidden terminal problem (the most severe one) through the exchange of specific frames that allow a pair of nodes to negotiate a transmission. However, this procedure is time–consuming and it still does not guarantee a complete immunity from collisions.

These approaches are called time domain strategies because a certain amount of time is employed for channel contention operations, namely sensing, backoff and exchange of RTS/CTS frame. The duration of this channel contention phase is long and random, representing a problem for applications that require low–latency and deterministic communication.

Frequency domain channel access

In an attempt to overcome the limitations of standard channel access schemes for wireless networks in the time domain, researchers have proposed to move the channel contention procedure to the frequency domain (Sen et al., 2010). Such an approach exploits the use of OFDM modulation at the PHY layer, employed by the majority of wireless standards, as reported in Chap. 3. OFDM provides an ordered set of subchannels or Subcarriers (SCs), equally spaced in frequency within a single wideband wireless channel.

The idea behind frequency domain contention is to let the nodes contend for the channel by randomly selecting one of the SCs and assign the channel to the node that has chosen, for example, the one with the lowest frequency. This strategy resolves contention in a short deterministic time, even for a large number of nodes, compared to conventional time domain schemes, such as CSMA/CA. The approach was upgraded and extended to handle multiple collision domains in Sen et al. (2011), where the Backoff to Frequency (BACK2F) protocol was introduced. A similar strategy was suggested in Feng et al. (2012), where the set of available SCs is divided into two subsets, one destined to random contention and the other to node identification. Here the ACK procedure was also moved to the frequency domain, allowing a further improvement of the efficiency.

Although this approach is promising in that it resolves contentions in a deterministic amount of time, it still suffers from certain issues that affect the MAC layer of wireless ad hoc networks, such as hidden terminal and ET. Moreover, none of the currently proposed frequency domain protocols is designed to handle channel access in FD wireless networks, while the availability of a large number of SCs in OFDM networks can be exploited to effectively identify and select FD opportunities. In addition, it has been suggested that FD communications could help limit the SC leakage problem, which affects the performance of channel access schemes based on frequency domain contention (Sen et al.,

2011).

Combining time and frequency domain strategies

In this section, a MAC layer protocol for ad hoc FD wireless networks based on time–frequency contention is proposed, that is capable of efficiently exploiting FD transmission opportunities and resolving collisions in a short and deterministic amount of time. To this end, a frequency domain approach based on multiple contention rounds in time, each using an OFDM symbol, is employed. This framework is exploited to advertise the transmission intentions of the nodes and to select, within each contention domain, the pair of nodes that will actually perform a data exchange. The presented scheme is fully distributed, effectively handles multiple contention domains, and preserves sufficient randomness to ensure fairness among different users.

The proposed approach is able to take the best out of the two strategies previously presented, namely time domain and frequency domain contention. Indeed, compared to frequency domain MAC protocols, such as Sen et al. (2011), the proposed scheme allows to eliminate the hidden terminal issue, exploiting the multiple round RTS/CTS procedure. Moreover, compared against previously reported time domain MAC protocols for FD wireless networks, such as Duarte et al. (2014), RCFD exhibits an increased efficiency as well as a reduced delay.

Protocol description

The RCFD algorithm is a channel access scheme based on a time and frequency domain approach. According to this strategy, not only the medium contention, but also transmission identification and selection are performed over multiple consecutive frequency domain contention rounds.

System model and assumptions

RCFD is designed for an ad hoc wireless network composed of N nodes with the same priority. Each node is assumed to have perfect FD capabilities, i.e., it can simultaneously receive a signal while transmitting in the same frequency band with perfect SIC. OFDM is adopted at the physical layer to transmit consecutive symbols over a set of S subcarriers. During the channel contention phase only, nodes transmit on single SCs while listening to the whole channel. In the data transmission phase, instead, only one pair of nodes transmit and receive in each collision domain, exploiting all SCs available in the selected channel, as generally done in existing IEEE 802.11 networks.

The proposed protocol relies on some assumptions that ensure its correct behavior. The validity of these assumptions as well as the possibility of relaxing them will be discussed further in this section. First, it is assumed that all nodes have data to send and try to access the channel simultaneously. The communication channel is assumed ideal (no external interference, fading or path loss), so that each node can hear every other node within its coverage range. However, there can be multiple collision domains, i.e., the range of a node may not include all the nodes in the network.

It is assumed that a unique association between each node and two OFDM subcarriers is initially established at network setup, maintained fixed throughout all operations and available to each node. More specifically, defining $\mathcal{S} = \{s_1, \ldots, s_S\}$ as the set of available SCs, it is split in two non–overlapping parts $\mathcal{S}_1$ and $\mathcal{S}_2$. Taking $\mathcal{N} = \{n_1, \ldots, n_N\}$ as the set of network nodes, a mapping is defined by the two functions

$$\mathcal{F}_1 : \mathcal{N} \to \mathcal{S}_1, \quad \mathcal{F}_2 : \mathcal{N} \to \mathcal{S}_2 \tag{5.1}$$

that uniquely link any node with an associated SC in each set. A simple implementation of such a map can be obtained by taking $\mathcal{S}_1 = \{s_1, \ldots, s_{S/2}\}$, $\mathcal{S}_2 = \{s_{S/2+1}, \ldots, s_S\}$ and defining $\mathcal{F}_1(n_i) = s_i$, $\mathcal{F}_2(n_i) = s_{i+S/2}$, $i = 1, \ldots, N$. It is worth stressing that the correspondence between a node and each of the two SCs must be unique, i.e., $\mathcal{F}_1(n_i) \neq \mathcal{F}_1(n_j)$ and $\mathcal{F}_2(n_i) \neq \mathcal{F}_2(n_j)$ for every $i \neq j$. Finally, it has to be noted that the assumed mapping imposes a constraint on the number of nodes in the network. Indeed, since each node must be uniquely associated with two OFDM SCs, the total number of nodes has to be less than or equal to $S/2$.

Channel contention scheme

The channel access procedure is composed of three consecutive contention rounds in the frequency domain. The first round starts after each node has sensed the channel and found it idle for a certain period of time T_{scan}. Each round consists in the transmission of an OFDM symbol and its duration is set to $T_{round} = T_{sym} + 2T_p$ to accommodate for signal propagation, which takes a time T_p each way (Sen et al., 2011). Therefore, the access procedure takes a fixed time of

$$T_{acc} = T_{scan} + 3T_{round} \tag{5.2}$$

As an example, if an IEEE 802.11g network is considered, standard values for these parameters are $T_{scan} = 28$ μs (the duration of a DIFS), $T_{sym} = 4$ μs, and $T_p = 1$ μs, thus obtaining $T_{acc} = 46$ μs.

In the following, the steps performed by every node in each contention round are outlined.

First round - randomized contention Every node that has data to send and has found the channel idle for a T_{scan} period, randomly selects a SC from the whole set $\mathcal{S}$ and transmits a symbol only on that SC, while listening to the whole channel band.

Let $\bar{s}_i$ denote the SC chosen by node n_i and $\mathcal{S}_i^1$ the set of SCs that actually carried a symbol during the first contention round, as perceived by node n_i.

Node n_i is defined as Primary Transmitter (PT) if and only if the following condition holds

$$\bar{s}_i = \min_j \left[s_j \in \mathcal{S}_i^1 \right] \tag{5.3}$$

i.e., the lowest–frequency SC among those carrying data is the one chosen by the node itself. It is noteworthy that, in a realistic scenario with multiple collision domains, several nodes in the network can be selected as PTs. Moreover, if multiple nodes in the same collision domain pick the same lowest–frequency SC, they are all selected as PTs. This potential collision will be resolved in the following contention rounds, as it will be detailed later on.

Second round - transmission advertisement (RTS) Only the nodes who identify themselves as PTs during the first round transmit during the second round. A PT node n_i that has data to send to node n_j transmits a symbol on two SCs, namely $s_a = \mathcal{F}_1(n_i) \in \mathcal{S}_1$ and $s_b = \mathcal{F}_2(n_j) \in \mathcal{S}_2$. In this way, n_i informs its neighbors that it is a PT and has a packet for n_j. This round is the so–called RTS part of the algorithm, as it resembles the time domain RTS procedure defined in the IEEE 802.11 standard. During the second round, all the nodes in the network (including the PTs) listen to the whole band. $\mathcal{S}_{h,1}^2 \subseteq \mathcal{S}_1$ and $\mathcal{S}_{h,2}^2 \subseteq \mathcal{S}_2$ are the sets of SCs that carried a symbol during the second contention round, as observed by a generic node n_h.

Node n_h is defined as RTS Receiver (RR) if and only if the following condition holds

$$n_h \text{ is not PT } \wedge \mathcal{F}_2(n_h) \in \mathcal{S}_{h,2}^2 \tag{5.4}$$

i.e., at least one PT node advertised, during the second round, that it has a packet for n_h. There can be multiple RRs in the network, but a node cannot be both PT and RR at the same time. Indeed, according to Eq. (5.4), even if a node that is PT receives an RTS (e.g., due to a first round collision where two nodes in the same domain selected the same lowest–frequency subcarrier), it does not take it into account and does not define

itself as RR.

Third round - transmission authorization (CTS)

Only the nodes selected as RR during the second round transmit in the third one. Any RR node n_h will select its CTS recipient as

$$n_l = \arg\min_{n_i} \left[\mathcal{F}_1(n_i) \ : \ \mathcal{F}_1(n_i) \in \mathcal{S}^2_{h,1} \right] \tag{5.5}$$

i.e., among the nodes that have sent an RTS to n_h, the one with the lowest corresponding SC is selected.[1] Node n_h then transmits a symbol on two SCs, namely $s_c = \mathcal{F}_1(n_h) \in \mathcal{S}_1$ and $s_d = \mathcal{F}_2(n_l) \in \mathcal{S}_2$. In this way, n_h informs n_l that its transmission is authorized. Since this round mimics the operation of the time domain CTS procedure, it is referred to as the CTS part of the RCFD algorithm. During the third round, all the nodes in the network (including the RRs) listen to the whole channel band. $\mathcal{S}^3_{i,1} \subseteq \mathcal{S}_1$ and $\mathcal{S}^3_{i,2} \subseteq \mathcal{S}_2$ are the sets of SCs that carried a symbol during the third round, as observed by a generic node n_i.

At the end of the third round, each node that has data to send needs to decide whether to transmit or not, according to the information gathered in the three rounds. Specifically, for a generic node n_i which has a packet for node n_j, three cases can be distinguished:

I. *Node n_i is a PT:*
It transmits if and only if both these conditions are verified

$$\begin{aligned} &\mathcal{F}_1(n_j) \in \mathcal{S}^3_{i,1} \\ &\mathcal{S}^3_{i,2} = \{\mathcal{F}_2(n_i)\} \end{aligned} \tag{5.6}$$

i.e., the intended receiver (node n_j) has sent a CTS and this is the only CTS within the contention domain of node n_i.

II. *Node n_i is an RR:*
It transmits (while receiving from the PT, thus enabling bidirectional FD) if and

[1] This choice may impair the fairness of the RCFD protocol if the subcarrier mapping is static. To avoid such a problem, periodic permutations of the maps $\mathcal{F}_1$ and $\mathcal{F}_2$ according to a common pseudorandom sequence can be scheduled. The exchange of broadcast messages advertising the new maps after each permutation might be needed to avoid synchronization issues among the nodes. This expedient is not implemented in the simulations that will be shown, which, however, report a good fairness level.

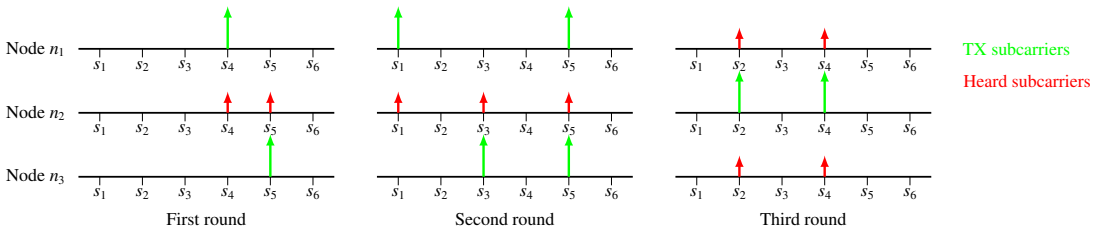


Figure 5.3: Outcomes of contention rounds for example scenario 1.

only if both these conditions are verified

$$\begin{aligned} \mathcal{S}_{i,1}^{2} &= \{\mathcal{F}_1(n_j)\} \\ \mathcal{S}_{i,1}^{3} &= \{\mathcal{F}_1(n_i)\} \end{aligned} \tag{5.7}$$

i.e., only the intended receiver (node n_j) has sent an RTS and no other node has sent a CTS (except node n_i itself).

III. *Node n_i is neither a PT nor an RR:*
It does not transmit.

It is worth to point out that not only may the nodes selected as PTs during the first round be granted access to the channels, but also an RR can transmit, if the conditions in case II are verified. This possibility is the key to enable FD transmission: a node that has a packet for another node from which it has received an RTS can send it together with the primary transmission (provided that no other CTSs from surrounding nodes were received).

It must be observed that the RCFD protocol only allows for bidirectional FD and does not take into account the relay FD opportunities displayed in Fig. 5.2. A modification of RCFD to accommodate for relay FD is left for future research.

Examples of operation

In order to better understand how the proposed MAC strategy works, two examples are provided, for a simplified system with $N = 3$ nodes and $S = 6$ OFDM subcarriers. The simplest scheme is adopted for SC mapping, i.e., $\mathcal{S}_1 = \{s_1, s_2, s_3\}$, $\mathcal{S}_2 = \{s_4, s_5, s_6\}$, $\mathcal{F}_1(n_i) = s_i$, $\mathcal{F}_2(n_i) = s_{i+3}$, $i = 1, 2, 3$.

Two different example scenarios are considered. Fig. 5.3 and Fig. 5.4 show the contention rounds for scenarios 1 and 2, respectively, while Fig. 5.5 reports the network topology and the transmission intentions. In both scenarios, node n_2 is within the

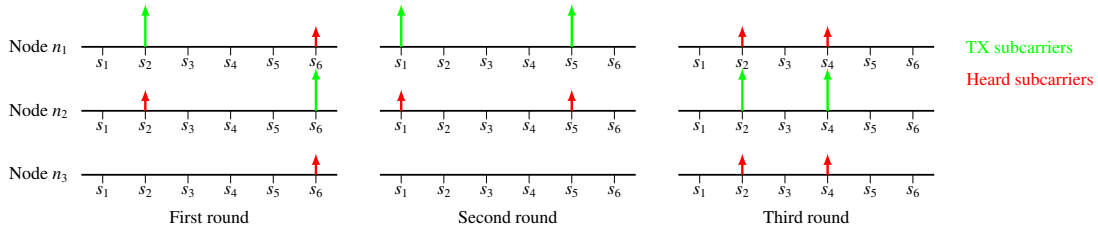


Figure 5.4: Outcomes of contention rounds for example scenario 2.

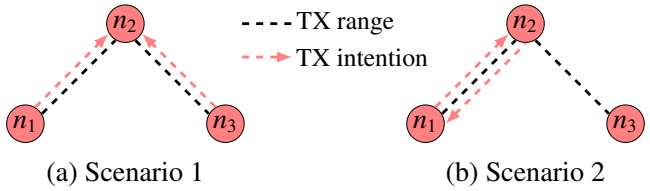


Figure 5.5: Topology and transmission intentions for the operation examples.

transmission range of nodes n_1 and n_3 that, however, cannot sense each other (two collision domains). In the first scenario, nodes n_1 and n_3 both intend to send a packet to n_2, resembling a typical hidden terminal situation. In the second one, nodes n_1 and n_2 have a packet for each other, representing a potential FD communication instance.

As seen in Fig. 5.3 for scenario 1, in the first round the two nodes with data to send randomly select two SCs as $\bar{s}_1 = s_4$ and $\bar{s}_3 = s_5$, with the result that both n_1 and n_3 are selected as PTs, since they cannot sense each other's transmissions. Consequently, in the second round they both transmit, causing n_2 to hear signals on SCs s_1, s_3 and s_5. According to Eq. (5.4), n_2 is selected as RR and transmits, during the third round, on SCs s_2 and s_4. Finally, according to Eq. (5.6), node n_1 is allowed to transmit, whereas the transmission by node n_3 is denied, since $\mathcal{S}^3_{3,2} = \{s_4\}$ and $\mathcal{F}_2(n_3) = s_6$. It can hence be observed that the hidden terminal problem has been identified and solved thanks to the RCFD strategy.

In scenario 2, as depicted in Fig. 5.4, nodes n_1 and n_2 participate in the first contention round, randomly selecting $\bar{s}_1 = s_2$ and $\bar{s}_2 = s_6$, therefore only n_1 is selected as PT. In the second round, n_1 transmits on SCs s_1 and s_5, thus node n_2 is selected as RR. Finally, in the third round n_2 transmits on SCs s_2 and s_4, providing a CTS to node n_1. Since the conditions in Eq. (5.6) are verified for n_1 and those in Eq. (5.7) are fulfilled for n_2, both nodes are cleared to transmit, thus enabling full–duplex transmission. If node n_2 had been selected as PT in the first round, the final outcome would have been the same

(n_1 selected as RR and subsequently cleared to transmit).

Protocol optimization and discussion

The assumptions on which the RCFD strategy is based are discussed, together with its limitations and some possible enhancements.

Enhancements to the subcarrier mapping scheme

The subcarrier mapping upon which the RCFD scheme relies imposes a limit on the number of nodes in the network, which has to be no higher than $S/2$.

It is worth stressing that the trend in wireless networks based on the IEEE 802.11 standard is to use wider channels, that offer an ever increasing number of SCs. As an example, IEEE 802.11ac introduces 160 MHz channels, that can accommodate 512 SCs and hence allow RCFD to reach up to 256 users (Perahia and Stacey, 2013).

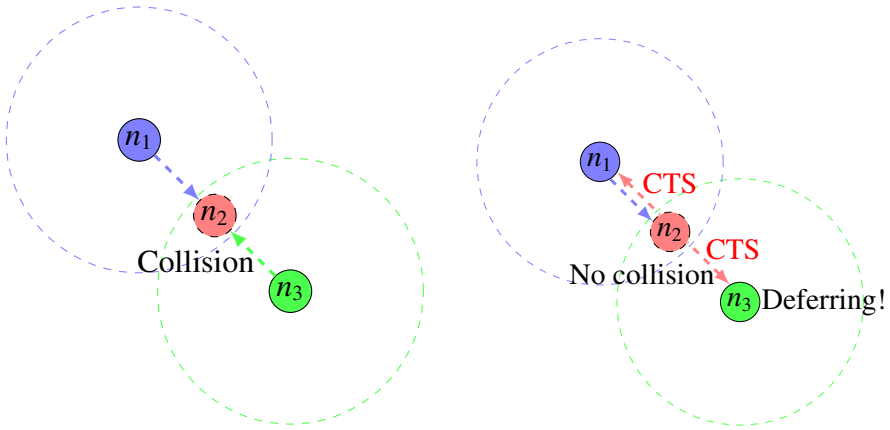
The number of nodes can be further increased even maintaining a fixed number of SCs if the information carried in each SC is exploited. In the presented version of the algorithm only the presence or absence of data on an SC was taken into account. A more refined version would discriminate between the actual content of the symbol transmitted in a specific SC, to be able to host multiple nodes within the same subcarriers. Each SC can carry $\log_2 m$ bits if an m-ary modulation is adopted and, in this way, the maximum number of users in the system can be increased to $m \cdot S/2$. As an example, if $S = 64$ SCs are available and a 64–QAM modulation is employed, a total of 2048 users can be hosted in the network.

Tab. 5.1 provides an example of extended subcarrier mapping in a system with $S = 4$ SCs which adopts a modulation of order $m = 4$, hence allowing the presence of 8 users. In this scenario, for instance, if node n_1 has to advertise a transmission to node n_6 in the second contention round, it would transmit bits 00 on SC s_1 (to advertise itself) and bits 01 on SC s_4 (to advertise the intended receiver).

Another possible issue of the proposed subcarrier mapping is that it must be established at network setup, representing a problem in dynamic ad hoc networks where nodes join and leave continuously. To overcome this issue, each node should keep track of the first available slots in the maps $\mathcal{F}_1$ and $\mathcal{F}_2$. Whenever a node leaves the network, it should send a broadcast message indicating its slots, so that all remaining nodes mark them as free and update the information on the first available slots. When a node joins the network, conversely, it sends a broadcast message and waits for a reply, which will assign it the first available slots. In networks with multiple collision domains, the broadcast messages need to be propagated so that all nodes update the information and share the

Table 5.1: Example of extended SC mapping

Node	$\mathcal{S}_1$		$\mathcal{S}_2$	
	SC Number	Data on SC	SC Number	Data on SC
n_1	s_1	00	s_3	00
n_2	s_1	01	s_3	01
n_3	s_1	10	s_3	10
n_4	s_1	11	s_3	11
n_5	s_2	00	s_4	00
n_6	s_2	01	s_4	01
n_7	s_2	10	s_4	10
n_8	s_2	11	s_4	11



(a) Standard procedure: collision

(b) Deferring: no collision

Figure 5.6: Example scenario of asynchronous channel access with potential collisions.

same version of the maps. Such a strategy will work with minor overhead if the network is not too dynamic.

Asynchronous channel access

An important assumption on which RCFD is based is that the channel access is synchronous, i.e., all nodes try to access the channel at the same time. This is not realistic, since in real networks nodes often generate packets, and therefore try to access the channel, in an independent manner. As a consequence, when the proposed algorithm is implemented in a network with multiple collision domains, a node may start a contention procedure while another node within its range is receiving data, thus causing a collision. Indeed, the scanning procedure performed before the contention rounds is only capable

of determining if a surrounding node is transmitting, not if it is receiving.

Fig. 5.6a reports an example of such a situation, where node n_3 tries to access the channel while node n_1 is already performing a data transmission to n_2, which is inside the coverage range of both nodes. When n_3 starts the first transmission round, it causes a collision with the ongoing transmission.

To cope with this issue, a simple yet effective modification to the original RCFD algorithm is carried out, so that an idle node (i.e., a node that does not have a packet to send, such as n_3 in Fig. 5.6b) which hears a CTS from a neighboring node refrains from accessing the channel until the end of the transmission is advertised through an ACK packet. To prevent freezing (in case the ACK is lost), a timeout can be started upon CTS detection and the node can again access the channel after its expiration. Fig. 5.6b shows that, if such a deferring policy is adopted, no collision happens in the previously described scenario.

Impact of fading, lock problems and collisions

In all the discussions so far an ideal channel was assumed. Real wireless communication environments are characterized by impairments such as fading, shadowing and path loss. For the proposed scheme, the case of selective fading, in which only narrow portions of the spectrum (corresponding to one or few subcarriers) are disturbed, is particularly challenging. Such a phenomenon could lead to sub–channel outage and the emergence of False Negatives (FNs), i.e., missed detection of data on a subcarrier (Sen et al., 2011).

The impact of FNs in the three contention rounds of RCFD can be summarized as follows:

1. *First round:* Multiple PTs can be selected in the same collision domain as a result of FNs; as a consequence, nodes that should be RR in the second round would be PT instead and would not send the CTS in the third round, thus leading to missed transmission opportunities.

2. *Second round:* A FN during the second round could lead to a node not receiving an RTS destined to it, again resulting in a missed opportunity for a transmission which, however, should have been authorized.

3. *Third round:* Again, a FN occurrence during the third round results in a missed CTS reception and a corresponding missed transmission opportunity.

In conclusion, FNs induced by sub–channel outage never result in a collision but only in possible missed transmission opportunities, thus causing underutilization of the channel and slightly degrading the efficiency of the protocol.

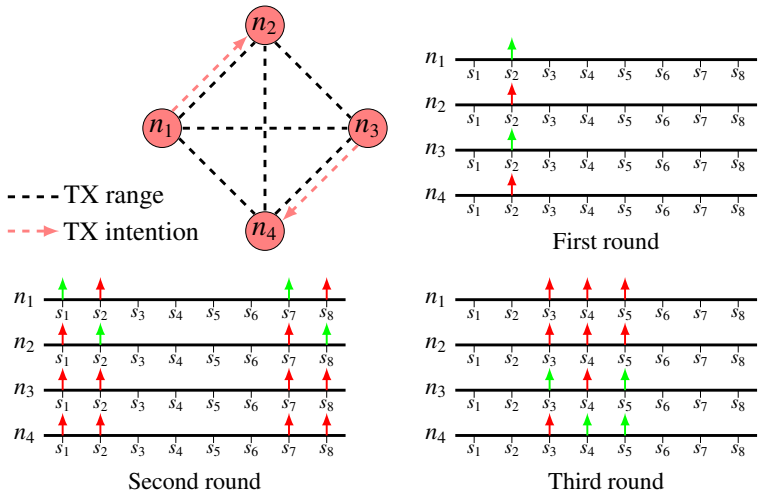


Figure 5.7: Example of a 4–node network in one collision domain where two nodes (n_1 and n_3) become PT simultaneously because they select the same subcarrier in the first round (s_2). The contention is solved in the third round where only n_1 receives the CTS and, hence, is cleared to transmit.

Similarly, channel underutilization may be caused by "lock" problems that arise for particular selections of subcarriers in the first round.[2] However, a different random selection is carried out at each transmission opportunity, thus preventing permanent lock problems. In general, RCFD is designed to ensure that collisions are avoided, at the cost of losing a transmission opportunity every now and then.

Another possible issue arises in RCFD when multiple nodes select the same SC in the first contention round, when the SC choice is random. This represents a problem in the BACK2F scheme (Sen et al., 2011), that was addressed by performing multiple rounds but still maintaining a residual collision probability. Conversely, in the proposed protocol, this could result in multiple PTs being present, only one of which is selected in the following rounds, thus preventing any possible collision. An example of how RCFD handles the issue of multiple PTs in the same collision domain is provided in Fig. 5.7.

Finally, real–world implementations of FD devices likely do not achieve perfect SI cancellation and may be impaired by RSI. If this interference is too high, it can impact RCFD in two ways. First, every bidirectional FD transmission will be less robust, leading to lower overall performance. Second, the detection of SCs in the three contention rounds for a node that is also transmitting in one or more rounds may be more difficult, and

[2]For example, consider a "line" network where adjacent nodes are in the same collision domain and they select SCs in the first round in ascending order. Only the first node will be the PT and transmit, while other concurrent transmissions may have been allowed.

false negatives such as those described at the beginning of this subsection may occur. However, it has to be noted that working implementations of FD devices able to reduce the residual SI to the noise floor can be found in the literature (Bharadia et al., 2013). Nevertheless, future activities are planned to assess the performance of RCFD under different levels of RSI.

Possible protocol improvements

The RCFD protocol in the presented form already yields significant performance benefits with respect to traditional channel access schemes.

Further improvements in channel utilization can be achieved if the ACK procedure is also moved to the frequency domain, as already suggested in Feng et al. (2012). The implementation of this enhancement would be straightforward, since a mapping between nodes and subcarriers is already established.

Moreover, as discussed in Sen et al. (2010) and Sen et al. (2011), the random selection of OFDM subcarriers implicitly defines an order among the nodes trying to access the channel, thus enabling the possibility of fast and efficient TDMA–like transmissions. Alternatively, the order among nodes can be exploited if the nodes in the network have different priorities. In this case, the first round of the RCFD algorithm can be modified by letting a high priority node randomly choose its SC among a subset of $\mathcal{S}$ which contains lower frequency SCs with respect to the set in which a low priority node picks its SC. This would guarantee to the former node a higher probability of being selected as a PT and, hence, a faster channel access.

For the sake of clarity, the version of RCFD evaluated in the next sections does not include these improvements, whose detailed design and performance evaluation are left for future research. However, it does include the extended SC mapping as well as the deferring policy to allow asynchronous access scheme.

Theoretical analysis

In order to validate the proposed protocol and highlight the benefits it is able to provide, its performance are compaired against those offered by standard MAC algorithms for wireless networks and other state–of–the–art strategies.

In this section a theoretical comparison based on the analytical evaluation of the *normalized saturation throughput* of different MAC algorithms is provided. This quantity is defined as the maximum load that a system is able to carry without becoming unstable (Bianchi, 2000). It can also be seen as the percentage of time in which nodes with full buffers utilize the channel for data transmission using a contention–based MAC scheme.

Table 5.2: System parameters for theoretical analysis

Parameter	Description	Value
T_{ack}	MAC–layer ACK transmission time	50 μs
T_{rts}	RTS frame transmission time	58 μs
T_{cts}	CTS frame transmission time	50 μs
T_{sifs}	SIFS	10 μs
T_{difs}	DIFS	28 μs
T_p	Propagation time over the air	1 μs
T_{slot}	MAC–layer slot time	9 μs
W	Initial value of backoff window	16
m	Maximum number of retransmission attempts	6
S	Number of available OFDM subcarriers	52
T_{round}	Duration of a contention round in the frequency domain	6 μs

In order to make the problem analytically tractable, some assumptions are made. A network of N nodes is considered, all within the same collision domain and with saturated queues, meaning that every node always has at least one packet to transmit. A First–in First–out (FIFO) policy is adopted at each node, meaning that only the packet at the head of the queue can be transmitted. Furthermore, an ideal communication channel is assumed, so that the only cause of transmission errors would be collisions among different packets. In the case of frequency–based channel access schemes, it is assumed that the exchange of data on subcarriers during the contention round works perfectly, regardless of the number of nodes in the network (if $N > S/2$ it can be assumed that an extended mapping scheme is adopted). Finally, both the transmission rate R and the payload size L (in bytes) are fixed.

Four different MAC layer protocols are compared with the proposed RCFD strategy. The baseline scheme is the IEEE 802.11 DCF proposed in the standard (IEEE 802.11-2016), both with and without the RTS/CTS option. The FD MAC strategy (Duarte et al., 2014) was selected among the various time–domain MAC protocols for FD networks since it is one of the most general approaches, and does not impose any assumption on network topology, traffic pattern or PHY configuration. Finally, the BACK2F scheme (Sen et al., 2011) has been chosen as a protocol that performs channel contention in the frequency domain.

In order to obtain a fair comparison, all the protocols are based on the same underlying physical layer, specifically that described by the IEEE 802.11g standard, which is very widespread. Tab. 5.2 reports the main parameters considered in this theoretical analysis.

Analysis for IEEE 802.11 and FD MAC

The starting point for the analysis is the work in Bianchi (2000), where the normalized saturation throughput was derived for the IEEE 802.11 DCF (with and without RTS/CTS). The main results of that study are outlined in the following and then extended to evaluate the normalized saturation throughput for the FD MAC algorithm (Duarte et al., 2014).

The IEEE 802.11 DCF is based on a CSMA/CA strategy, where nodes listen to the channel before transmitting. If they find it busy, they wait until it becomes idle, and then defer transmission for an additional random backoff period in order to avoid collisions. The first analysis step is, hence, the introduction of a discrete–time Markov model to describe the behavior of a single station during backoff periods. This model was then used to derive the probability τ that a single station transmits in a randomly chosen slot and the probability p that a transmission results in a collision, as functions of the system parameters, such as the initial value of the backoff window W and the maximum number of backoff stages m. Subsequently, two probabilities were computed, namely P_{tr}, the probability that at least a transmission attempt takes place in a slot, and P_s, the probability that this transmission is successful, expressed as functions of the number of nodes in the network N, and of the probabilities τ and p. Specifically, the number of stations that transmit in a given slot is a binomial random variable B of parameters N and τ and the probabilities P_{tr} and P_s can be expressed as

$$P_{tr} = P\left(B \geq 1\right) = 1 - \left(1 - \tau\right)^{N} \tag{5.8}$$

$$P_s = P\left(B = 1 | B \geq 1\right) = \frac{N\tau\left(1-\tau\right)^{N-1}}{1-\left(1-\tau\right)^{N}} \tag{5.9}$$

Finally, the saturation throughput can be computed as

$$\eta_{DCF} = \frac{P_{tr} P_s T_d}{\left(1 - P_{tr}\right) T_{slot} + P_{tr} P_s T_S + P_{tr}\left(1 - P_s\right) T_C} \tag{5.10}$$

where T_d is the payload transmission time, T_{slot} is the slot time in IEEE 802.11, T_S is the slot duration in case of a successful transmission and T_C is the slot duration in case of a collision. The values for T_S and T_C, as computed in Bianchi (2000), are

$$T_S = T_{difs} + T_d + T_{sifs} + T_{ack} + 2T_p \tag{5.11}$$

$$T_C = T_{difs} + T_d + T_p \tag{5.12}$$

for the standard IEEE 802.11 DCF without RTS/CTS and

$$T_S = T_{difs} + T_{rts} + T_{cts} + T_d + 3T_{sifs} + T_{ack} + 4T_p \tag{5.13}$$

$$T_C = T_{difs} + T_{rts} + T_p \tag{5.14}$$

in case the RTS/CTS option is enabled. The meaning and the values of parameters T_{difs}, T_{sifs}, T_{rts}, T_{cts}, T_{ack} and T_p are reported in Tab. 5.2, whereas the transmission time T_d for a packet of length L sent at rate R can be derived from the IEEE 802.11 specifications (IEEE 802.11-2016).

The FD MAC algorithm, presented in Duarte et al. (2014), builds on the IEEE 802.11 DCF with the use of RTS and CTS frames, with a substantial difference: when node n_j receives an RTS from node n_i, it checks at the head of its transmission queue if there is a packet destined to n_i and, if present, starts to transmit it immediately after the CTS frame, with a waiting period of T_{sifs}. Other minor modifications to the DCF include the possibility for a node to receive both a data frame and an ACK frame within a NAV interval and the possibility to send an ACK while waiting for another ACK (Duarte et al., 2014).

The analysis presented for the IEEE 802.11 DCF in Bianchi (2000) is extended to account also for the FD MAC, taking into account that a successful FD transmission can occur in two different cases. The first one is when only two nodes grab the channel simultaneously and have packets for each other, which happens with probability

$$\frac{P\left(B=2|B\geq 1\right)}{\left(N-1\right)^2} = \frac{N\tau^2\left(1-\tau\right)^{N-2}}{2\left(N-1\right)\left(1-\left(1-\tau\right)^N\right)} \tag{5.15}$$

since the probability that a generic node has a packet for another specific node is $1/\left(N-1\right)$. A successful FD communication takes place also if a single node grabs the channel, which happens with probability expressed by Eq. (5.9), and the target receiver has a packet for it at the head of the queue, which happens with probability $1/\left(N-1\right)$. Hence, the probability that a successful FD transmission takes place is given by

$$P_{s,fd} = \frac{P\left(B=2|B\geq 1\right)}{\left(N-1\right)^2} + \frac{P\left(B=1|B\geq 1\right)}{N-1} = \frac{N\tau\left(1-\tau\right)^{N-2}\left(2-\tau\right)}{2\left(N-1\right)\left(1-\left(1-\tau\right)^N\right)} \tag{5.16}$$

A successful HD transmission happens when a single node grabs the channel but the target receiver does not have a packet for it, which occurs with probability

$$P_{s,hd} = P\left(B=1|B\geq 1\right)\left(1-\frac{1}{N-1}\right) = \frac{N\left(N-2\right)\tau\left(1-\tau\right)^{N-1}}{\left(N-1\right)\left(1-\left(1-\tau\right)^N\right)} \tag{5.17}$$

Consequently, the saturation throughput is given by

$$\eta_{FD} = \frac{T_d P_{tr} P_{s,hd} + 2T_d P_{tr} P_{s,fd}}{(1 - P_{tr}) T_{slot} + P_{tr} P_s T_S + P_{tr} (1 - P_s) T_C} \tag{5.18}$$

where P_{tr}, T_S and T_C are expressed by Eq. (5.8), (5.13) and (5.14) respectively.

Analysis for BACK2F

The Markov model introduced in Bianchi (2000) is no longer useful with the BACK2F scheme described in Sen et al. (2011). In this channel access scheme, indeed, there cannot be any idle slots (i.e., $P_{tr} = 1$) and the only case in which a transmission is not successful is when there is a collision on the SC selection after the second contention round in the frequency domain. An original Markov model is hence introduced to derive the success probability P_S, i.e., the probability that no collisions happen, as a function of the number of nodes N and the number of available OFDM subcarriers S.

Markov chain model of the channel contention procedure Consider a discrete–time Markov chain that models the three-dimensional process $\{x(t), c(t), y(t)\}$, where $x(t)$ represents the number of nodes winning the first contention round of BACK2F in time slot t, $c(t)$ represents the lowest–frequency SC during the first contention round in the same time slot and $y(t)$ represents the number of nodes winning the second contention round. The processes $x(t)$ and $y(t)$ take values in the set $\{1, \ldots, N\}$, while $c(t)$ can range from 0 to $S - 1$.[3] Trivially, it must hold $y(t) \leq x(t)$, since only the nodes that have won the first round can take part in the second one, and also $x(t) = N$ if $c(t) = S - 1$. Moreover, if $c(t) = S - 1$, it means that all the nodes have won the first contention round, i.e., $x(t) = N$. Taking these constraints into account, the number of reachable states is $N \cdot (N - 1) \cdot (S - 1)/2 + N$.

It can be proved that the proposed chain is time–homogeneous, irreducible and aperiodic and, hence, a stationary distribution can be found as

$$\pi_{i,a,j} = \lim_{t \to \infty} P\{x(t) = i, c(t) = a, y(t) = j\} \tag{5.19}$$

for $i = 1, \ldots, N$, $a = 0, \ldots, S - 1$ and $j = 1, \ldots, i$.

Derivation of the transition probabilities The stationary distribution of Eq. (5.19) is derived from the transition probabilities between the different states of the Markov chain.

[3] Without loss of generality it is assumed that $\mathcal{S} = \{0, 1, \ldots, S - 1\}$.

These probabilities are of the form

$$p_{i,a,j|k,b,l} = P\{x(t) = i, c(t) = a, y(t) = j \,|\, x(t-1) = k, c(t-1) = b, y(t-1) = l\} \tag{5.20}$$

Through some computations, $p_{i,a,j|k,b,l}$ can be factorized in three terms

$$p_{i,a,j|k,b,l} = p_{j|i,a,k,b,l} \cdot p_{i|a,k,b,l} \cdot p_{a|k,b,l} \tag{5.21}$$

where

$$p_{j|i,a,k,b,l} = P\{y(t) = j \,|\, x(t) = i, c(t) = a, x(t-1) = k, c(t-1) = b, y(t-1) = l\} \tag{5.22}$$

$$p_{i|a,k,b,l} = P\{x(t) = i | c(t) = a, x(t-1) = k, c(t-1) = b, y(t-1) = l\} \tag{5.23}$$

$$p_{a|k,b,l} = P\{c(t) = a | x(t-1) = k, c(t-1) = b, y(t-1) = l\} \tag{5.24}$$

Exact expressions for these three terms are derived in the following for all possible values of the parameters i, a, j, k, b, l, according to the structure of the BACK2F algorithm, reported in Algorithm 2 for convenience.

Derivation of $\mathbf{p_{j|i,a,k,b,l}}$ It can first be observed that

$$\begin{aligned} p_{j|i,a,k,b,l} &= P\{y(t) = j \,|\, x(t) = i, c(t) = a, x(t-1) = k, c(t-1) = b, y(t-1) = l\} \\ &= P\{y(t) = j \,|\, x(t) = i\} \end{aligned} \tag{5.25}$$

since the number of winning nodes at the second round only depends on the number of nodes that have won the first round in the same time slot. It can be further stated that Eq. (5.25) is meaningful only for $j \leq i$, which leaves only the two following scenarios.

Scenario I: $j = i$

In this case i nodes are randomly choosing among S subcarriers. The probability that they all pick the same one is given by $1/S^{i-1}$.

Scenario II: $j < i$

The probability that j nodes out of i pick the same SC c and that all the other nodes pick SCs with higher frequency than c is given by

$$\binom{i}{j} \left(\frac{1}{S}\right)^j \left(1 - \frac{c+1}{S}\right)^{i-j} \tag{5.26}$$

Algorithm 2 BACK2F channel access algorithm (Sen et al., 2011).

```
1:  procedure BACK2F(packet)
2:      myback ← rnd [0, S − 1]
3:      wait for T_difs
4:      if channel is busy then
5:          goto line 2
6:      else
7:          transmit on SC myback in round 1
8:          minback ← lowest–frequency SC with signal
9:          myback ← myback − minback
10:         if myback > 0 then                          ▷ Lost round 1
11:             goto line 2
12:         else
13:             myback2 ← rnd [0, S − 1]
14:             transmit on SC myback2 in round 2
15:             minback2 ← lowest–frequency SC with signal
16:             if myback2 = minback2 then              ▷ Won round 2
17:                 transmit packet
18:                 goto line 1
19:             else                                    ▷ Lost round 2
20:                 goto line 2
21:             end if
22:         end if
23:     end if
24: end procedure
```

This probability has to be summed over all possible SCs except the last one (which would result in all the nodes picking the same one, i.e., $j = i$)

$$\sum_{c=0}^{S-2} \binom{i}{j} \left(\frac{1}{S}\right)^{j} \left(1 - \frac{c+1}{S}\right)^{i-j} \tag{5.27}$$

Summing up all the scenarios, the following expression for $p_{j|i,a,k,b,l}$ is obtained:

$$p_{j|i,a,k,b,l} = \begin{cases} \sum_{c=0}^{S-2} \binom{i}{j} \left(\frac{1}{S}\right)^{j} \left(1 - \frac{c+1}{S}\right)^{i-j} & \text{if } j < i \\ \frac{1}{S^{i-1}} & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \tag{5.28}$$

Derivation of $\mathbf{p_{i|a,k,b,l}}$ To compute this second term, the probability that exactly i nodes win the first round at time slot t given that SC a is the lowest–frequency one

and that, in the previous time slot, k nodes won the first round (with SC b) and l nodes won the second round must be derived. Again, the problem is split in multiple scenarios.

Scenario I: $k \neq l$

In this scenario, at the end of time slot $t-1$ the following groups of nodes can be distinguished:

A) $N-k$ nodes that have lost round 1 and, hence, have $myback > 0$ (line 10 in Algorithm 2).

B) $k-l \neq 0$ nodes that have lost round 2 and, hence, have $myback = 0$ (line 12 in Algorithm 2).

C) l nodes that have won round 2 and, after transmitting, have $myback$ randomly distributed between 0 and $S-1$ (lines 18 and 2 in Algorithm 2).

Therefore, the following observations can be made:

- The lowest–frequency SC at time t is 0 (chosen by at least the nodes of group B), hence $p_{i|a,k,b,l}$ is always 0 when $a \neq 0$.
- There are at least $k-l$ nodes (group B) that have $myback = 0$ and win round 1, hence, $i \geq k-l$.
- The maximum number of first round winners is k, since $N-k$ nodes (group A) have $myback > 0$, hence, $i < k$.

  The probability that m of the l nodes of group C pick 0 as a SC and hence win round 1 at time slot t is

$$\binom{l}{m}\left(\frac{1}{S}\right)^{m}\left(1-\frac{1}{S}\right)^{l-m} \tag{5.29}$$

  and the corresponding number i of first–round winners is $i = k-l+m$, hence, $p_{i|a,k,b,l}$ for the case of $a = 0$ and $k \neq l$ is obtained by replacing m in Eq. (5.29) with $i-k+l$.

Scenario II: $k = l \neq N$

In this scenario, at the end of time slot $t-1$ the following groups of nodes can be distinguished:

A) $N-k$ nodes have lost round 1 and, hence, will have $myback > 0$. The maximum value of $myback$ for this node is $S-b-1$, according to line 9 in Algorithm 2 and taking into account that $minback = b$ (lowest–frequency SC at round 1 in time slot $t-1$).

B) k nodes have won round 2 and, after transmitting, will have *myback* randomly varying between 0 and S-1 (lines 18 and 2 in Algorithm 2).

The case $a = 0$ is trivial, since the maximum number of first round winners is k (analogously to scenario I) and the probability that i nodes out of k (group B) select $myback = 0$ (given that there is at least one node that selects it) is

$$\frac{\binom{k}{i}\left(\frac{1}{S}\right)^{i}\left(1-\frac{1}{S}\right)^{k-i}}{1-\left(1-\frac{1}{S}\right)^{k}} \tag{5.30}$$

Another trivial case is $a = S - b - 1$: in this situation, the $N - k$ nodes of group A all win the first round at t (hence $i \geq N - k$) and the probability that m nodes out of the remaining k (group B) select SC $S - b - 1$ is

$$\binom{k}{m}\left(\frac{1}{b+1}\right)^{m}\left(1-\frac{1}{b+1}\right)^{k-m} \tag{5.31}$$

with $i = N - k + m$.

The case of $0 < a < S - b - 1$, instead, is non–trivial, since the nodes from both groups can select a as a SC. In detail, the probability that n nodes from group A and $i - n$ nodes from group B select SC a (given that at least one node selects it) is given by

$$\frac{\left[\binom{N-k}{n}\left(\frac{1}{S-b-a}\right)^{n}\left(1-\frac{1}{S-b-a}\right)^{N-k-n}\right]\cdot\left[\binom{k}{i-n}\left(\frac{1}{S-a}\right)^{i-n}\left(1-\frac{1}{S-a}\right)^{k-i+n}\right]}{1-\left(1-\frac{1}{S-a}\right)^{k}\left(1-\frac{1}{S-b-a}\right)^{N-k}} \tag{5.32}$$

The expression in Eq. (5.32) has to be summed for all possible values of n, taking into account that $0 \leq n \leq i$ by definition, and also $n \leq N - k$ and $i - n \leq k$. Therefore, the probability that i nodes win the first round at time t when $k = l \neq N$ and $0 < a < S - b - 1$ is

$$\frac{\sum_{n=\max(i-k,0)}^{min(N-k,i)}\left[\binom{N-k}{n}\left(\frac{1}{S-b-a}\right)^{n}\left(1-\frac{1}{S-b-a}\right)^{N-k-n}\right]\cdot\left[\binom{k}{i-n}\left(\frac{1}{S-a}\right)^{i-n}\left(1-\frac{1}{S-a}\right)^{k-i+n}\right]}{1-\left(1-\frac{1}{S-a}\right)^{k}\left(1-\frac{1}{S-b-a}\right)^{N-k}} \tag{5.33}$$

Scenario III: $k = l = N$

In this scenario there is only one group of N nodes, which have all won the second round in time slot $t-1$ and hence can select *myback* in the range $[a, S-1]$. The probability that exactly i nodes select $myback = a$ (given that at least one selects it) is given by

$$\frac{\binom{N}{i}\left(\frac{1}{S-a}\right)^{i}\left(1-\frac{1}{S-a}\right)^{N-i}}{1-\left(1-\frac{1}{S-a}\right)^{N}} \tag{5.34}$$

Summing up all the scenarios, the following expression for $p_{i|a,k,b,l}$ is obtained

$$p_{i|a,k,b,l} = \begin{cases} \binom{l}{i-k+l}\left(\frac{1}{S}\right)^{i-k+l}\left(1-\frac{1}{S}\right)^{k-i} & \text{if } k \neq l, a = 0, k-l \leq i \leq k \\ \frac{\binom{k}{i}\left(\frac{1}{S}\right)^{i}\left(1-\frac{1}{S}\right)^{k-i}}{1-\left(1-\frac{1}{S}\right)^{k}} & \text{if } k = l \neq N, a = 0, i \leq k \\ \binom{k}{i-N+k}\left(\frac{1}{b+1}\right)^{i-N+k}\left(1-\frac{1}{b+1}\right)^{N-i} & \text{if } k = l \neq N, a = S-b-1, i \geq N-k \\ \frac{\left[\binom{N-k}{n}\left(\frac{1}{S-b-a}\right)^{n}\left(1-\frac{1}{S-b-a}\right)^{N-k-n}\right]\cdot\left[\binom{k}{i-n}\left(\frac{1}{S-a}\right)^{i-n}\left(1-\frac{1}{S-a}\right)^{k-i+n}\right]}{1-\left(1-\frac{1}{S-a}\right)^{k}\left(1-\frac{1}{S-b-a}\right)^{N-k}} & \text{if } k = l \neq N, 0 < a < S-b-1 \\ \frac{\binom{N}{i}\left(\frac{1}{S-a}\right)^{i}\left(1-\frac{1}{S-a}\right)^{N-i}}{1-\left(1-\frac{1}{S-a}\right)^{N}} & \text{if } k = l = N \\ 0 & \text{otherwise} \end{cases} \tag{5.35}$$

Derivation of $\mathbf{p_{a|k,b,l}}$ To compute the third and last term, the probability that SC a is the lowest–frequency one at the first contention round during time slot t, given that, in the previous time slot, k nodes won the first round (with SC b) and l nodes won the second round must be derived. The same three scenarios considered in the previous derivation, with the same groups of nodes, are taken into account.

Scenario I: $k \neq l$

Tthe only possible value for a in this scenario is 0, hence $p_{0|k,b,l} = 1$ and 0 otherwise.

Scenario II: $k = l \neq N$

The case $a = 0$ is trivial, since it can only happen for the k nodes of group B and it happens with probability

$$1-\left(1-\frac{1}{S}\right)^{k} \tag{5.36}$$

The case $a > 0$, instead, is non–trivial. Moreover, as discussed previously, $a \leq S-b-1$. Let the random variable $X_i, i = 1, \ldots, S-b-1$ denote the number of nodes in group A that select SC i at the first round and $Y_j, j = 0, \ldots, S-1$ the number of nodes in group B that select SC j at the first round. Both these groups

of random variables follow a multinomial distribution with constant probabilities $p_i = \frac{1}{S-b-1}$ for the first group and $p_j = \frac{1}{S}$ for the second group. The probability that a is the lowest–frequency SC is expressed as

$$p_{a|k,b,l} = p_{X,a} \cdot p_{Y,a} + p_{X,a} \cdot p_{Y,\bar{a}} + p_{X,\bar{a}} \cdot p_{Y,a} \tag{5.37}$$

where:

$$\begin{aligned} p_{X,a} &= P\{X_1 = 0, \ldots, X_{a-1} = 0, X_a \neq 0\} \\ &= P\{X_1 = 0, \ldots, X_{a-1} = 0\} - P\{X_1 = 0, \ldots, X_a = 0\} \\ &= \left(1 - \frac{a-1}{S-b-1}\right)^{N-k} - \left(1 - \frac{a}{S-b-a}\right)^{N-k} \end{aligned} \tag{5.38}$$

$$\begin{aligned} p_{X,\bar{a}} &= P\{X_1 = 0, \ldots, X_a = 0\} \\ &= \left(1 - \frac{a}{S-b-a}\right)^{N-k} \end{aligned} \tag{5.39}$$

$$\begin{aligned} p_{Y,a} &= P\{Y_0 = 0, \ldots, Y_{a-1} = 0, Y_a \neq 0\} \\ &= P\{Y_0 = 0, \ldots, Y_{a-1} = 0\} - P\{Y_1 = 0, \ldots, Y_a = 0\} \\ &= \left(1 - \frac{a}{S}\right)^{k} - \left(1 - \frac{a+1}{S}\right)^{k} \end{aligned} \tag{5.40}$$

$$\begin{aligned} p_{Y,\bar{a}} &= P\{Y_0 = 0, \ldots, Y_a = 0\} \\ &= \left(1 - \frac{a+1}{S}\right)^{k} \end{aligned} \tag{5.41}$$

The expression for $p_{a|k,b,l}$ when $k = l \neq N$ and $a > 0$ can hence be obtained by inserting Eq. (5.38), (5.39), (5.40) and (5.41) in Eq. (5.37)

$$\left(1 - \frac{a}{S}\right)^{k} \left(1 - \frac{a-1}{S-b-1}\right)^{N-k} - \left(1 - \frac{a+1}{S}\right)^{k} \left(1 - \frac{a}{S-b-1}\right)^{N-k} \tag{5.42}$$

Scenario III: $k = l = N$

In this scenario there is only one group of N nodes, which have all won the second round at time slot $t-1$ and hence can select *myback* in the range $[0, S-1]$. Let the random variable $Z_i, i = 0, \ldots, S-1$ denote the number of nodes that select SC i at the first round, multinomially distributed with constant probability $p_i = \frac{1}{S}$.

The probability that a is the lowest–frequency SC is expressed as

$$\begin{aligned} p_{a|k,b,l} &= P\{Z_0 = 0, \ldots, Z_{a-1} = 0, Z_a \neq 0\} \\ &= P\{Z_0 = 0, \ldots, Z_{a-1} = 0\} - P\{Z_0 = 0, \ldots, Z_a = 0\} \\ &= \left(1 - \frac{a}{S}\right)^N - \left(1 - \frac{a+1}{S}\right)^N \end{aligned} \tag{5.43}$$

Summing up all the scenarios, the following expression for $p_{a|k,b,l}$ is obtained

$$p_{a|k,b,l} = \begin{cases} 1 & \text{if } k \neq l, a = 0 \\ 1 - \left(1 - \frac{1}{S}\right)^k & \text{if } k = l \neq N, a = 0 \\ \left(1 - \frac{a}{S}\right)^k \left(1 - \frac{a-1}{S-b-1}\right)^{N-k} - \left(1 - \frac{a+1}{S}\right)^k \left(1 - \frac{a}{S-b-1}\right)^{N-k} & \text{if } k = l \neq N, 0 < a \leq S-b-1 \\ \left(1 - \frac{a}{S}\right)^N - \left(1 - \frac{a+1}{S}\right)^N & \text{if } k = l = N \\ 0 & \text{otherwise} \end{cases} \tag{5.44}$$

Computation of the saturation throughput Taking into account the three–dimensional process $\{x(t), c(t), y(t)\}$, a collision in a time slot can happen only if two or more nodes win the second contention round, i.e., if $y(t) > 1$. The success probability can hence be computed as

$$P_s = \sum_{i=1}^{N} \sum_{a=0}^{S-1} \pi_{i,a,1} \tag{5.45}$$

Once this probability is obtained, the saturation throughput is given by

$$\eta_{B2F} = \frac{P_s T_d}{P_s T_S + (1 - P_s) T_C} \tag{5.46}$$

Considering the structure of the BACK2F protocol, the values of T_S and T_C are equal to

$$T_S = T_{difs} + 2T_{round} + T_d + T_{sifs} + T_{ack} + 2T_p \tag{5.47}$$

$$T_C = T_{difs} + 2T_{round} + T_d + T_p \tag{5.48}$$

where T_{round} is the duration of a contention round in the frequency domain, reported in Tab. 5.2.

Analysis for RCFD

In the RCFD protocol, similarly to what happens in BACK2F, there are no idle slots. Moreover, the RTS/CTS exchange in the frequency domain prevents any possibility of collision. As a consequence, $P_{tr} = 1$ and $P_s = P_{s,hd} + P_{s,fd} = 1$, where

$$P_{s,hd} = 1 - \frac{1}{N-1}, \quad P_{s,fd} = \frac{1}{N-1} \tag{5.49}$$

The saturation throughput hence becomes

$$\eta_{RCFD} = \frac{T_d P_{s,hd} + 2T_d P_{s,fd}}{T_S} \tag{5.50}$$

where, in this case

$$T_S = T_{difs} + 3T_{round} + T_h + T_d + T_{sifs} + T_{ack} + 2T_p \tag{5.51}$$

In both this analysis and the one of BACK2F the scanning time T_{scan} used in Eq. (5.2) is equal to T_{difs}, to provide a fair comparison among all the MAC protocols.

Numerical results

The saturation throughput has been theoretically derived for the different MAC protocols as a function of several system parameters.This metric is now numerically evaluated for different network configurations and system parameters. Tab. 5.2 reports the simulation parameters in this evaluation, which are adopted from the IEEE 802.11g standard (IEEE 802.11-2016).

Fig. 5.8 shows the saturation throughput for all MAC algorithms versus the number of nodes in the network. The payload length has been kept fixed at $L = 1000$ Bytes, while the transmission rate is $R = 6$ Mbps, yielding a data transmission time of roughly $T_d = 1.4$ ms. It can be observed that the RCFD strategy outperforms all other MAC algorithms for any number of nodes. The two schemes that consider FD transmissions (RCFD and FD MAC) are able to provide a normalized throughput higher than one, for a small number of nodes. BACK2F and IEEE 802.11 RTS/CTS do not show a significant variation with the number of nodes, with the first one providing a higher throughput (close to 1) and performing close to RCFD for a large number of nodes. The standard IEEE 802.11 DCF provides the worst performance, strongly affected by the number of nodes, as expected.

It is worth noting that the sharp decrease in throughput presented by FD–capable

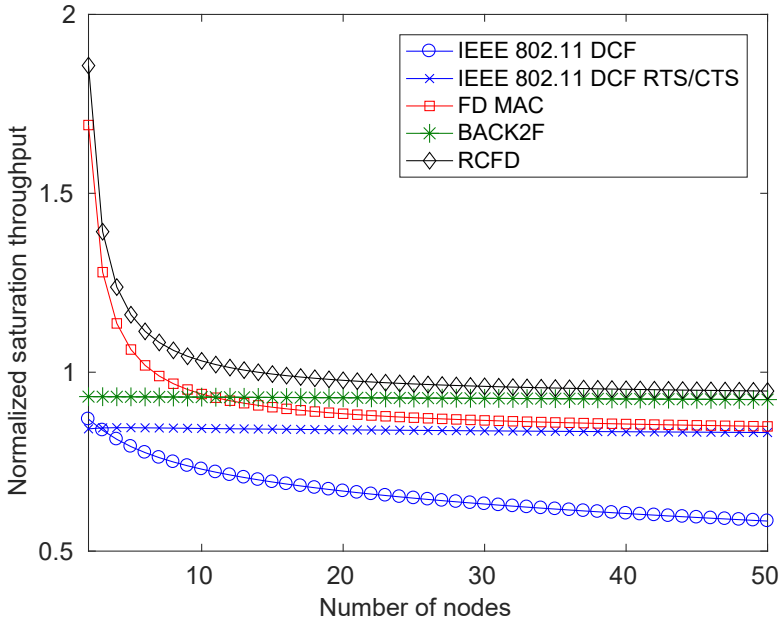


Figure 5.8: Theoretical saturation throughput versus number of nodes, with $L = 1000$ bytes long packets and $R = 6$ Mbps data rate.

MAC protocols (RCFD and FD MAC) is due to the FIFO assumption. Indeed, in both cases, assuming that a node n_i gets the channel, a FD transmission happens only if the packet at the head of the queue of the receiver n_j is destined to n_i, which happens with probability $1/(N-1)$. The throughput curves for these algorithm, hence, follow a hyperbolic shape. The FIFO assumption was considered in this analysis for the sake of tractability and will be relaxed in the following simulations.

Another evaluation is reported in Fig. 5.9, where the number of nodes and the data rate are fixed at $N = 10$ and $R = 6$ Mbps, respectively, and the payload length L varies from 100 to 2300 bytes. Again, the proposed RCFD technique provides the best performance for all possible payload sizes. The techniques based on time domain RTS/CTS (IEEE 802.11 and FD MAC) perform very poorly for short packets, since in that case the overhead represented by the exchange of RTS and CTS frames has a very significant impact. The techniques that include frequency–based contention (RCFD and BACK2F) are characterized by a similar trend, even if the first one always provides a higher throughput, thanks to its FD capabilities. The standard IEEE 802.11 DCF without RTS/CTS, finally, yields the worst results, since it clearly suffers from the occurrence of collisions.

In order to make an assessment of the numerical results based on the theoretical models presented in this section, a set of network simulations have been performed using

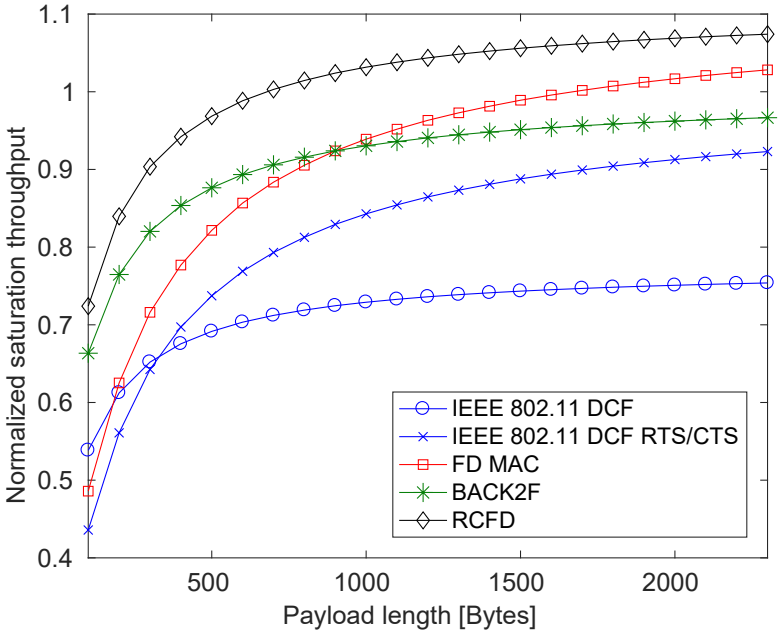


Figure 5.9: Theoretical saturation throughput versus packet payload size for a network with $N = 10$ nodes using an $R = 6$ Mbps data rate.

Table 5.3: Comparison of normalized saturation throughput in analysis and simulations for FD MAC, BACK2F and RCFD channel access schemes

Algorithm	**N = 2**	**N = 10**	**N = 20**	**N = 50**
FD MAC Analysis	1.6908	0.9390	0.8840	0.8485
FD MAC Simulations	1.4281	0.8458	0.7929	0.7377
BACK2F Analysis	0.9319	0.9304	0.9287	0.9235
BACK2F Simulations	0.9349	0.9305	0.9301	0.9253
RCFD Analysis	1.8570	1.0316	0.9773	0.9474
RCFD Simulations	1.8514	0.9306	0.9301	0.9300

the ns3 platform (ns3), configured according to the following assumptions:

- N nodes are randomly deployed in the same collision domain;
- each node randomly generates packets for every other node in the network and the transmission queue (which follows a FIFO behavior) is always saturated;
- the communication channel is ideal, with collisions being the only source of errors;
- the values of transmission rate ($R = 6$ Mbps) and payload size ($L = 1000$ bytes) are fixed.

The results, which refer to the simulation throughput averaged over 10 different simulation runs, are reported in Tab. 5.3, where they are compared with the numerical values of Fig. 5.8. It can be observed that the results of the analysis and simulations are close. Moreover, the simulations confirm that RCFD outperforms the other channel access schemes for any network size, as the analysis suggested.

Simulation assessment

The results of the theoretical analysis show a clear prevalence of the proposed RCFD algorithm over other MAC layer schemes considered. However, the analysis and simulations were conducted under some possibly limiting assumptions, the most important one being that all nodes are within the same collision domain. In order to relax this assumption, the five aforementioned MAC strategies have been compared through ns3, for the case of a wireless network with multiple collision domains.

Simulations setup

The standard distribution of ns3 already contains models for the IEEE 802.11 DCF, both with and without RTS/CTS, as defined in the standard. However, the modules for the MAC algorithms proposed in the literature, namely FD MAC, BACK2F and the proposed RCFD, were not available and therefore had to be purposely developed. Moreover, the standard ns3 *wifi* module only allows half–duplex communications, preventing a node from transmitting if it is receiving. In order to be able to simulate a network with full–duplex nodes, the patch discussed in (Zhou and Srinivasan, 2014) was adopted, which allows to simulate an FD wireless network with ns3. It is worth stressing that, for the algorithms based on frequency domain operations (BACK2F and RCFD), the exchange of data over OFDM subcarriers during the contention rounds is assumed to be ideal, i.e., when a node transmits on a subcarrier all the other nodes in its collision domain are able to detect it.

Two different scenarios have been simulated: a *structured scenario* and a *random scenario*, described in detail in the following.

Setup for the structured scenario The simulated network for the structured scenario is depicted in Fig. 5.10. It is an ad hoc wireless network composed of fixed nodes placed on a grid. The distance between two adjacent nodes in the same row or column is d. The coverage range of each node is a circle of radius $r = d\sqrt{2}$ and, hence, includes all its one–hop neighbors. Within this area, the node can transmit and receive packets as

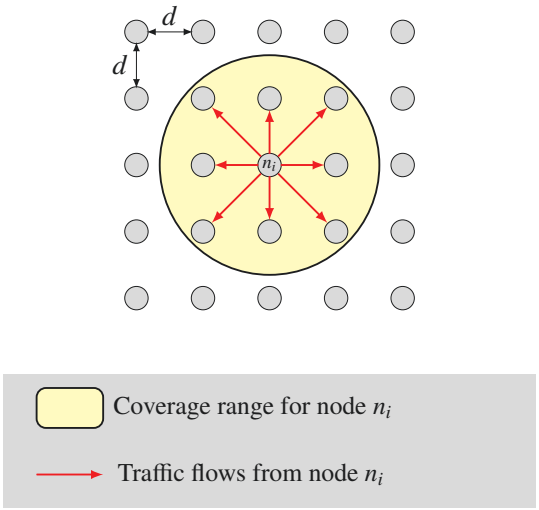


Figure 5.10: Simulated network for the structured scenario.

well as overhear transmissions. To implement this channel model, the *RangePropagationLossModel* of ns3 has been adopted, combined with a purposely implemented error model. According to these models, a transmission between two nodes is successful only if the distance is below r and there is no collision, and it fails with probability 1 otherwise (regardless of the adopted transmission rate). In this way, the impact of collisions on the network can be accurately analyzed for the different channel access strategies, isolating it from all the other factors that can affect the performance, such as path loss, fading, performance of different modulation and coding schemes, etc.

The total number of nodes in the network is $N = g^2$, where g is the grid size, and simulations have been conducted for several values of g.

Setup for the random scenario In the random scenario, N nodes are randomly deployed within a square of size l. The coverage range r of a node is determined as the maximum range which allows a success transmission probability above 90% for a packet of size L transmitted with rate R and assuming no fading.

The channel model used in this scenario combines the *LogDistancePropagationLossModel* for path loss and the *NakagamiPropagationLossModel* to emulate Rayleigh fading. The *NistErrorRateModel* validated in Pei and Henderson (2010) was adopted, that takes into account the different robustness levels of each modulation and coding scheme.

The goal of the random scenario is to investigate how the RCFD algorithm would perform in a more realistic ad hoc wireless network in comparison to the other channel

access techniques.

Traffic model and metrics for both scenarios In each node, several applications are installed, one for each node within its coverage range, as shown in Fig. 5.10 for the structured scenario. The starting time of each application, t_s, is distributed as an exponential random variable of parameter λ_s truncated after $t_{s,max}$, while the stop time coincides with the end of the simulation.

An *OnOffApplication* model is adopted where the duration of the ON and OFF periods are also exponentially distributed, with mean T_{ON} and T_{OFF}, respectively. During the ON period, the applications generates Constant Bitrate (CBR) traffic with source rate R_s. All packets have the same length L and the data rate at the physical layer, R, is constant.

Network operations have been simulated for a total of T seconds (with the initial transient period removed), for different values of the network size N. Given a certain parameter configuration, each simulation has been repeated a total of N_S times and results have been averaged.

Two performance metrics were considered, namely the *normalized system throughput*, Γ, and the *average delay*, Δ. The normalized system throughput is the ratio of the total number of payload bits successfully delivered by all the nodes in the network over the simulation time T, and the offered traffic G. The offered traffic is given by

$$G = R_s \cdot N_a \cdot \frac{T_{ON}}{T_{ON} + T_{OFF}} \tag{5.52}$$

where N_a is the total number of running applications in the network, which is a function of the network size N and the coverage radius r.

The average delay, on the other hand, is the arithmetic mean of the delay experienced by each packet in the network, defined as the time elapsed from the instant in which the packet is generated by the application to the instant in which the packet is successfully delivered or discarded.[4]

Tab. 5.4 reports all the parameters adopted in the simulations.

It is worth noting that the simulation–based results are complementary with respect to those derived from the theoretical analysis, since the latter were based on the assumption of a single collision domain, whereas the former take into account multiple collision domains.

[4]A packet is discarded in three cases: (1) the transmission keeps failing after $N_{tx,max}$ transmission attempts; (2) the packet transmission queue has exceeded the maximum size Q_{max}; (3) the time elapsed from the packet generation has exceeded the threshold Δ_{max}.

Table 5.4: Simulation parameters

Parameter	Description	Value
d	Distance between two adjacent nodes in the structured scenario	100 m
l	Side of deployment area in the random scenario	500 m
λ_s	Parameter of application starting time	0.5 s^{-1}
$t_{s,max}$	Maximum application starting time	5 s
T_{ON}	Average time during which each application is ON	0.1 s
T_{OFF}	Average time during which each application is OFF	0.1 s
R_s	Application source rate during the ON period	1 Mbit/s
T	Duration of each simulation	20 s
$N_{tx,max}$	Maximum number of retransmissions at the MAC layer	7
Q_{max}	Transmission queue size (packets)	1000
Δ_{max}	Maximum interval after which a packet is discarded	1 s
L	Payload length for packets	$\{200, 500, 1000\}$ bytes
R	Data rate at the PHY layer	$\{6, 18, 54\}$ Mbit/s
N_S	Number of simulations for each configuration	10

Simulation results for the structured scenario

In order to provide a comprehensive assessment of the presented protocol, in the network simulations its performance in the structured scenario have been evaluated for two opposite cases:

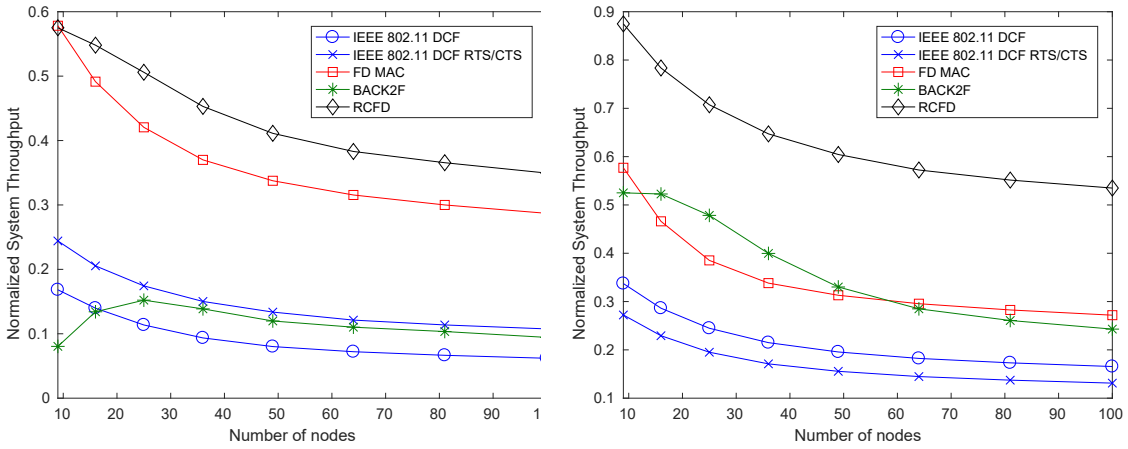
I. *Long packet transmission time:* in this case large payload packets ($L = 1000$ Bytes) were exchanged at the lowest possible rate provided by IEEE 802.11g, namely $R = 6$ Mbit/s, resulting in a very long packet transmission time.

II. *Short packet transmission time:* in this case small payload packets ($L = 200$ Bytes) were exchanged at the highest possible rate, namely $R = 54$ Mbit/s, with a corresponding short packet transmission time.

In each case, the aforementioned performance metrics for the considered MAC algorithms have been evaluated for different values of the grid size parameter g, ranging from 3 ($N = 9$ nodes) to 10 ($N = 100$ nodes).

Fig. 5.11a shows the normalized system throughput Γ for case I. The RCFD strategy outperforms the other MAC protocols for any network size. The FD MAC algorithm is able to achieve similar performance when the number of nodes is small, but its throughput significantly degrades as the network size increases. The BACK2F protocol presents a significantly lower Γ, due to its difficulties in handling multiple collision domains. Finally, the IEEE 802.11 strategies based on time domain channel contention perform poorly.

The same metric Γ is reported in Fig. 5.11b for the second case. Again, RCFD performs much better than all other strategies. It can be observed, in particular, that the schemes relying upon the exchange of RTS/CTS frames (FD MAC and IEEE 802.11) perform much worse than in the previous case, since these frames represent a significant overhead, given the lower time needed for the actual transmission of data frames. BACK2F, which



(a) Case I: $R = 6$ Mbit/s, $L = 1000$ bytes

(b) Case II: $R = 54$ Mbit/s, $L = 200$ bytes

Figure 5.11: Simulated normalized system throughput Γ for the structured scenario.

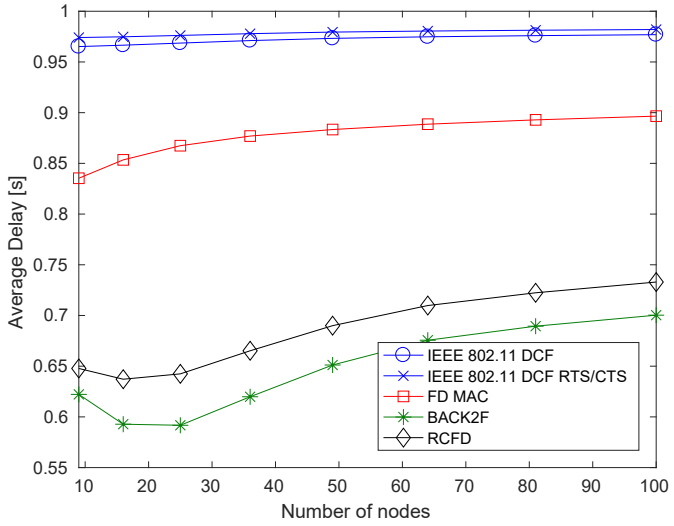
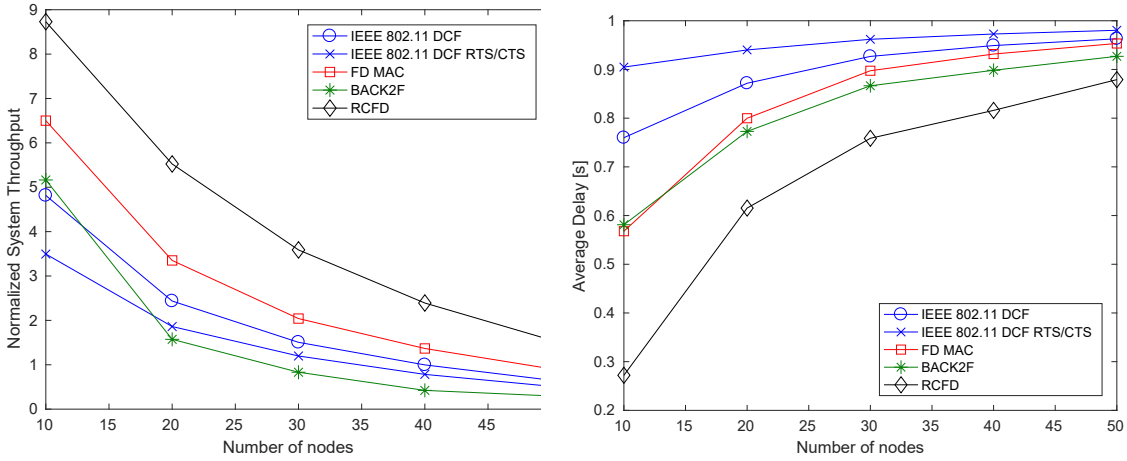


Figure 5.12: Simulated average delay Δ for the structured scenario, case II ($R = 54$ Mbit/s, $L = 200$ bytes).

instead relies on frequency domain contention as RCFD, performs much better than in the previous case, reaching similar performance as FD MAC, despite not being a full–duplex MAC protocol.

It is worth noticing that the normalized throughput values are higher in Fig. 5.11b with respect to Fig. 5.11a. Indeed, the higher PHY rate allows to exchange an increased amount of data in the same time.

The average delay Δ simulated in case II for all the MAC protocols is shown in



(a) Normalized system throughput Γ **(b)** Average delay Δ

Figure 5.13: Simulated results for the random scenario ($R = 18$ Mbit/s, $L = 1000$ Bytes).

Fig. 5.12. The strategies that include frequency domain channel contention strongly outperform those based on a time domain approach. In particular, BACK2F slightly outperforms the RCFD strategy, mostly due to the lower number of contention rounds in the frequency domain (2 against 3).

Simulation results for the random scenario

In the random scenario, the performance of the considered MAC algorithms have been evaluated for different network size values, ranging from $N = 10$ to $N = 50$ nodes. The payload size has been fixed to $L = 500$ bytes and the PHY layer transmission rate to $R = 18$ Mbps, providing an intermediate case between the two extremes analyzed in the structured scenario. Under this configuration, the coverage radius of each node was set to $r = 60$ m in order to provide 90% transmission success probability.

Fig. 5.13a shows the normalized system throughput Γ for the different MAC algorithms. Also in this case, RCFD is able to significantly outperform all the other schemes. As in case I of the structured scenario, FD MAC provides the closest performance, while the throughput of the BACK2F algorithm suffers from the presence of multiple collision domains and significantly degrades with the network size.

The average delay Δ for the random scenario is reported in Fig. 5.13b. Again, RCFD significantly outperforms all the other schemes, confirming that this strategy represents a very interesting opportunity for real–world applications. Among the other algorithms, BACK2F emerges as the one able to guarantee the lowest delay, thanks to the channel contention in the frequency domain.

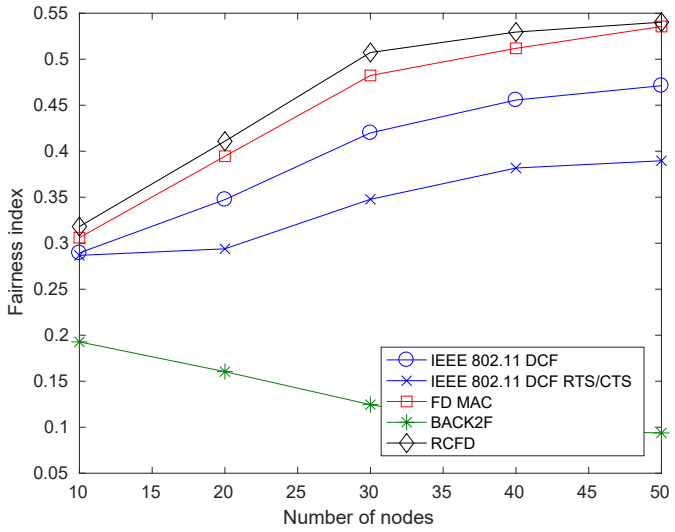


Figure 5.14: Simulated fairness index J for the random scenario ($R = 18$ Mbit/s, $L = 1000$ Bytes).

In order to provide a final insight, the fairness of the compared MAC protocols is reported in Fig. 5.14 for the random scenario, measured in terms of Jain's fairness index (Jain et al., 1984), defined as

$$\mathcal{J}(p_1, \ldots, p_N) = \frac{\left(\sum_{i=1}^{N} p_i\right)^2}{N \cdot \sum_{i=1}^{N} p_i^2} \tag{5.53}$$

where p_i is the number of packets successfully received by node n_i. It can be observed that, also in terms of fairness, RCFD outperforms all other protocols.

Conclusions

The currently employed channel access schemes for wireless networks present several issues and relatively low performance. The introduction of full–duplex wireless communication can lead to increased performance but also poses additional challenges to transmission scheduling, and no standard MAC protocol has emerged so far as the best solution for FD wireless networks. The proposed RCFD, a full–duplex MAC protocol based on a time–frequency channel access procedure, addresses these issues. Theoretical analyses and network simulations show that this strategy provides excellent performance in terms of both throughput and packet transmission delay, also in the case of dense networks, compared to other standard and state–of–the–art MAC layer schemes.

5.3 Considerations on industrial full–duplex networks

The possibilities offered by FD wireless, as explained in Sec. 5.1, can be profitably exploited not only for home/office communications but also for industrial communication networks, which represent the topic of this thesis. The benefits of employing FD–capable devices in an industrial context do not limit to the doubling of network capacity and can lead to many different performance improvements according to the traffic profile.

As a first example, consider *cyclic industrial traffic*, where packets are exchanged periodically between a central master node (typically, the controller in a NCS) and several distributed slave nodes (tipically, sensors and actuators). In the simplest case in which all the nodes are FD and each slave need to both receive and send data to the master, bidirectional FD communication can be established, hence allowing a reduction in the data exchange time and an ultimate reduction of cycle time. A relay FD configuration can instead be adopted if only the master is FD–capable and/or the slaves need only to send or receive a packet. In this case, indeed, the master can send a packet to one slave while receiving a packet from another slave. A proper scheduling must ensure that the pair of slaves simultaneously active do not interfere with each other.

In the case of *acyclic industrial traffic*, where different slaves can access the channel randomly to send a packet, the RCFD protocol described in Sec. 5.2 can be adopted instead of other contention–based MAC protocols. Indeed, this protocol allows a fixed and short channel access time, favouring the low latency and determinism generally targeted in ICNs. Moreover, the protocol allows to solve the hidden terminal effect which can be detrimental also in industrial applications (Willig et al., 2005). The RCFD protocol can also be upgraded by considering the random selection of OFDM subcarriers in the first round as an implicit ordering of the nodes attempting to access the channel, hence allowing contention–free channel access. Finally, messages exchanged in an acyclic way are typically characterized by different priorities (e.g., different types of alarms and warnings), that can be ensured by appropriately restricting the set of SCs among which a node randomly selects during the first contention round of RCFD.

These examples suggest that FD wireless networks can be an interesting topic for industrial communications that should be investigated more in detail. Further advances in SIC methods are also important, since reliable and cost–effective FD implementations are required in order to ensure success in the competitive industrial market.

6

High–performance Wireless Networks for Control

A fundamental limitation of traditional wireless solutions for industrial control is that they are based on general purpose bottom layers. For example, WirelessHART (WirelessHART) and ISA 100.11a (ISA-100.11a-2009) are based on the IEEE 802.15.4 PHY layer, and hence they share the same bandwidth, modulation schemes, operating frequency, etc. Keeping the bottom layers of general purpose wireless standards can allow faster standardization and easier intercompatibility. However, it represents a fundamental bottleneck to network performance, which may not be good enough to cope with the most critical industrial control applications, such as mining, robotics and power systems (Pang et al., 2017).

In the light of these considerations, a novel approach is presented in this chapter, labeled as WirelessHP. This approach totally differs from the strategy traditionally adopted in the design of industrial wireless networks, in the sense that it proposes a completely customized protocol stack, where each layer is optimized towards the specific requirements of industrial control applications, rather than attempting to satisfy all generic requirements of different application scenarios.

This chapter is mainly based on the works in Luvisotto et al. (2017a), Luvisotto et al. (2017b) and Pang et al. (2017).

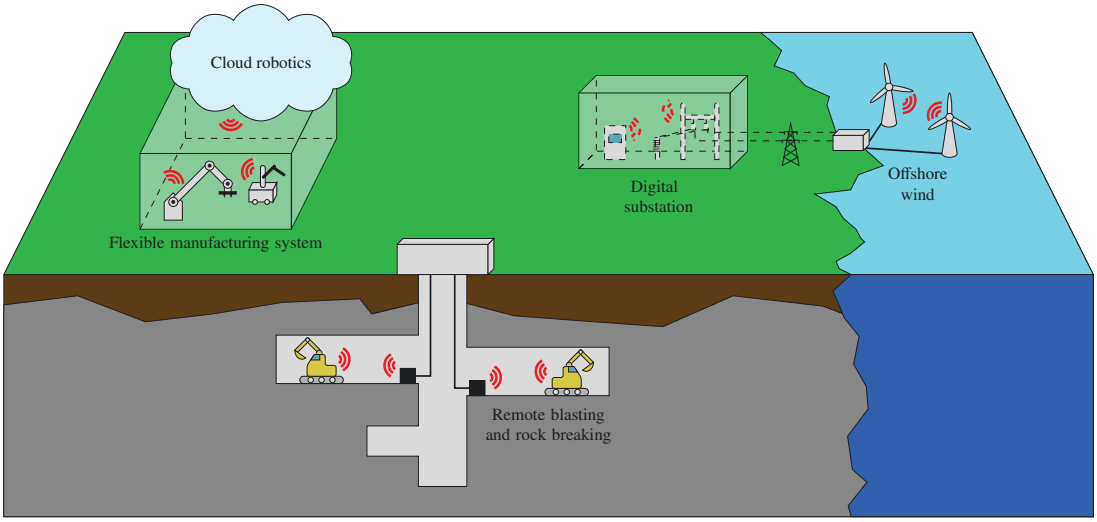


Figure 6.1: Some examples of critical NCSs with ultra-high performance requirements: (a) robotics and factory automation, (b) power systems automation, and (c) the mining industry.

6.1 Application scenarios and requirements

The need of customized networks for critical NCS stems from different industrial use cases characterized by ultra–high performance requirements, some of which are represented in Fig. 6.1.

A first possible application environment is represented by the mining sector, where remote blasting and rock breaking procedures are increasingly used to enhance performance and ensure the safety of workers (Mishra et al., 2017). Robotics and factory automation also include critical scenarios, such as Flexible Manufacturing Systems (FMSs), able to automatically adapt and react to changes in the environment, production flow and product types. FMSs will rely on the cooperation among intelligent robots, often mounted over Automatic Guided Vehicles (AGVs), and integrate cloud robotics, enabling the centralized management of distributed resources (Xu, 2012). Power systems automation also presents demanding use cases, both in power distribution (e.g., digital substations in the smart grids (Parikh et al., 2013)) and in power generation (e.g., integration of wind parks). Finally, the synchronized control of complex devices employed in power electronics applications, such as multilevel converters (Toh and Norum, 2013), also require deterministic and fast data exchange between multiple elements.

These scenarios represent some examples of applications in which ultra–high performance networks are required, and are generally served by fast and robust Ethernet networks based on optical fiber media. In order to replace some of these wired links, each

Table 6.1: Example of system–level requirements for different industrial communications scenarios (10^2 bits packets)

Scenario	# of nodes	Update rate	Goodput	System range
BA	10^2-10^3	10^{-1} Hz	10^3-10^4 bps	10^1-10^2 m
PA	10^2-10^3	10^1 Hz	10^5-10^6 bps	10^1-10^2 m
FA	10^2-10^3	10^3 Hz	10^7-10^8 bps	10^1-10^2 m
PSA	10^1-10^2	10^4 Hz	10^7-10^8 bps	10^2-10^3 m
PEC	10^2-10^3	10^5 Hz	10^9-10^{10} bps	10^1-10^2 m

layer of the WirelessHP protocol stack needs to be designed carefully. In the following, a detailed breakdown of the application–level requirements and the kind of performance that they impose on wireless links is given.

Requirements from the application perspective

A set of quantitative performance metrics which reflects the requirements of critical industrial control scenarios can be defined. They include the update rate (or cycle time), the number of nodes, the goodput and the communication range. The first metric describes how often an actuator/sensor receives updated commands from and/or sends new sensing data to the central controller and represents a fundamental parameter for the design of the industrial control application. The second metric describes the network size, i.e., the number of sensors/actuators that are attached to a single controller. The goodput expresses how many information bits are globally exchanged over time between the nodes and the controller. Finally, the communication range determines the area to be covered by the wireless network.

Tab. 6.1, which represents an extended and more quantitative version of Tab. 2.1, proposes some typical range of values for these metrics in different industrial scenarios. It can be observed that update rate and goodput increase considerably moving from the simplest scenarios (those related to BA) to the most challenging ones (PSA and PEC). Conversely, number of nodes and system range are similar in different scenarios, with the exception of PSA applications, that involve a smaller number of nodes covering a wider area. The WirelessHP scenario deals with the most critical industrial scenarios, i.e., those characterized by the highest values of update rate and throughput, such as the most advanced FA, PSA and PEC applications.

Finally, together with the four aforementioned metrics, a WirelessHP system should also guarantee a high reliability level. From an application perspective, a PER around 10^{-9} is perceived as tolerable (Gerlach-Erhardt, 2009).

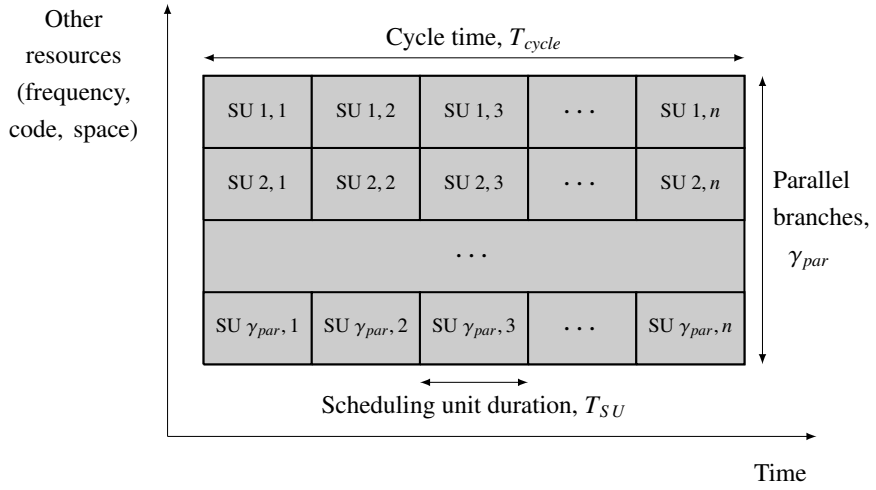


Figure 6.2: Cyclic communication schedule over parallel branches.

Link–level design and the bottleneck of timing

The metrics presented in Tab. 6.1 describe the performance of a WirelessHP network from the perspective of the control application and are relevant to the engineers who deploy it in specific applications. However, they are still too generic for guiding the design of the physical and data–link layers. Consequently, requirements on the link–level, i.e., considering the Point–to–point (P2P) link between two nodes, are derived. The rationale is that, if the links are able to guarantee the required link–level performance, the system–level behavior will meet the overall requirements of the applications.

In order to clarify this aspect, consider the timing analysis of a WirelessHP network that should be employed in one of the scenarios reported in Tab. 6.1. According to the considerations in Chap. 2, a star topology network is considered, whose operations are cyclic, i.e., the nodes send data to and/or receive data from the controller in a specified order which is repeated over time. The update rate indicated as a fundamental performance metric in Tab. 6.1 can be computed as the inverse of the minimum cycle time, T_{MCT}, i.e., the minimum time required for the controller to communicate to every node in the network.

The most important quantity to be defined in order to compute the minimum cycle time is the Scheduling Unit (SU) T_{SU}, defined as the minimum time required for the unidirectional transmission of a fixed amount of data over a wireless link. This time includes any hardware processing delay, the time required to access the wireless channel, the time to actually transmit the packet and the time to receive the ACK, if necessary,

i.e, $T_{SU} = T_{proc} + T_{access} + T_{TX} + T_{ACK}$. In order to complete a cycle, a number of scheduling units at least equal to the number of nodes in the network, N_{nodes}, must be scheduled. However, in order to increase the reliability of the system, the same data can be mapped to multiple scheduling units, thus increasing the cycle time of a factor k_{red} (redundancy level). Finally, multiple scheduling units can be transmitted in parallel, assuming the adoption of a TDMA scheme on γ_{par} different space and/or frequency resources, as represented in Fig. 6.2.

Taking all these factors into account, the minimum cycle time is given by

$$T_{MCT} = \frac{k_{red} \cdot N_{nodes} \cdot T_{SU}}{\gamma_{par}} \tag{6.1}$$

From this analysis it appears that in order to design a network which satisfies specific system–level requirements, such as update rate and number of nodes, the designer should reduce as much as possible the duration of the scheduling unit and provide the highest possible degree of parallelization and the lowest redundancy level which allows to achieve the desired reliability. Link–level performance, and in particular the scheduling unit, hence represent the bottleneck in the realization of high–performance wireless solutions.

Derived link–level expectations

Following the aforementioned discussion, three relevant link–level metrics can be outlined:

1. *SU* - The minimum time required for the transmission of a fixed amount of a data.

2. *Data rate* - The number of information bits that a single link can carry over a unit of time.

3. *Link range* - The maximum distance between two nodes over which reliable communication is feasible. This value is generally lower than the system range presented in Tab. 6.1. It is easier to achieve high data rate, low latency and high reliability on a short–range link and deal with applications that require large coverage, such as PSA, by considering multi–hop communications (even though the overall latency will be increased).

In order to provide clear targets for WirelessHP systems and to easily measure the gap between currently available wireless technologies and the desired performance, a specific sets of link–level requirements for two different scenarios, baseline and target, are considered here and shown in Tab. 6.2. The baseline requirements can lead to satisfactory performance in most of the use cases of the WirelessHP scenario, whereas the target

Table 6.2: Link–level requirements for WirelessHP scenarios (10^2 bits packets)

Scenario	Data rate	SU	Link range
Baseline	500 Mbps	500 ns	3 m
Target	2 Gbps	200 ns	10 m

requirements would allow to match the desired performance also in very extreme use cases, such as the most advanced PEC applications.

Those requirements have been computed starting from the system–level performance metrics reported in Tab. 6.1 for specific use cases. In particular, the SU for the baseline scenario (500 ns) has been derived from Eq. (6.1) by considering $N_{nodes} = 50$ nodes, a redundancy level $k_{red} = 8$ and a parallelization level $\gamma_{par} = 2$, thus yielding a target cycle time of 100 μs (update rate of 10^4 Hz), which may be representative of a typical PSA deployment. With the chosen redundancy level, a link PER of 10^{-1} at PHY layer is sufficient to guarantee 10^{-8} application layer PER.[1] In contrast, the scheduling unit for the target scenario (200 ns) would be suitable for an advanced PEC application, which requires a cycle time of 10 μs for a network of 100 nodes, achieved with a redundancy level $k_{red} = 2$ and a parallelization level $\gamma_{par} = 4$. In this case, in order to achieve 10^{-8} application layer PER, the link PER at the PHY layer should be 10^{-4}.

6.2 Analysis of the state–of–the–art

The extremely high network performance required by the most critical industrial control scenario cannot be adequately addressed by currently available wireless standards, that are conceived for home/office data exchange and, hence, are not optimized for the industrial traffic. In this section, the performance and limitations of the most advanced wireless standards are reviewed. Subsequently, some emerging research trends that can be of interest for the design of WirelessHP networks are analyzed.

Review of high–performance wireless standards

The suitability of the more advanced standards for high–performance wireless communications in a WirelessHP scenario is assessed here. Since the focus is on high–performance wireless, low–rate standards such as IEEE 802.15.4 or IEEE 802.15.1, will not be covered.

[1] This assumes that the different redundant transmissions experience a fully independent channel realization, which is often not the case in practical deployments. To overcome this issue and still guarantee high reliability at application layer, the link PER should be increased with respect to the theoretical value.

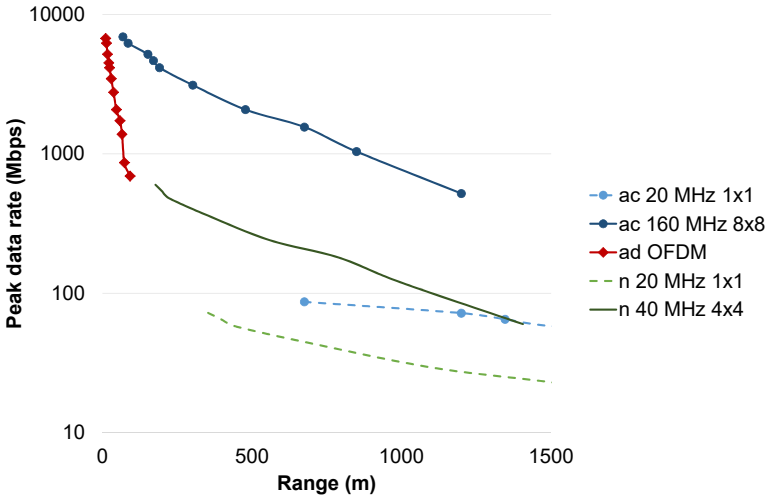


Figure 6.3: Data rate versus free–space range for high–performance WLAN standards (assuming maximum EIRP is used according to US regulations).

For the same reason, IWSN standards such as WirelessHART, WIA-PA and WISA are not discussed here.

WLAN standards

The IEEE 802.11 standard for WLANs includes several amendments, which are described in detail in Sec. 3.1. Here the focus is restricted only on the most recently approved amendments that brought a significant performance improvement, namely IEEE 802.11n, IEEE 802.11ac and IEEE 802.11ad.

Fig. 6.3 shows the theoretical free–space link range of WLAN standards versus their data rates. For IEEE 802.11n and IEEE 802.11ac, since they have several PHY layer configurations, two extreme cases were considered. To calculate the range, the receiver sensitivity R was first considered, as reported in the standard for each modulation and coding scheme (IEEE 802.11-2016). R is the minimum received power that allows a PER lower than 1% for a frame of 4096 bytes assuming 5 dB implementation loss and 10 dB noise figure. The range d was then derived from the following equation

$$d^{\alpha} = \frac{P_{tx} \cdot G_{tx} \cdot G_{rx} \cdot \lambda^2}{R \cdot F \cdot 16\pi^2} \tag{6.2}$$

assuming maximum EIRP $P_{tx} \cdot G_{tx}$,[2] 10 dBi receive gain G_{rx}, a path loss exponent α of 2,

[2] EIRP limits for the US were used, i.e. 36 dBm at 2.4 GHz, 53 dBm at 5 GHz and 43 dBm at 60 GHz. Limitations for Europe are significantly different, namely 20 dBm at 2.4 GHz, 36 dBm at 5 GHz and 57 dBm at 60 GHz, meaning that in Europe the range of IEEE 802.11n/ac will be shorter and that of

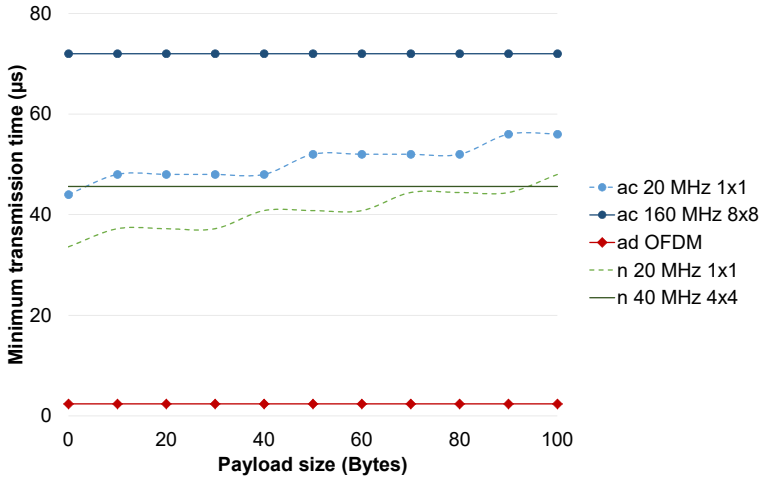


Figure 6.4: Transmission time versus payload size for high–performance WLAN standards.

a fading margin F of 12 dB and center frequencies of 2.4 GHz for IEEE 802.11n, 5 GHz for IEEE 802.11ac and 60 GHz for IEEE 802.11ad. It can be seen that IEEE 802.11ad achieves very high data rates (over 1 Gbps) at the expense of a reduced range (less than 100 m), while IEEE 802.11ac with 8×8 MIMO and 160 MHz channels manages to maintain such rates up to almost 1000 m. Lower–bandwidth versions of IEEE 802.11n/ac, conversely, do not reach 100 Mbps but are able to cover more than 1000 m range when the maximum allowable EIRP is adopted under free–space line–of–sight conditions. In realistic Non Line–of–Sight (NLOS) propagation environments with shadowing, the range may be considerably lower.

In Fig. 6.4 the transmission time of a packet with different payload size L is reported for the high–performance WLAN standards. This metric represents a lower bound to the scheduling unit, which was described as a critical performance indicator in WirelessHP scenarios. Only payload lengths lower than 100 bytes have been considered, as industrial communications are characterized by the exchange of short packets. The transmission time for a packet transmitted with OFDM can be expressed as

$$T_{TX} = T_{preamble} + T_{sym} \cdot \left\lceil \frac{L + L_{header}}{N_{DBPS}} \right\rceil \tag{6.3}$$

where $T_{preamble}$ is the time required to transmit the preamble, T_{sym} is the transmission time of a single OFDM symbol, L_{header} is the length of the various headers. N_{DBPS} is the number of bits carried in each symbol and depends on the modulation, coding, number of OFDM subcarriers and number of MIMO streams. It can be observed that

IEEE 802.11ad will be longer.

IEEE 802.11ad provides by far the best performance, with a transmission time of roughly 3 μs. This is due to both the reduced OFDM symbol duration T_{sym} and the shorter preamble time.

The limit performance of the IEEE 802.11 standard will be updated in the future as new amendments will be published, such as IEEE 802.11ax and IEEE 802.11ay. For a detailed description of these amendments, please refer to Sec. 3.1.

WPAN and cellular standards

Besides IEEE 802.11, other international wireless standards define high–performance networks, such as the IEEE 802.15.3 standard for HR-WPANs and the most recent cellular standards defined by 3GPP. These standards are described in detail in Sec. 3.3 and only some key features are recalled here.

The first active HR-WPAN amendment is IEEE 802.15.3c, published in 2009 for the 57-66 GHz mmWave spectrum. It presents different PHY layer modes, based on single carrier or OFDM, reaching a peak data rate of 5.78 Gbps. Another amendment, IEEE 802.15.3e, has been released in 2017. Also deployed in the mmWave spectrum, IEEE 802.15.3e is designed for high–rate close–proximity communications, with rates as high as 100 Gbps and typical communication distance of 10 cm. However, this amendment is not considered in this section, since the use case, namely fast multimedia data exchange at close proximity, is not involved with industrial applications.

The most advanced deployed cellular standard, instead, belong to 4G and it is called LTE-A. Several dedicated frequency bands between 700 MHz and 3.6 GHz are employed, with a maximum transmission bandwidth of 100 MHz. MIMO is used for both multiplexing and diversity and the peak data rates are 3 Gbps in downlink and 1.5 Gbps in uplink The next advancement in cellular standards will be 5G, whose first deployments will start around 2020. The 5G vision foresees a flexible system that can support many different use cases with distinct requirements. The most interesting use cases for WirelessHP are those related to URLLC, where the latency and reliability requirements are much higher than in traditional broadband cellular communications. Several key technologies are considered to reach these requirements, such as the use of mmWave spectrum, massive MIMO (64×64 and more), use of unlicensed spectrum, new waveforms and enhancements at the network layer. It is worth emphasizing that the direct application of cellular networks to WirelessHP scenarios is not a realistic option, as the network overhead involving base stations and routers does not allow to reduce latency at μs–level. Nonetheless, some emerging technologies in cellular URLLC may be applicable for the design of WirelessHP networks.

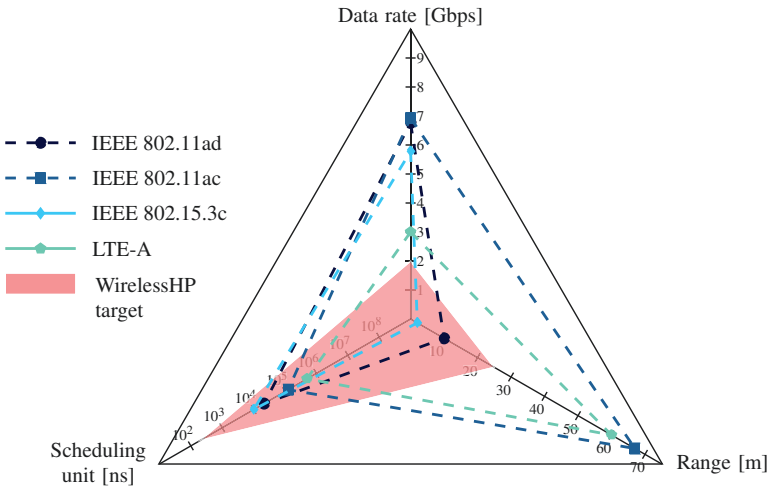


Figure 6.5: Link–level performance of most advanced wireless standards against target requirements for WirelessHP.

Table 6.3: Link–level performance for high–performance standards and required performance by WirelessHP (10^2 bits packets)

Standard	Data rate	SU	Link range
IEEE 802.11ad	6.76 Gbps	21.47 μs	10 m
IEEE 802.11ac	6.93 Gbps	126 μs	67 m
IEEE 802.15.3c	5.78 Gbps	10.08 μs	2 m
LTE-A	3 Gbps	500 μs	60 m
WirelessHP baseline	500 Mbps	500 ns	3 m
WirelessHP target	2 Gbps	200 ns	10 m

Assessment of current high–performance standards

The plot in Fig. 6.5 assesses the performance of the most advanced current wireless standards, namely IEEE 802.11ac/ad, IEEE 802.15.3c and LTE-A, in the three link–level metrics presented for WirelessHP. These performance are compared with the target requirements reported in Tab. 6.2. The exact performance values are also reported in Tab. 6.3.

For WLAN and WPAN standards, the SU is computed for a 10^2 bits packet, while for LTE-A it is equal to the minimum TTI assigned to an user. The link range is computed according to Eq. (6.2) for the WLAN and WPAN standards, considering the highest possible data rate, and derived from typical values for LTE-A.

It can be observed that, in terms of data rate and link range, the standards are almost matching the WirelessHP requirements. In particular, all the considered standards offer

peak data rates higher than the 2 Gbps required by the target scenario and each of them, with the exception of IEEE 802.15.3c, is able to guarantee the 3 m link range required by the baseline scenario, with LTE-A and IEEE 802.11ac providing the highest range. However, the gap in terms of SU is still around two orders of magnitude, also considering the baseline requirements, thus suggesting that latency is the bottleneck in the design of WirelessHP networks. Interestingly, and confirming the trend reported in Fig. 6.4, the mmWave standards (IEEE 802.11ad and IEEE 802.15.3c) are those offering the shortest SU.

Future wireless standards, especially 5G, will offer better performance, although it is unlikely that they will be able to meet the target WirelessHP requirements. A clean–slate system design, hence, appears the only way to achieve WirelessHP communications.

Emerging trends in fundamental technologies

In recent years, the focus in the design of wireless networks is shifting from the old vision centered on data rate towards flexible systems that can simultaneously satisfy many conflicting requirements, such as low latency, high reliability, high throughput for a large number of users and so on. As a consequence, the research community is investing a significant effort in proposing new ideas to meet this new vision. Some of the most interesting trends in this regard are discussed in the following.

New results on modulations and coding

OFDM modulation has long been the preferred scheme for wireless communications, from WLANs to cellular networks. However, triggered by the recent proposal of simpler frequency–domain equalization schemes, single carrier modulation is gaining renewed interest, as it allows higher energy efficiency thanks to its reduced Peak–to–Average Power Ratio (PAPR) compared to OFDM (Benvenuto et al., 2010). Single carrier can also ensure lower latency, although a proposed architecture with block transmission and frequency–domain equalization is likely to have similar latency as OFDM.

As for multicarrier schemes, new alternatives to OFDM are being proposed, such as Generalized Frequency–Division Multiplexing (GFDM), Filter Bank Multicarrier (FBMC), Universally Filtered Multicarrier (UFMC) and Filtered Orthogonal Frequency–Division Multiplexing (F-OFDM) (Banelli et al., 2014). All of them are based on digital filtering performed on single subcarriers or on a set of them. This feature allows reduced Out–of–band Emissions (OOBE) compared to OFDM, allowing to use more efficiently the available bandwidth. The drawback of these strategies is that the sharp spectral rolloff of the filters calls for long filter lengths, which impacts the latency (Schaich et al.,

2014). This is particularly true for the schemes that filter each subcarrier individually, namely GFDM and FBMC, which provides the lowest OOBE. Finally, all the filtered multicarrier schemes present an increased hardware complexity with respect to OFDM.

As for channel coding, the most advanced proposals are Turbo, LDPC and Polar codes (proposed for 5G). However, all these schemes require high codeword length in order to reach very high efficiency and may be less efficient when the codewords are short, as is the case of short WirelessHP packets. Traditional schemes such as convolutional and block codes may offer both higher efficiency and shorter decoding latency (Yoo et al., 2010).

The transmission of short packets with low latency and high reliability requirements is starting to gain attention also among information theory researchers. Traditional PHY layer metrics such as ergodic capacity and outage probability are no longer significant when the packet size is small and new metrics must be introduced, such as the maximum coding rate that can yield a certain error level (Durisi et al., 2016). Moreover, the control overhead represented by preamble and headers is typically disregarded in traditional information theory, whereas it becomes non negligible when the payload size is low, especially since overhead bits are typically encoded in a suboptimal way, for example through repetition coding. Preliminary results suggest that joint encoding of the data bits and the control bits (preamble, headers, trailers) may result in high performance improvement (Durisi et al., 2016). Finally, diversity in time, frequency and space is identified as the key to achieve reliability. However, adding diversity comes at the price of increased overhead, hence, compromises have to be found (Durisi et al., 2016).

mmWave communications

The term mmWave broadly refers to the portion of the spectrum between 30 and 300 GHz, although the majority of current activities are focused on the unlicensed band around 60 GHz. The high available bandwidth in this range of the spectrum allows to reach extremely high data rates.

The propagation characteristics above 30 GHz are significantly different from those below this threshold. Free–space path loss is 20-40 dB highere in the 60 GHz band. Oxygen absorption, which peaks at both 60 and 120 GHz, adds an additional 7-15.5 dB/km loss. Rain attenuation can be also significant, as is shadowing caused by large objects such as human bodies. All these characteristics make long–distance communication in the 60 GHz band practically infeasible, unless very high transmitting power and high–gain antennas are employed. Measurement campaigns with commercial mmWave hardware has shown that Gbps transmission is feasible within a few meters (Ansari et al., 2015).

When considering small–scale effects, the mmWave channel is still a multipath channel, but the delay spread is lower (typically less than 30 ns) compared to the sub-30 GHz spectrum (Park and Rappaport, 2007).

The MAC layer design for mmWave systems reflects the propagation issues, since it must handle directional communications, which bring issues such as neighbor discovery, deafness and enhance hidden terminal problems. Several directional MAC layer schemes have been proposed in the literature and a comprehensive review can be found in Niu et al. (2015). Joint PHY/MAC techniques that combine NLOS transmission and relaying can be employed to mitigate shadowing effects.

mmWave communications represent a significant opportunity in the WirelessHP scenario and some examples are already being proposed, such as that presented in Yamamoto et al. (2015), where an experimental setup for the wireless control of an Insulated Gate Bipolar Transistor (IGBT) is demonstrated. This PEC application makes use of 60 GHz transceivers to send signals modulated by a 400 MHz amplitude modulation from a controller circuit to a power circuit composed of some IGBTs. A simplified protocol stack to allow for real–time communications is employed and parallel transmission in multiple channels is considered for reliability. The results show an extremely low latency of up to 400 ns together with a reduced jitter of up to 250 ns, which are close to WirelessHP requirements.

URLLC communications in 5G

The URLLC scenario for 5G networks has similar requirements to those of WirelessHP. A vast number of ideas is being proposed to cope with low latency and high reliability requirements.

The main idea is to review the design of the radio frame, which in LTE has a fixed 10 ms duration and it is divided in TTIs of 0.5 ms (Ford et al., 2017). A reduced duration of OFDM symbols is proposed, together with a removed or reduced cyclic prefix, in order to lower the TTI. Furthermore, flexible TTI duration is proposed to serve users with different requirements, exploiting the non–orthogonality of F-OFDM. The use of low–latency convolutional codes is also envisioned. Finally, network latency is also improved by enhancements of the network structure, such as virtualization and device–to–device communications (Ford et al., 2017). The latter will shorten the distance and the number of hops that messages have to travel and, hence, reduce end–to–end latency.

The key to increase reliability is again diversity, with a preference for spatial and frequency diversity, which do not impact latency. In particular, the use of MIMO for

diversity allows to reach very high reliable performance (Johansson et al., 2015).

The application of 5G URLLC technology to FA is the subject of the work in Yilmaz et al. (2015). An assessment by simulations is conducted based on real measurements in industrial environments. It is shown that a PER of 10^{-9} can be reached in 99.999% of the environment with latency lower than 1 ms if a 2×8 MIMO system is employed where the base station exploits an high number of antennas to achieve diversity.

High–performance industrial wireless systems proposals

Some proposals can be found in the literature containing complete designs for wireless systems to be employed in industrial control applications.

The work in Wei et al. (2013) proposes RT-WiFi, a MAC layer protocol for real–time IEEE 802.11 networks which provides complete integration with both existing PHY and upper layers and can be seen as a representative implementation of the WIA-FA standard. A flexible TDMA scheme is proposed at the MAC layer, with slot time as low as 200 μs, and the use of IEEE 802.11 Timing Synchronization Function (TSF) for timing purposes. In–slot or out–of–slot retransmissions are envisioned to improve reliability and mechanisms for coexistence with other networks, such as clear channel assessment and reduced inter–frame spacing, are proposed (even though they clearly impair the determinism of the system). When applied over the IEEE 802.11g PHY layer, the protocol allows to significantly reduce both the average (around 500 μs) and maximum latency on packet delivery, while maintaining a low loss ratio.

An alternative approach is proposed in Dombrowski and Gross (2015), where a token–passing MAC scheme named EchoRing is presented. The protocol adopts cooperative relaying, where the station which has the token can select another station to assist transmission. At the PHY layer, low–order modulations and coding are used in a 10 MHz bandwidth. An experimental assessment shows that the protocol is able to outperform other token–passing schemes as well as the standard IEEE 802.11 MAC, reaching a latency of 10 ms and a PER than 10^{-6}.

The work in Swamy et al. (2015) presents a scheduled TDMA protocol, OccupyCoW, which makes use of cooperative relaying to reach very high levels of reliability, while maintaining a fixed cycle time of 2 ms in a network of 30 nodes. The cycle is divided in initial downlink (broadcast) and uplink phases, a scheduling phases and two additional downlink and uplink phases where relaying operations can take place. In order to support simultaneous packet transmission, a cyclic–delay diversity technique is adopted at the PHY layer. Through theoretical and simulation analysis, it is shown that the required SNR to reach a 10^{-9} error rate is very low (around 0 dB) when 3–hops relaying is

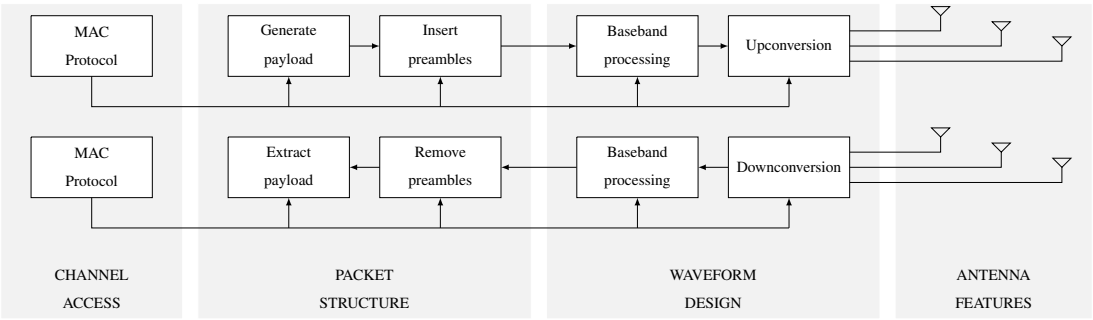


Figure 6.6: Main areas and blocks in the transmit and receive paths of a wireless link.

considered. An updated version of the same work proposes the use of network coding to enhance relaying performance, especially in the 2–hops scenario (Swamy et al., 2016).

Although these results allow to reach considerably higher performance compared to existing standards, they are still not sufficient for the WirelessHP scenario. Indeed, the presented approaches are only able to provide update rates of 250 Hz (Wei et al., 2013), 20 Hz (Dombrowski and Gross, 2015) and 500 Hz (Swamy et al., 2015), far below the WirelessHP requirements reported in Tab. 6.1. The main reason is that, while all these systems present significant enhancements to the MAC layer, their PHY is not customized and limited in bandwidth, thus not allowing to significantly decrease the SU duration.

6.3 Directions towards high–performance industrial wireless

The existing high–performance wireless standards and the various proposals available in the literature contain a vast set of options that could be considered when designing a new wireless communications system. A subset of them can be useful in the WirelessHP scenario and in this section they will be reviewed, targeting at low latency and high reliability. The design options are grouped in four main areas: channel access, packet structure, waveform design and antenna features. Each area pertains to a different step in the transmit–receive path of a wireless node, as depicted in Fig. 6.6, and has an impact on both latency and reliability. As a conclusion, a summary of the main options to be adopted in a WirelessHP system can be found in Tab. 6.4.

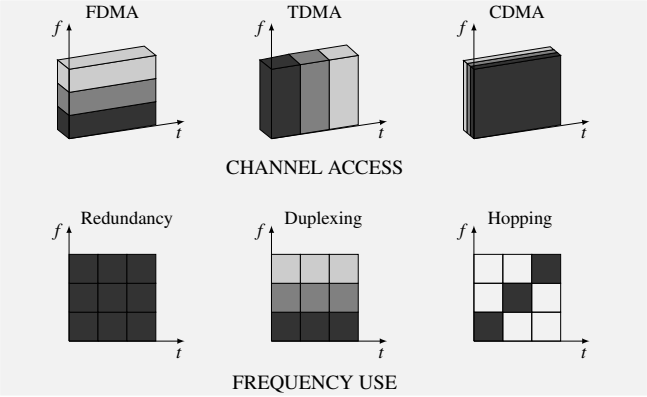


Figure 6.7: Main aspects of the channel access area.

Channel access

This area is concerned with the choice of an effective strategy to control channel access among multiple users and with mechanisms to use the available radio resources, as shown in Fig. 6.7.

Several channel access strategies have been proposed and employed in the design of wireless networks. The CSMA/CA scheme used in many wireless networks is not suitable for WirelessHP, due to its high latency and lack of determinism. Token passing schemes, such as the one proposed in Dombrowski and Gross (2015), provide an high degree of determinism but the token circulation adds overhead and there are robustness issues linked to the loss of the token. A TDMA system appears to be the best option, since it allows for bounded channel access delay. A strict timing synchronization and centralized scheduling are required for this scheme to work. Frequency–Division Multiple Access (FDMA) can also be considered in addition to TDMA, where different users can be scheduled on the same timeslot but in separate frequency bands. In addition to timing, frequency synchronization is also required if FDMA is employed together with TDMA. CDMA also allows for simultaneous channel access by multiple users, but it requires more bandwidth than the one actually necessary for data communication between two users, due to the adoption of spread–spectrum techniques.

Assuming that a high bandwidth is available, it can be used for a single transmission or split in multiple subchannels of lower bandwidth. In this case, redundant copies of the same message can be sent on different parallel subchannels to enhance reliability through frequency diversity, and/or simultaneous downlink/uplink transmissions can be scheduled on different subchannels (frequency duplexing). Finally, a frequency hopping strategy can

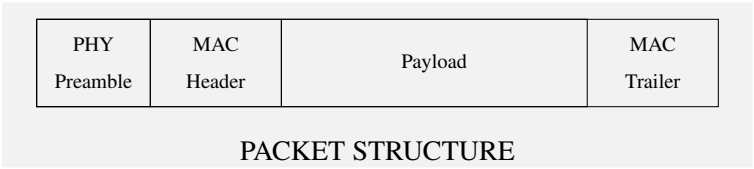


Figure 6.8: Packet structure for WirelessHP.

be considered, where the subchannel adopted for data transmission is changed according to a predefined cyclic pattern, in order to combat external narrowband interference and frequency–varying fading.

It is noted that radio regulations mandate restrictions on channel usage in order to ensure fair and efficient spectrum usage. This typically implies limits on transmission power, duty cycle, and channel access method. The precise limits depend on the frequency bands, and are more restrictive in the unlicensed bands where a large number of systems have to co–exist. The design of channel access for WirelessHP has to take these restrictions into account.

Packet structure

This area is involved with the definition of the different parts of a packet to be used in WirelessHP, as reported in Fig. 6.8. In WirelessHP networks, the different protocol layers are not designed separately, but a cross–layer design approach should be pursued. The main focus is on PHY and MAC, while the impact of upper layers (network and transport) must be minimized, since many of their functionalities are in general not required. Consequently, a typical WirelessHP packet is composed by an application layer payload of a few bits, MAC header and trailer and the PHY layer preamble. Preamble and header have to be jointly encoded with the data, as suggested by information–theoretical analyses (Durisi et al., 2016), and should be as short as possible, to limit the latency. The header and trailer can be significantly shortened with respect to those adopted in the standards. With fixed configurations, header bits defining configuration options can be omitted and only node addresses and error–control bits such as Cyclic Redundancy Check (CRC) need to be included. The length of the PHY layer preamble can also be significantly reduced. Typical functions of the preamble are packet detection, Automatic Gain Control (AGC) calibration, channel estimation and time/frequency synchronization. The first function is still required but may be simpler in a strictly synchronized TDMA system, where nodes know with a good approximation when to expect the packet. With

fully static deployment, the other functions, instead, need to be performed only in the network initialization phase and then sporadically updated, thus requiring a lower number of bits for the preamble.

When a packet is lost, the typical approach is to retransmit it, thus achieving temporal diversity. While retransmissions allow to increase reliability, they have a strong impact on both latency and determinism, and, hence, they should be performed as fast as possible in order to keep this impact low. However, if the new transmission is performed in the same frequency band and the channel is slowly varying, it is likely to have the same outcome than the first one, thus wasting channel resources. Furthermore, a typical control application involves the exchange of two types of messages, commands from the controller to the nodes and sensing data from the nodes to the controller, with different reliability requirements. In particular, commands must be acknowledged while, in case a sensing message is lost, it is often better to send the updated sensing data in the following cycle rather than retransmitting old data. Considering all these aspects, adaptive retransmissions in WirelessHP is not a good option and a fixed and low number (possibly zero) of in–slot retransmissions should be considered. HARQ can be employed for retransmission since it can bring a limited overhead, even at the expense of increased complexity. In the WirelessHP scenario, piggybacked ACK is the most suitable choice for commands, while sensing messages are either not acknowledged at all or a negative acknowledgement can be sent only after a certain number of different consecutive packets have not been received.[3] Finally, the frame aggregation and block ACK features that have been introduced in WLANs and WPANs have not to be considered in a context where packets are short and low latency is crucial. Additional reasons to avoid frame aggregation are its complexity and the higher frame error probabilities that would be experienced.

Waveform design

This area deals with all the digital and analog processing operations that transform a packet of bits into a waveform to be sent over the air, as depicted in Fig. 6.9.

The first aspect taken into account is the frequency spectrum. The majority of current standards adopt either the 2.4/5 GHz bands or the mmWave 60 GHz band. The latter option appears as more valuable, due not only to the huge available bandwidth but also to the reduced delay spread, which allows to have shorter guard intervals against ISI and, hence, to reduce latency. However, attenuation is higher, so that achievable range is

[3]Since communications are tightly scheduled, the controller knows exactly when to expect a sensing message and, therefore, can detect communication failures.

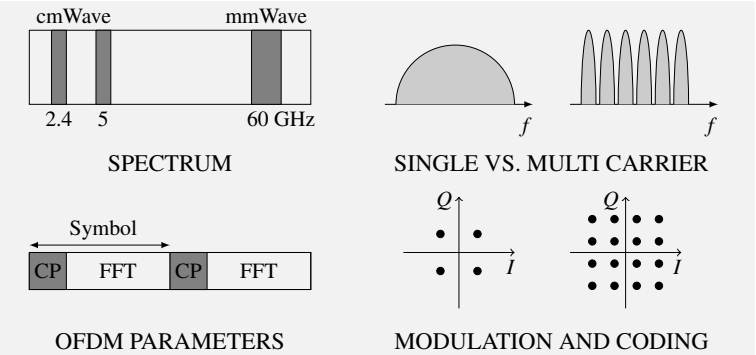


Figure 6.9: Main aspects of the waveform design area.

lower. The 2.4, 5 and 60 GHz bands are typically employed because they are allocated for unlicensed use.

Once a suitable frequency band has been identified, the next step is to choose between single and multi–carrier systems. Single carrier provides higher energy efficiency and shorter latency, although the latter is true only if block transmission and frequency–domain equalization are not adopted. The choice will hence be driven by the effectiveness of channel equalization approaches. Among the multi–carrier options, traditional OFDM should be preferred to strategies based on filtering, as it exhibits a reduced latency. If OFDM is chosen, there are several parameters that can be tuned towards reduced latency. The first one is the cyclic prefix length, which should be as short as possible while allowing to avoid ISI (i.e., it should be equal to the maximum delay spread of the propagation channel). The FFT size is also a relevant parameter, as it is directly proportional to the symbol time: increasing the FFT size will allow to fit more bits in one OFDM symbol, while decreasing it will shorten the symbol time. The optimal value for the FFT size should be found given the number of bits to transmit.

The modulation order and coding rate can also be selected at the design stage through an optimization approach rather than being configurable at each packet transmission, as in WLAN standards. Indeed, high–order modulations and code rates allow very high data rates but suffer from poor reliability and may be unnecessary if the number of bits per packet is low. The possibility of adapting the modulation and coding to the status of the channel through consecutive transmissions is often proposed in literature and considered by the standards in order to increase reliability (Tramarin et al., 2015), as discussed in Chapter 4 of this thesis. Such an approach can still be valuable, provided that the duration of the packet transmission does not vary significantly over different

Table 6.4: Directions for the configuration of a WirelessHP system

Cathegory	Options	Pros	Cons	Directions for WirelessHP
Channel access	CSMA/CA Token passing TDMA FDMA CDMA	Distributed and fair Deterministic Deterministic Multiuser Multiuser	No determinism Overhead and less robust Time synchronization Frequency synchronization Low spectral efficiency	TDMA + FDMA
Frequency use	Frequency redundancy Frequency duplexing Frequency hopping	Improved reliability Simultaneous uplink/donwlink Robust to external interference	Low spectral efficiency Low spectral efficiency Low spectral efficiency	Frequency redundancy and/or duplexing
Preamble/headers	Short Long	Reduced latency Improved reliability	Decreased reliability Increased latency	Short preamble/headers jointly encoded with data
Retransmissions	Fixed ARQ HARQ	Deterministic Improved reliability Maximum reliability	Decreased reliability Increased latency Overhead and complexity	Fixed and low number of retransmissions
Acknowledgements	No ACK Immediate ACK Block ACK Piggybacked ACK	Shorter latency Fast response Reduced overhead No overhead, shorter latency	Decreased reliability Increased overhead Increased overall latency Response packet required	No ACK for sensing data, piggybacked for control data
Frame aggregation	No aggregation Aggregation	Increased overhead (headers) Reduced overhead	Decreased latency Increased latency	No aggregation
Spectrum	μWave (2.4-5 GHz) mmWave (60 GHz)	Long range, mature hardware Huge bandwidth, short delay spread	Small available bandwidth Short range, hardware issues	Flexible band
Waveform	Single carrier Multicarrier OFDM Filtered multicarrier	High energy efficiency, lower latency Exploits frequency selectivity Reduced OOBE	Low spectral efficiency Low energy efficiency Increased latency, complexity	Single carrier or OFDM (depending on channel equalization)
Cyclic prefix	Short Long	Reduced latency Improved resilience to ISI	Increased exposure to ISI Increased latency	Shortest possible cyclic prefix
FFT size	Low High	Short symbol time More data in a symbol	Less data in a symbol Long symbol time	Optimized FFT size
Modulation order	High Low	More data in a symbol Improved reliability	Decreased reliability Less data in a symbol	Optimized modulation order
Code rate	High Low	More data in a symbol Improved reliability	Decreased reliability Less data in a symbol	Optimized code rate
Rate selection	Fixed Adaptive	Improved determinism Improved reliability	Less adaptive to the channel Requires channel knowledge	Fixed rate optimized at network setup
Channel coding	LDPC, Turbo, Polar Convolutional, block	High spectral efficiency Shorter latency	High decoding latency Lower spectral efficiency	Traditional codes (block and convolutional)
MIMO	Spatial multiplexing Spatial diversity	High spectral efficiency Improved reliability	Non efficient with short packets Lower spectral efficiency	Spatial diversity
Beamforming	Online Offline	Adaptive to channel Improved latency	Increased latency Sensitive to channel variations	Offline beamforming
Relaying	Distributed antennas Cooperative relaying	Improved coverage and reliability Improved coverage and reliability	Requires multiple antennas Increased latency	Cooperative relaying (only in extreme cases)

cycles (otherwise determinism in data delivery timing could be impaired) and that the overhead due to the exchange of channel status information is limited.[4] In a practical setup, modulation and coding should be optimized at the network setup and calibration stage, before the actual start of operations, preferably after nodes have done long–term measurements that should render a stable representation of the channel status. During the operational phase, the channel status should be monitored in a non–invasive way and

[4]In this sense, the adoption of a strategy that estimates the channel status at the receiver without the exchange of additional information, such as the RSIN–E algorithm discussed in Sec. 4.2, could be interesting.

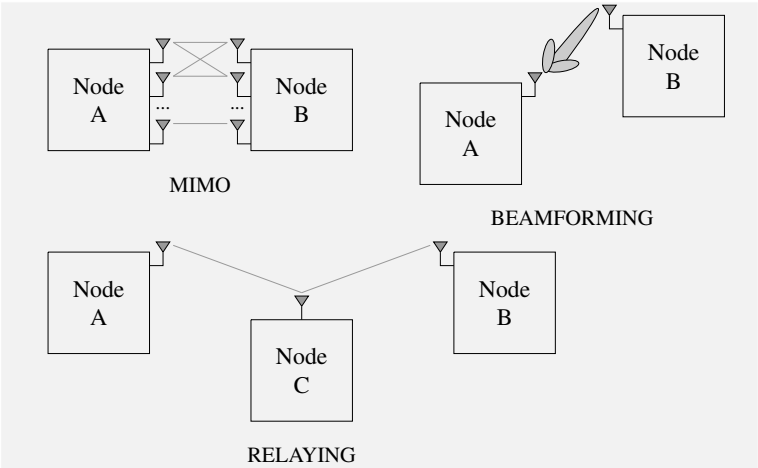


Figure 6.10: Main aspects of the antenna features area.

modulation/coding could be updated to increase reliability. This update process will be characterized by slow changes over time, following the the slow–varying nature of the WirelessHP channel. As for the channel coding strategy, convolutional and block codes are preferable to LDPC, Turbo and Polar codes, due to their reduced latency.

Antenna features

This last area is involved with all the possibilities that arise when multiple antennas are available at the controller and/or at the nodes for MIMO transmission, beamforming and relaying, as shown in Fig. 6.10.

Specifically, a MIMO architecture can be targeted at spatial multiplexing or spatial diversity. Spatial multiplexing schemes require a symmetric configuration and may not be that useful when packet size is short, thus MIMO is better suited in the WirelessHP scenario for spatial diversity. This can be achieved even when the configuration is asymmetric towards the receiver, using Maximum–ratio Combining (MRC), or towards the transmitter, using STBC (Tramarin et al., 2016b).

Beamforming is essential to establish directional communication in mmWave, where nodes are generally equipped with antenna arrays which can generate a precise beam pattern. mmWave standards, such as IEEE 802.11ad and IEEE 802.15.3c, envision beamforming procedures at the beginning of each network cycle, to cope with dynamic network topologies and rapidly changing channels. However, in the WirelessHP scenario, the topology is fixed and the channel is slowly varying, hence online beamforming, which significantly increases latency, is not required and can be configured during network

initialization.

Distributed antenna architectures can be employed to enhance reliability and coverage. A first scheme can be designed when the controller is equipped with multiple antennas, which are spread at different locations and connected by radio cables. An alternative distributed antenna scheme to enhance reliability, which does not necessarily require nodes to have multiple antennas, is cooperative relaying, where nodes that overhear a frame not destined to them can forward it to the desired receiver according to a specified relaying scheme, thus effectively acting as additional antennas. However, this feature should be considered only in extreme cases where coverage and reliability are vital while the latency requirements are more relaxed (e.g., the scenario proposed in Swamy et al. (2015)).

6.4 Design of a low–latency PHY

According to the WirelessHP vision, each layer in the protocol stack must be optimized towards the achievement of the required high performance for wireless networks to be employed in industrial applications. Specifically, the PHY layer must be redesigned with respect to general–purpose wireless standards, such as IEEE 802.11, with the goal of achieving the lowest possible latency for small–size packets.

Indeed, the analysis of currently available high–performance wireless standards in Sec. 6.2 revealed that the data rate and range requirements of Tab. 6.2 are already satisfied, while the SU, which is linked to the cycle time and hence the communication latency, is the most critical performance metric. While some guard margins should be reserved for synchronization, propagation and elaborations, the majority of time during a SU is devoted to packet transmission. The goal hence becomes to achieve the shortest possible transmission time for small–size packets, with payloads of 100 bits or lower. The value of 100 bits is considered indicative for the targeted applications, and includes bits reserved for destination and source addresses, status and command data, and error checking.

In this section, a low–latency PHY design is proposed, based on OFDM. In detail, the design starts from the IEEE 802.11 PHY and aims at reducing its inefficiencies in the exchange of short packets.

Packet transmission time in IEEE 802.11 OFDM

As detailed in Sec. 3.1, most versions of the IEEE 802.11 standard use OFDM at the PHY layer, from the legacy IEEE 802.11a and g to the more recent IEEE 802.11n, ac

and ad.

In an OFDM–based PHY, the transmitted waveforms are grouped in OFDM symbols. Within each symbol, information bits are mapped onto N_{FFT} subcarriers in frequency domain, which correspond to N_{FFT} samples in time domain via the FFT. The last N_{cp} time–domain samples are repeated and appended at the beginning of the symbol, forming the Cyclic Prefix (CP), that allows to avoid ISI in multipath fading channels. The transmission time of an OFDM symbol is hence given by

$$T_{sym} = T_s \cdot (N_{cp} + N_{FFT}) \tag{6.4}$$

where T_s is the sample time. The occupied bandwidth B is generally equal to the sample frequency $F_s = 1/T_s$, while the subcarrier spacing, i.e., the bandwidth assigned to a single subcarrier, is given by $\Delta_{SC} = B/N_{FFT}$.

A Physical layer Packet Data Unit (PPDU) is formed by a certain number of OFDM symbols, the first ones used for the preamble and the last ones for the data.[5] The number of symbols destined to the PHY preamble, N_{sym}^{pre} varies among the different versions of the IEEE 802.11 standard, as reported in Fig. 6.11. The number of symbols used for data, N_{sym}^{data}, instead, depends on the amount of data to send L (in bits), the number of subcarriers used for data N_{dsc}, the modulation order M, the channel coding rate R_c and the number of spatial streams N_{ss}, according to

$$N_{sym}^{data} = \left\lceil \frac{L}{N_{dsc} \cdot \log_2 M \cdot R_c \cdot N_{ss}} \right\rceil \tag{6.5}$$

The overall packet transmission time can hence be computed as the total number of OFDM symbols times the symbol duration, which, taking Eq. (6.4) into account, yields[6]

$$T_{pkt} = T_s \cdot (N_{cp} + N_{FFT}) \cdot \left(N_{sym}^{pre} + N_{sym}^{data} \right) \tag{6.6}$$

Tab. 6.5 reports the packet transmission time when $L = 100$ bits, no channel coding is employed ($R_c = 1$) and only 1 spatial stream is used, for different versions and configurations of the IEEE 802.11 standard and for low–order modulations. It is assumed that only the PHY of IEEE 802.11 is employed while the upper layers are customized in order to compress headers and trailers (IEEE 802.11 data–link header is 28 bytes long

[5] The "data" part includes the payload as well as all the headers and trailers added by upper layers, i.e., is the Physical layer Service Data Unit (PSDU).

[6] It is assumed here that OFDM symbols used in the preamble have the same structure of those used for data, specifically they contain the CP. In some OFDM systems, where preamble symbols do not have a CP, Eq. (6.6) should be slightly modified.

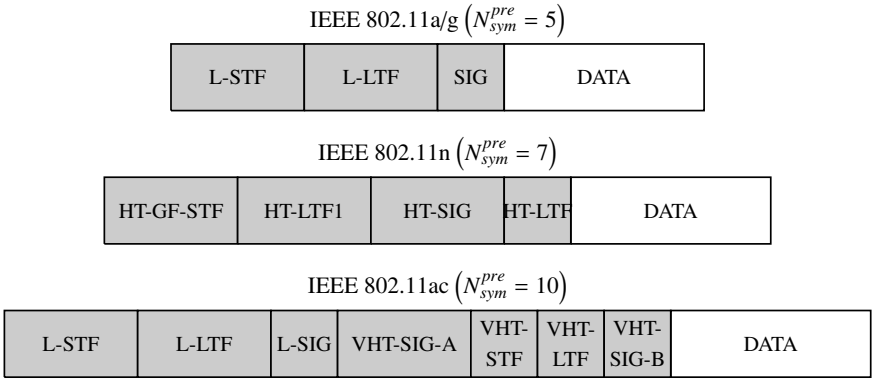


Figure 6.11: PPDU format in different versions of the IEEE 802.11 standard. Preamble symbols are in light gray, while data symbols are in white. For IEEE 802.11n/ac the number of spatial streams is set to 1 and, for IEEE 802.11n, the GF preamble is considered (IEEE 802.11-2016).

Table 6.5: Transmission time for 100 bits packets with IEEE 802.11 OFDM PHY

Standard	**B**	$\mathbf{N_{FFT}}$	$\mathbf{N_{dsc}}$	$\mathbf{N_{cp}}$	$\mathbf{N_{sym}^{pre}}$	**TX time, $\mathbf{T_{pkt}}$**		
						$M = 2$	$M = 4$	$M = 8$
IEEE 802.11a/g	5 MHz	64	48	16	5	128 μs	112 μs	96 μs
IEEE 802.11a/g	10 MHz	64	48	16	5	64 μs	56 μs	48 μs
IEEE 802.11a/g	20 MHz	64	48	16	5	32 μs	28 μs	24 μs
IEEE 802.11n	20 MHz	64	52	16	7	40 μs	36 μs	32 μs
IEEE 802.11n	40 MHz	128	108	32	7	36 μs	32 μs	32 μs
IEEE 802.11ac	80 MHz	256	234	64	10	44 μs	44 μs	44 μs
IEEE 802.11ac	160 MHz	512	468	128	10	44 μs	44 μs	44 μs

(IEEE 802.11-2016), hence the PSDU length L would be much higher than 100 bits).

It can be observed from the table that there is a large gap between the IEEE 802.11 transmission time of short packets and the WirelessHP target requirement for the SU (200 ns), confirming the fact that current wireless standards can not provide satisfactory performance. Moreover, it can be noticed that the most recent standards, such as IEEE 802.11n and ac, although employing much higher bandwidth up to 160 MHz, show a slightly increased packet transmission time with respect to the old IEEE 802.11a/g. New standards, indeed, are not optimized at all for small packet transmissions, due to both their high number of preamble symbols, as reported in Fig. 6.11, and their high FFT size, which is suboptimal when only few bits have to be transmitted.

All the versions of the IEEE 802.11 standard discussed so far are working in the 2.4 and 5 GHz unlicensed frequency bands. In 2012 the IEEE 802.11ad amendment, working in the unlicensed mmWave 60 GHz spectrum, was released. The PPDU format

in IEEE 802.11ad is slightly different from those shown in Fig. 6.11. OFDM is still employed to transmit data, with $N_{FFT} = 512$, $N_{dsc} = 336$, $N_{cp} = 128$ and a bandwidth of $B = 2.16\ GHz$. However, the preamble is composed of a single carrier part, which requires a fixed transmission time of 1.89 μs, followed by one OFDM symbol. The packet transmission time of 100 bits packets with IEEE 802.11ad is hence of 2.38 μs (regardless of the modulation order), which is closer to the WirelessHP target of 200 ns, although still one order of magnitude higher.

WirelessHP low–latency PHY

A recent trend in the design of high–rate wireless PHY layers is the adoption of single carrier modulation as an alternative to OFDM (Benvenuto et al., 2010). Single carrier transmission has low latency, since it allows stream–, rather than block–, processing, and high energy efficiency, since it has low PAPR. However, in this section a WirelessHP PHY based on OFDM is proposed, since this modulation technique allows for easier channel equalization, easier compliance to spectrum mask regulations and partial reuse of existent designs (e.g., IEEE 802.11). In order to achieve the packet transmission time requirement for WirelessHP, an optimized OFDM PHY design has to minimize the inefficiencies that affect short packet transmission in IEEE 802.11.

Reducing preamble length

The impact of preamble on PHY performance is commonly disregarded in general purpose wireless communications, since its duration is negligible with respect to the entire packet (Durisi et al., 2016). In short–packet communications, however, the impact of preamble is of primary importance and its duration must be limited as much as possible.

To better clarify this concept, consider the preamble overhead (in samples) defined as

$$O_{pre} = N_{sym}^{pre} \cdot (N_{cp} + N_{FFT}) \tag{6.7}$$

In IEEE 802.11a/g the preamble overhead is $O_{pre} = 400$ samples, while the total number of samples to transmit a packet of $L = 100$ bits with $M = 8$ is 480, which means that 83% of the transmitted samples are used for preamble.

Reducing the number of preamble symbols, N_{sym}^{pre}, is hence a key step towards the increase of efficiency and the reduction of packet transmission time. However, in order to ensure a reliable packet decoding process, a customized preamble must support the main functions accomplished by the IEEE 802.11 preamble, which are detailed in the following.

- *Packet detection and timing synchronization:* These functions are concerned with identifying the beginning of a packet and achieving sample–level synchronization. The first task is generally realized by exploiting the correlation between repeated identical sequences, such as those contained in the L-STF part of the IEEE 802.11a/g preamble (Schmidl and Cox, 1996). The second task, accomplished by the L-LTF part, relies on the correlation between the received samples and the known transmitted ones (Heiskala and Terry, 2001).

- *Frequency offset estimation:* Carrier Frequency Offset (CFO) is a mismatch between transmitter and receiver oscillator frequencies due to Doppler effects and non–idealities of components. OFDM systems are particularly sensitive to CFO, since a strict frequency synchronization is required to ensure subcarriers orthogonality. Hence, an IEEE 802.11a/g OFDM receiver estimates the CFO by exploiting the correlation between repeated identical sequences in both L-STF and L-LTF and compensates for it before decoding.

- *Channel estimation:* This function is concerned with estimating the response of the wireless channel in view of performing channel equalization. In IEEE 802.11a/g, frequency–domain channel estimation is performed by demodulating the received L-LTF symbols and performing element–wise division by the transmitted ones.

- *Information about length and coding:* In order to ensure a correct decoding process, an OFDM receiver must know the length of the PSDU as well as the modulation and coding schemes adopted in the packet generation process. In IEEE 802.11a/g, this information is contained in the SIG field of the preamble.

In order to accomplish all the listed functions, the IEEE 802.11 amendments use a quite long preamble, as reported in Fig. 6.11. However, some key assumptions typical of industrial wireless communications can be exploited to design a simplified and reduced preamble still able to carry out all the necessary functions. A central assumption is the predictability of traffic patterns: industrial communications are generally tightly scheduled, as described in Sec. 2.2, and, hence, a node knows with high precision the time instant at which a packet destined to it is supposed to arrive. Consequently, simpler packet detection and timing synchronization algorithms can be designed, that do not need to correlate long sequences. Another important fact is the low temporal variability of the industrial wireless channel (Tanghe et al., 2008), which can be exploited to simplify the channel estimation procedures. Finally, the messages exchanged in industrial control applications will be of predefined length and the modulation and coding options are likely to be selected during the network calibration phase and remain fixed, as detailed

in Tab. 6.4, hence there is no need of including this information in the preamble of each packet.

The WirelessHP PHY design proposed in this section adopts a reduced preamble of $N_{sym}^{pre} = 1$ symbol, which represents a substantial reduction with respect to the preambles adopted in the IEEE 802.11 amendments and reported in Fig. 6.11. It will be shown in the following that a receiver is able to perform packet detection, timing synchronization, CFO and channel estimation with this one-symbol preamble, by exploiting the mentioned a priori information.

Optimizing OFDM parameters

The preamble overhead is not the only source of inefficiency in OFDM communications. There are three other causes of overheads that will be detailed in the following.

- *CP:* every data symbol contains a CP of N_{cp} samples, resulting in an overhead of

$$O_{cp} = N_{sym}^{data} \cdot N_{cp} \tag{6.8}$$

- *Unused subcarriers:* in the OFDM encoding process, the set of subcarriers onto which information bits are mapped does not include all the N_{FFT} subcarriers, as some of them are reserved for special use: N_{psc} pilot subcarriers are used to transmit pilot data employed to correct residual phase errors before decoding; N_{gsc} guard subcarriers at the edges of the symbol are nulled in frequency to avoid out–of–band emissions; N_{dcsc} subcarriers in the middle of the symbol are nulled in frequency to avoid Direct Current (DC) offset (Yih, 2009). The number of data subcarriers can hence be computed as

$$N_{dsc} = N_{FFT} - N_{psc} - N_{gsc} - N_{dcsc} \tag{6.9}$$

  while the overhead due to unused subcarriers is

$$O_{usc} = N_{sym}^{data} \cdot (N_{psc} + N_{gsc} + N_{dcsc}) \tag{6.10}$$

- *Padding bits:* the total number of data subcarriers is always $N_{sym}^{data} \cdot N_{dsc}$, but the information bits are mapped only onto $\left\lceil \frac{L}{\log_2 M \cdot R_c \cdot N_{ss}} \right\rceil$ subcarriers, and the two quantities do not necessarily coincide.[7] The subcarriers in excess are padded with

[7] For example, if $L = 100$ bits are transmitted with $M = 2$, $R_c = 1$ and $N_{ss} = 1$ in IEEE 802.11ac with 160 MHz bandwidth, they are mapped only onto 100 of the 468 available data subcarriers.

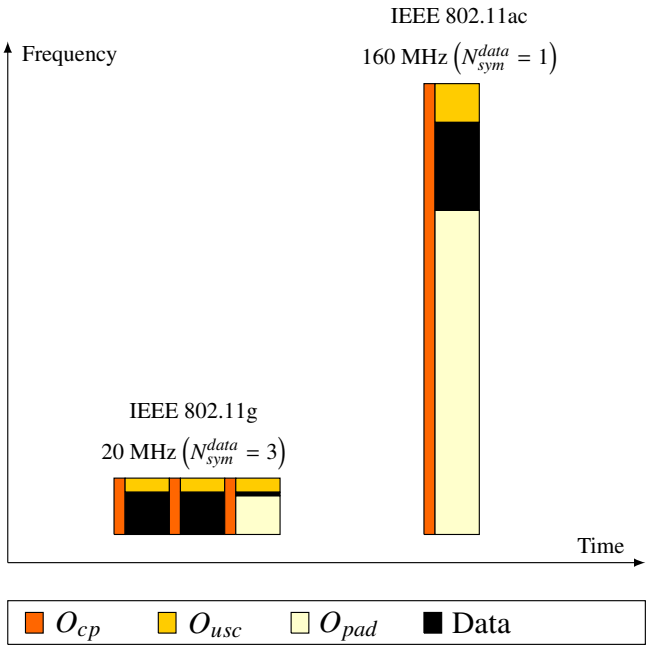


Figure 6.12: OFDM overheads in data symbols for IEEE 802.11g (with 20 MHz bandwidth) and IEEE 802.11ac (160 MHz). Fixed parameters: $M = 2$, $R_c = 1$, $N_{ss} = 1$, $L = 100$ bits.

zero bits, thus yielding an overhead of

$$O_{pad} = N_{sym}^{data} \cdot N_{dsc} - \left\lceil \frac{L}{\log_2 M \cdot R_c \cdot N_{ss}} \right\rceil \tag{6.11}$$

Fig. 6.12 provides a visual representation of the three different sources of overhead in OFDM (CP, unused subcarriers and padding bits) versus the amount of modulated data symbols (represented in black), for two extremes of the IEEE 802.11 standard: 802.11g with 20 MHz bandwidth and 802.11ac with 160 MHz bandwidth. It can be observed that the data occupy only a portion of the packet, particularly in IEEE 802.11ac, where the majority of samples (368 out of 640) are wasted for padding.

The WirelessHP PHY proposed in this section aims at finding the optimal OFDM parameters that can minimize the total packet transmission time. Some parameters cannot be changed because they are imposed by the application (the PSDU size L), by the hardware capabilities (the sampling time T_s and the number of spatial streams N_{ss}), or by preamble design (N_{sym}^{pre}).It is also assumed that the values of modulation and coding (M and R_c), as well as the number of pilot and DC null subcarriers (N_{psc} and N_{dcsc}), are optimized at a later stage and hence can not be changed here. This leaves only three

parameters to be optimized, namely the FFT size N_{FFT}, the cyclic prefix length N_{cp} and the number of guard subcarriers N_{gsc}, yielding the following constrained optimization problem

$$\underset{N_{FFT},N_{cp},N_{gsc}}{\arg\min} \quad T_{pkt} \tag{6.12}$$

where T_{pkt} is the quantity expressed by Eq. (6.6).

The proposed optimization problem is an integer programming problem, as the variable parameters are forced to assume integer values. Moreover, there are a set of constraints to be considered, which are detailed in the following.

- *CP long enough:* the CP is inserted before any OFDM symbol to combat ISI in multipath fading channels. In order to do it efficiently, the cyclic prefix duration must exceed the maximum delay spread of the channel (Tse and Viswanath, 2005), i.e.,

  $$T_s \cdot N_{cp} \geq T_{ds}^{max} \tag{6.13}$$

  The maximum delay spread depends on the structure of the propagation environment and can be assessed through measurement campaigns.

- *Subcarrier spacing shorter than coherence bandwidth:* the coherence bandwidth of a wireless channel (B_c) is the range over which its frequency response can be considered flat. B_c can be approximated as the inverse of the maximum delay spread (Tse and Viswanath, 2005). While the overall transmission bandwidth in OFDM is generally greater than the coherence bandwidth, it is important that the channel experienced by a single subcarrier is flat, i.e., $\Delta_{sc} \leq B_c$. Given the subcarrier spacing of $\Delta_{SC} = B/N_{FFT}$, this yields the following constraint on FFT size

  $$T_s \cdot N_{FFT} \geq T_{ds}^{max} \tag{6.14}$$

- *Guard bandwidth large enough:* guard bandwidth ratio is defined as the ratio between guard bandwidth, given by $\Delta_{sc} \cdot N_{gsc}$, and the total transmission bandwidth $B = \Delta_{sc} \cdot N_{FFT}$. The standards do not fix an explicit value for the guard bandwidth ratio, as it depends on regulation spectrum masks, OFDM windowing, etc.[8] However, it is assumed that all these regulations can be summarized in a minimum guard bandwidth ratio value GBR_{min}, yielding a constraint on the number of guard subcarriers

  $$N_{gsc} \geq GBR_{min} \cdot N_{FFT} \tag{6.15}$$

[8] For example, in IEEE 802.11a/g with 20 MHz bandwidth $N_{FFT} = 64$ and $N_{gsc} = 11$, yielding a guard bandwidth ratio of approximately 0.172.

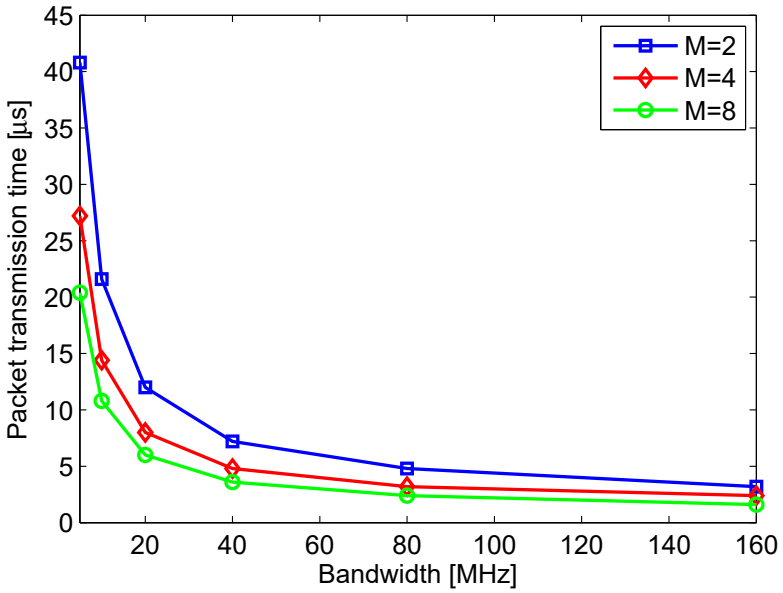


Figure 6.13: Packet transmission time in the 2.4/5 GHz WirelessHP PHY for different modulation orders and bandwidth with $L = 100$ bits packets. A maximum delay spread of $T_{ds}^{max} = 400$ ns is assumed.

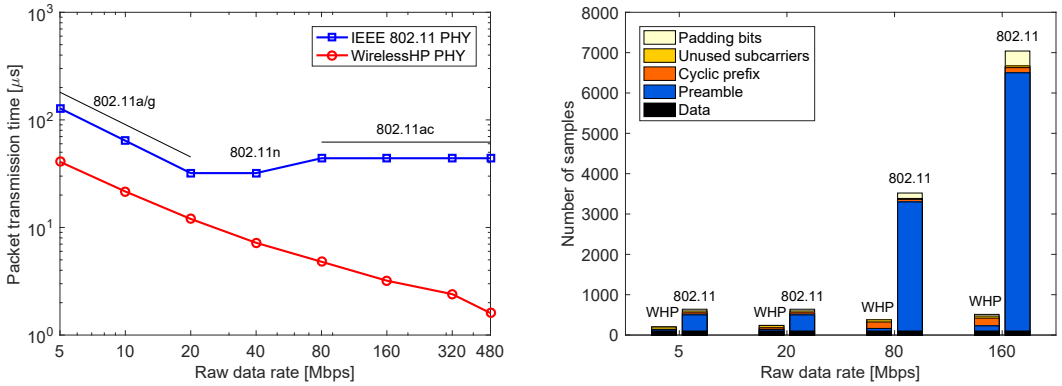
- *FFT size as a power of 2:* computationally efficient algorithms to perform FFT and its inverse require the number of samples to be a power of 2 (Brigham, 1974), hence it must hold

$$N_{FFT} = 2^m, \quad m \in \mathbb{N} \tag{6.16}$$

The WirelessHP PHY design proposed in this section is based on the solution of the integer programming problem in Eq. (6.12) subject to the constraints in Eq. (6.13), (6.14), (6.15), and (6.16).

Performance with optimized design

The performance of the WirelessHP PHY obtained through the solution of the optimization problem presented in Eq. (6.12) and the reduced one–symbol preamble are reported here. Some parameters are kept fixed throughout the evaluation. Specifically, the PSDU length is fixed to $L = 100$ bits, a representative value for most critical control applications, low–order modulations are employed ($M = 2$, 4 and 8) to achieve highly–reliable communications, and neither MIMO ($N_{ss} = 1$) nor channel coding ($R_c = 1$) are employed, to maintain the structure of the system as simple as possible. Finally, the OFDM parameters (except the FFT size) have been kept close to the values used in the IEEE 802.11g/n standard, i.e., $N_{psc} = 4$, $N_{dcsc} = 1$ and $GBR_{min} = 0.1875$. The



(a) Packet transmission time

(b) Breakdown of all transmitted samples per packet, highlighting the overhead

Figure 6.14: Comparison between IEEE 802.11 and WirelessHP PHY for $L = 100$ bits packets and different raw data rates.

first results are relevant to the 2.4/5 GHz spectrum, where the maximum transmission bandwidth is of 160 MHz (in the 5 GHz band only) and a conservative assumption for the maximum delay spread is $T_{ds}^{max} = 400$ ns (i.e., the minimum guard interval duration in IEEE 802.11n/ac). Finally, the number of preamble symbols is $N_{pre}^{sym} = 1$, as motivated previously.

Fig. 6.13 reports the packet transmission time with the WirelessHP PHY for different values of bandwidth (up to the maximum value of 160 MHz) and low–order modulations of $M = 2, 4$ and 8. The curves have been obtained by solving the optimization problem of Eq. (6.12) and plotting the value of T_{pkt} corresponding to the optimal parameter choice. It can be seen that the WirelessHP packet transmission time scales with the bandwidth (differently from the IEEE 802.11 PHY, as it is reported in Tab. 6.5), reaching the lowest values with $B = 160$ MHz, of 3.2 μs, 2.4 μs and 1.6 μs for $M = 2$, 4 and 8 respectively. It can also be observed that the reduction in packet transmission time obtained by increasing the modulation order is significant when the bandwidth is low, while it becomes almost negligible when high transmission bandwidth is available.

The increased efficiency of WirelessHP with respect to the IEEE 802.11 PHY is further analyzed in Fig. 6.14. The first plot, Fig. 6.14a, reports the transmission time for $L = 100$ bits packets versus the raw data rate. This metric, defined as $R = B \cdot \log_2 M$ and measured in bit/s, allows representing modulation and bandwidth simultaneously. The transmission time for such short packets in IEEE 802.11, as it was already reported in Tab. 6.5, does not scale with raw data rate and does not show significant improvements between IEEE 802.11a/g and IEEE 802.11n/ac, as the increased number of subcarriers

is compensated by a larger preamble. With the proposed PHY design, instead, the transmission time scales with the raw data rate and reaches almost 1 μs for $B = 160$ MHz and $M = 8$, which is almost 28 times lower than what IEEE 802.11ac achieves with the same settings.

A more detailed insight is given by Fig. 6.14b, which allows to see the role of the transmitted samples, distinguishing between data and the different types of overheads found in OFDM systems, for selected values of raw data rate. In all the considered values the modulation is fixed to $M = 2$ and, hence, the number of symbols onto which data are mapped is always 100 (for both IEEE 802.11 and WirelessHP). The number of overhead samples, conversely, varies significantly with the data rate and, most notably, with the transition from IEEE 802.11 to WirelessHP. As an example, for $R = 160$ Mbps, the IEEE 802.11 design requires 6940 overhead samples versus the 371 needed by WirelessHP. Finally, the figure allows to see that the preamble overhead, expressed by Eq. (6.7), is the major source of inefficiency in IEEE 802.11. In the WirelessHP design, instead, the preamble overhead is more relevant when R is low, whereas the cyclic prefix is the major source of overhead if the raw data rate is high.

In the 2.4/5 GHz spectrum, WirelessHP is able to achieve packet transmission times around 1 μs, which is still 5 times longer than the target SU value of 200 ns. However, it is not possible to decrease this time even further because of the limited bandwidth and of the high delay spread,[9] which, taking into account the constraints of Eq. (6.13) and (6.14), basically force every OFDM symbol to be not shorter than 800 ns. Moving to the mmWave spectrum around 60 GHz would allow to overcome both these limitations. In this frequency band, indeed, an higher transmission bandwidth is available, up to around 2 GHz. Moreover, the delay spread is generally lower than that observed at 2.4/5 GHz, as confirmed by the guard interval duration in IEEE 802.11ad, which is of 48.4 ns compared to the 400 ns in IEEE 802.11n/ac.

Fig. 6.15 reports the packet transmission time of the WirelessHP PHY for different bandwidth and modulation orders, where all the parameters have been kept to the previous values except the bandwidth (which is equal to the sampling frequency F_s) and the maximum delay spread, fixed to $T_{ds}^{max} = 50$ ns. It can be observed that again the packet transmission time scales with the available bandwidth, reaching 228 ns (close to the WirelessHP target SU value) for $B = 2\ GHz$ and $M = 4$ and 8. In comparison, the IEEE 802.11ad PHY, also operating in the mmWave spectrum, requires a bandwidth of $B = 2.16\ GHz$ to achieve a transmission time of 2.38 μs. Similarly to what is shown in

[9] With more advanced equalization, it would be possible to shorten or omit the cyclic prefix, see e.g., Farhang-Boroujeny and Ding (2001). However, such advanced equalizers are considerably more complex and, hence, will not be considered here.

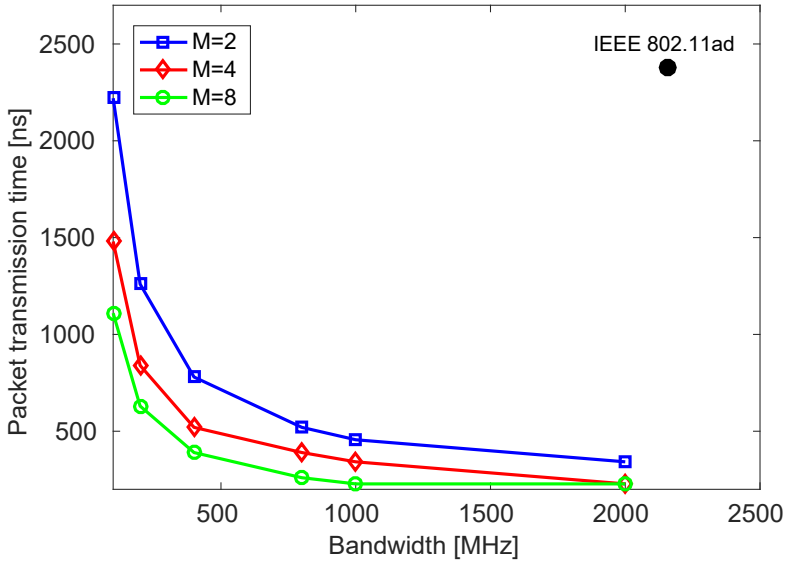


Figure 6.15: Packet transmission time in the mmWave WirelessHP PHY for different modulation orders and bandwidth with $L = 100$ bits packets, compared with packet transmission time in IEEE 802.11ad (same for all modulation and coding schemes). A maximum delay spread of $T_{ds}^{max} = 50$ ns is assumed.

Fig. 6.13, also in the mmWave spectrum the reduction in packet transmission time due to the increase of modulation order is significant when the bandwidth is low and becomes irrelevant for high bandwidth, with $M = 4$ and $M = 8$ yielding the same transmission time if $B = 2$ GHz is employed.

Comparison with other industrial wireless solutions

A direct comparison of the proposed PHY with other industrial wireless standards and proposal is not feasible, since the latter include a complete protocol stack, while the former is limited to the PHY only. However, in order to provide a qualitative comparison, a relation between packet transmission time and SU for WirelessHP can be stated as

$$T_{SU} = (1 + \eta) \cdot T_{pkt} \tag{6.17}$$

where η is an overhead (expressed as fraction of the packet transmission time) that accounts for the delays that are not strictly related to the transmission of payload bits, such as propagation time, processing and synchronization margins, ramp-up/ramp-down time, etc. The exact value of η depends on the upper layers that will be designed on top of the proposed WirelessHP PHY, however a typical value can range between 60 and 100%.

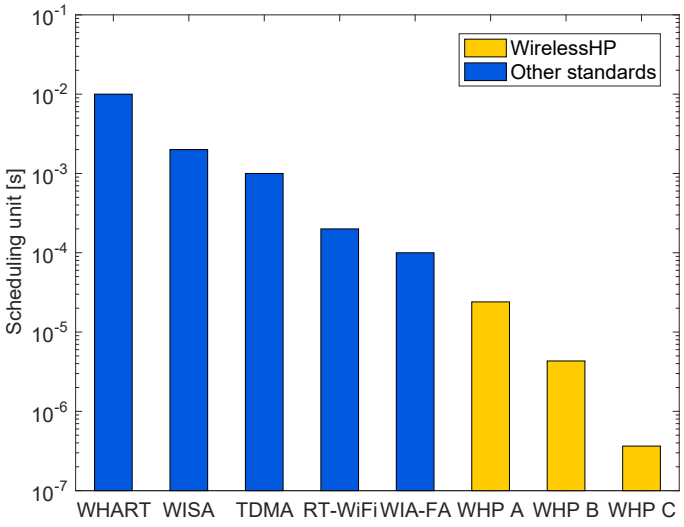


Figure 6.16: SU of WirelessHP implementations versus state–of–the–art industrial wireless standards and proposals.

The design of the upper layers will also impose the choice of specific parameter values for the WirelessHP PHY, such as modulation order M, code rate R_c, number of spatial streams N_{ss} and bandwidth B. In order to provide an exhaustive representation, the SU for a payload of $L = 100$ bits and three specific set of parameters is considered:

A) A WirelessHP implementation in the 2.4/5 GHz band, with $M = 2$, $R_c = 1$, $N_{ss} = 1$, $B = 20$ MHz and $\eta = 100\%$ was chosen as the representation of a low–performance system. The optimized parameters $N_{FFT} = 32$, $N_{gsc} = 6$ and $N_{cp} = 8$ yield a packet transmission time of $T_{pkt} = 12$ μs, and hence a SU of $T_{SU} = 24$ μs.

B) A WirelessHP implementation in the 2.4/5 GHz band, with $M = 4$, $R_c = 1$, $N_{ss} = 1$, $B = 160$ MHz and $\eta = 80\%$ was chosen to represent a medium–performance case. The optimized parameters $N_{FFT} = 64$, $N_{gsc} = 12$ and $N_{cp} = 64$ yield a packet transmission time of $T_{pkt} = 2.4$ μs, and hence a SU of $T_{SU} = 4.32$ μs.

C) A WirelessHP implementation in the mmWave band, with $M = 8$, $R_c = 1$, $N_{ss} = 1$, $B = 2$ GHz and $\eta = 60\%$ was chosen as the representation of a high–performance system. The optimized parameters $N_{FFT} = 128$, $N_{gsc} = 24$ and $N_{cp} = 100$ yield a packet transmission time of $T_{pkt} = 228$ ns, and hence a SU of $T_{SU} = 364.8$ ns.

The SU achieved with the three proposed WirelessHP implementations (WHP A, WHP B and WHP C) is compared with the slot time of several industrial wireless standards

and proposals in Fig. 6.16. Specifically, the list include WirelessHART (WHART), which has a minimum slot time of 10 ms (WirelessHART), WISA (Scheible et al., 2007) (2 ms), the modification to IEEE 802.15.4 proposed in (Lennvall et al., 2016) and indicated with TDMA (1 ms), the RT-WiFi proposed in Wei et al. (2013) (200 μs) and the WIA-FA standard (100 μs) (IEC 62948-2017).[10]

It can be noticed that implementations based on WirelessHP offer a significant improvement in terms of SU (and hence latency) with respect to the state–of–the–art industrial wireless solutions (the plot is in logarithmic scale), even when the basic 20 MHz PHY is considered with $M = 2$ modulation and $\eta = 100\%$ overhead (WHP A). Moreover, it is evident that the migration to the mmWave spectrum (implementation WHP C) allows a notable SU reduction, allowing to approach very closely the target of 200 ns.

Experimental validation

The WirelessHP PHY layer design presented in this section has been implemented on an experimental demonstrator based on Software Defined Radio (SDR) platforms. Although the adopted hardware was limited to work in the sub–4 GHz spectrum and with a very low bandwidth of 5 MHz, the obtained results provide a first proof of the feasibility of the proposed design.

The experimental setup adopted in this section is schematically represented in Fig. 6.17. Two Ettus USRPs model N210, each mounting an SBX-40 daughterboard allowing operations in the 0.4-4.4 GHz band, are used to setup a unidirectional wireless link. Each USRP is connected to a Windows PC through a Gigabit Ethernet cable. Two Matlab programs were developed to run on the Windows PCs: *OFDM_TX* handles the generation and encoding, while *OFDM_RX* performs the baseband processing and decoding of WirelessHP packets. For debugging purposes, a Tektronix MDO4104-6 oscilloscope is attached to the antenna of the transmitting USRP through a power splitter and a Rohde&Schwarz FSH6 spectrum analyzer is employed to monitor the wireless medium.

The USRP N210 platforms employ a Xilinx Spartan 3A-DSP Field–Programmable Gate Array (FPGA) module which performs digital/analog conversion and sample buffering, but cannot perform the baseband processing. This limitation, coupled with the limited real–time performance of the Windows PCs, restricted the achievable sample frequency F_s and hence the bandwidth B. Indeed, with the considered setup, it was only possible to run the tests at $B = 5$ MHz. At the receiving side, time domain samples were

[10] The WIA-FA standard does not specify a minimum slot time, however some preliminary implementations reported a value of 100 μs.

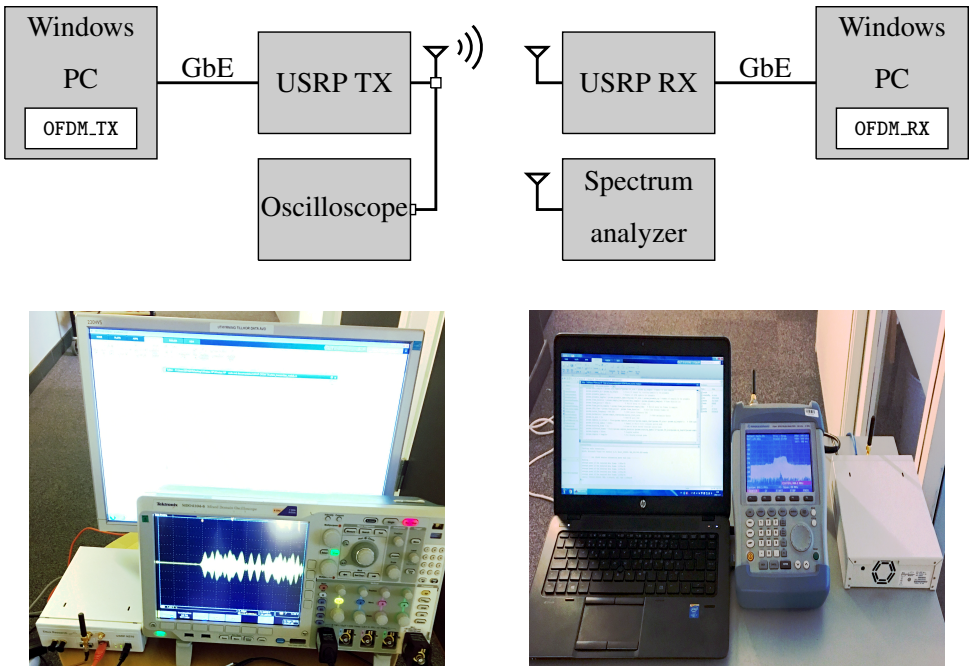


Figure 6.17: Narrowband experimental demonstrator based on USRP N210.

recorded and processed at non–real–time speed, due to the limited processing power of the PC. In order to avoid interference from other co–located WLANs, the tests were performed in the 868 MHz unlicensed spectrum, where a conservative assumption of $T_{ds}^{max} = 400$ ns was considered for the delay spread.

During the tests, the transmitting node sent WirelessHP packets with a PSDU size of $L = 104$ bits and a repetition period of $T = 100$ μs. The values of the other PHY layer parameters were mostly the same as in the theoretical evaluation: a single spatial stream ($N_{ss} = 1$), $N_{psc} = 4$ pilot subcarriers, $N_{dcsc} = 1$ DC null subcarrier and a minimum guard bandwidth ratio of $GBR_{min} = 0.1875$. While in the theoretical analysis channel coding was not considered, here a high–rate convolutional channel coding scheme was applied ($R_c = 5/6$), which, however, did not influence the results in terms of packet transmission time since the additional bits used for coding would be padded if coding were not employed. Only low–orderPSK modulations were considered, namely BPSK ($M = 2$), Quadrature Phase–Shift Keying (QPSK) ($M = 4$) and 8-PSK ($M = 8$). The optimized packet durations obtained through the WirelessHP design are $T_{pkt} = 47.6$ μs, 27.2 μs and 20.4 μs for BPSK, QPSK and 8-PSK respectively. These values were obtained by using an optimized FFT size of $N_{FFT} = 32$ with $N_{gsc} = 6$ guard subcarriers and $N_{cp} = 2$

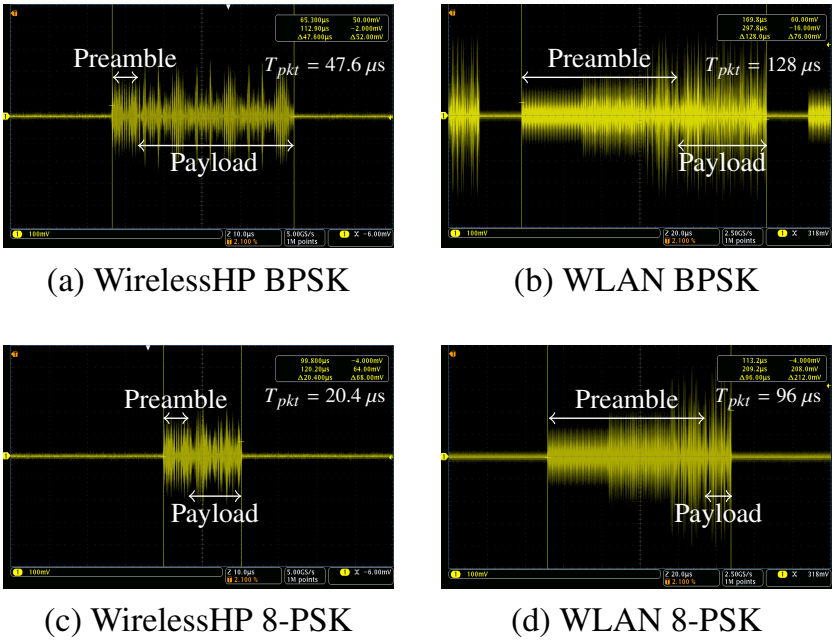


Figure 6.18: Captured WirelessHP and WLAN (IEEE 802.11g) waveforms for $L = 104$ bits packets transmitted with BPSK and 8-PSK with 5 MHz bandwidth.The zoom level in the WLAN waveforms (b) and (d) is doubled with respect to the WirelessHP waveforms (a) and (c).

samples for CP.

A first set of results is presented in Fig. 6.18, where the waveforms captured by the oscilloscope when BPSK and 8-PSK were used for the transmission of WirelessHP packets are presented. The figure also includes the waveforms obtained if the IEEE 802.11g PHY were used instead of the proposed one.[11] It can be noticed that there is a significant reduction in packet transmission time from IEEE 802.11 to WirelessHP: almost 3 times for BPSK (47.6 μs vs. 128 μs) and almost 5 times for 8-PSK (20.4 μs vs. 96 μs). The decrease in packet transmission time is in great part due to the reduction of the preamble, which can be observed to be quite dominant in the IEEE 802.11 waveforms (especially when 8-PSK is employed).

The reliability level of the proposed WirelessHP PHY is presented in Fig. 6.19, which shows the results of an experimental campaign aimed at assessing the BER at different

[11] In order to provide a fair comparison, the same bandwidth $B = 5$ MHz was used for WirelessHP and IEEE 802.11g: the parameters for IEEE 802.11g with 5 MHz bandwidth are reported in the first row of Tab. 6.5. Due to the increased packet transmission time with IEEE 802.11, the packet transmission period has also been increased from 100 μs to 150 μs.

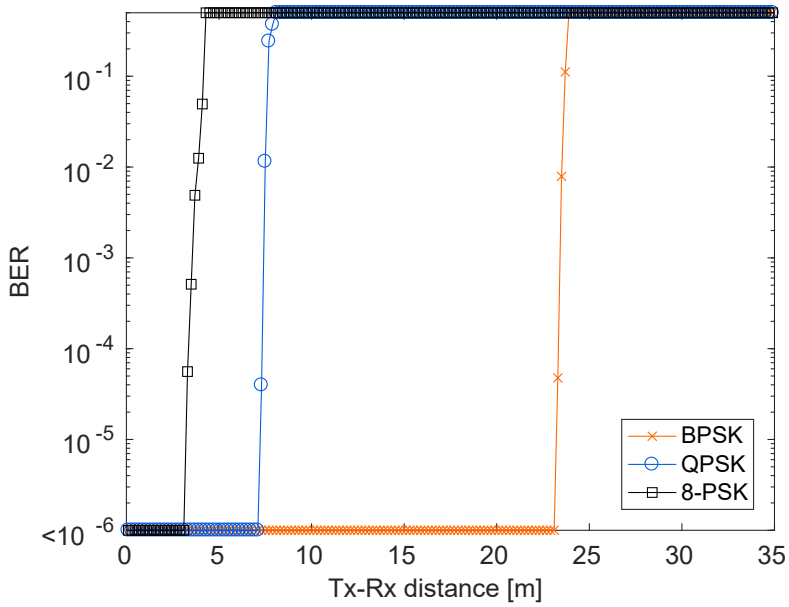


Figure 6.19: BER vs. communication distance for the 5 MHz WirelessHP PHY with different low–order modulations. Each measurement involved the transmission of 10^4 packets of 104 bits.

communication distances. In order to perform this campaign, a total of 10^4 WirelessHP packets (i.e., more than 10^6 payload bits) have been sent for each modulation order M and for different distances between transmitter and receiver, ranging from 0 to 35 meters with 0.2 meters granularity. The results show that the lowest modulation (BPSK) is able to achieve zero errors (which indicates a BER lower than 10^{-6}) up to more than 20 meters, whereas QPSK and 8-PSK show zero errors only up to 7 and 4 meters respectively. These results confirm that the proposed PHY is able to guarantee reliable communication within the short–range applications envisioned by the WirelessHP scenario. Specifically, it is confirmed that a preamble of only one OFDM symbol is sufficient to support reliable packet reception.

A further experimental result, presented in Fig. 6.20, is concerned with the robustness of the proposed WirelessHP PHY to CFO. In order to perform this test, a controlled mismatch was artificially inserted between the center frequencies of the transmitting and receiving USRPs, with a maximum value of ±45 ppm. The tests have been performed with a fixed distance between transmitter and receiver of 1.5 meters and again involved the transmission of 10^4 packets for each frequency offset and modulation value. The results show that all the considered modulations do not exhibit any significant error until roughly ±30 ppm, which represents a good robustness level.[12] When an higher offset is

[12]It has to be noted that devices compliant to the IEEE 802.11 standard in the 2.4 GHz band must have a maximum tolerance on clock frequency of ±20 ppm (IEEE 802.11-2016).

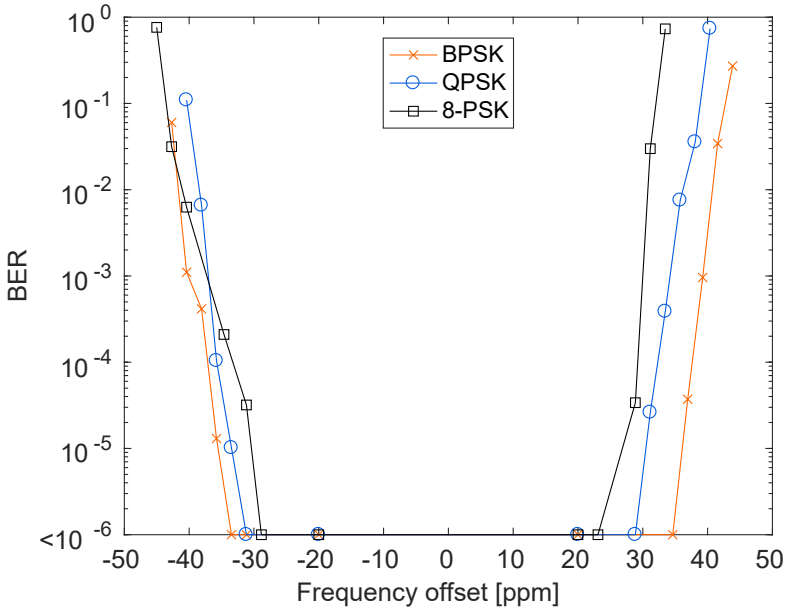


Figure 6.20: BER vs. artificially introduced frequency offset between transmitter and receiver oscillators for the 5 MHz WirelessHP PHY with different low–order modulations and a fixed communication distance of 1.5 meters.

introduced, the higher–order modulation (8-PSK) is the most sensitive to synchronization errors, as expected.

6.5 Concluding remarks and future activites

The most critical industrial control applications require ultra high–performance wireless networks, targeting at Gbps data rate and 10 μs–level cycle time, that are named as WirelessHP. An analysis of currently available and future wireless standards suggested that none of them is able to meet the required performance. Consequently, a clean–slate system design is necessary, building on emerging ideas and trends proposed in the scientific literature. To this aim, fundamental directions towards joint PHY and MAC layer design of such a system have been outlined.

Subsequently, a preliminary proposal for PHY design has been presented, describing a low–latency PHY for WirelessHP, specifically aimed at reducing latency through the minimization of transmission time for very short packets. The design is based on the IEEE 802.11 OFDM PHY, but it is significantly optimized by both reducing the PHY layer preamble and optimizing the OFDM parameters. Theoretical analysis shows that the proposed PHY is able to greatly reduce the packet transmission time with respect to IEEE 802.11, down to almost 1 μs in the 2.4/5 GHz band and 200 ns in the mmWave

band for 100 bits packets. The feasibility of the proposed design is confirmed through an SDR–based experimental demonstrator. With this platform, reliable communication can be established up to 20 meters distance if BPSK is employed.

However, the demonstrator platform presents several limitations. As a first issue, the current version of the demonstrator is only able to work at a limited bandwidth (5 MHz), which does not allow a significant reduction of the packet transmission time. Upgrading the available hardware would allow to increase the bandwidth and speed up the packet exchange by performing all the baseband processing in the FPGA. Moreover, when the SU becomes very short (some μs or lower), the packet transmission time may not be the dominant source of delay and some other aspects, such as synchronization and processing times, must be managed. Furthermore, if operation in the mmWave band is considered, directional communication should be established through beamforming techniques in order to overcome the high path–loss, requiring the development of high–accuracy beamforming strategies. Finally, the reliability level reported achieved by the demonstrator (BER lower than 10^{-6}), although high, may not be good enough for the most critical industrial applications, which could require a BER of 10^{-10} or lower. In order to achieve this target, methods for increasing reliability should be considered, such as enhanced channel coding, spatial diversity schemes, improved channel estimation and synchronization algorithms and optimized analog RF front end stages. The effect of packet size on reliability will also be investigated by taking into account other possible packet lengths typical of specific industrial control scenarios.

Finally, the design of a low–latency PHY is just the first step towards the realization of a complete WirelessHP stack. In the future, a thorough optimization of the upper layers must be carried out, to ensure deterministic communication and higher reliability. Moreover, collaboration from industry and academia should be sought, towards the realization of dedicated WirelessHP devices, the definition of new industrial wireless standards to provide the required performance, and, possibly, the allocation of dedicated frequency bands for such applications.

7

LoRaWAN for Industrial IoT

LPWANs have recently emerged as appealing communication systems in the context of the IoT. Particularly, they revealed effective in typical IoT applications such as environmental monitoring and smart metering. Such networks, however, have a great potential also in the industrial scenario and, hence, in the context of the IIoT, which represents a dramatically growing field of application. This chapter focuses on a specific LPWAN, namely LoRaWAN, and provides an assessment of its performance for typical IIoT employments such as those represented by indoor industrial monitoring applications. In detail, after a general description of LoRaWAN, a discussion on how to set some of its parameters in order to achieve the best performance in the considered industrial scenario is carried out. Subsequently the outcomes of a performance assessment, based on realistic simulations, are presented, aiming at evaluating the behavior of LoRaWAN for industrial monitoring applications. Moreover, a comparison with the IEEE 802.15.4 network protocol, often adopted in similar application contexts, is proposed.

7.1 Industrial IoT and LPWANs

The general vision of IoT and the benefits that it could bring in the industrial world are discussed at the beginning of this section. Subsequently, the main LPWAN solutions are briefly discussed, with a specific focus on LoRaWAN.

IoT and Industrial IoT

The IoT concept was introduced towards the turn of the century (Ashton) to indicate an interconnected system of uniquely identifiable objects equipped with Radio Frequency Identification (RFID) technology. Nowadays the IoT paradigm has expanded, embracing a wide range of communication technologies, software architectures and applications, and can be best defined as *"a network of networks where, typically, a massive number of objects/things/sensors/devices are connected through communications and information infrastructure to provide value–added services"* (Perera et al., 2015). Over the past decade, several IoT solutions have been developed by both industry and academia for various kinds of applications, including smart wearables, smart home and smart cities, to name a few (Perera et al., 2015).

In the next years, the IoT vision is expected to be applied not only to the consumer market but also to productive sectors, dramatically changing manufacturing, energy, transportation, agriculture and other industrial applications, in what has already been termed as IIoT (Wan et al., 2016). According to a report by the World Economic Forum (O'Halloran and Kvochko, 2015), the IIoT revolution will impact economical sectors that account for nearly two-thirds of the global Gross Domestic Product (GDP), changing the basis of competition and redrawing industry boundaries. New connected ecosystems will emerge, allowing significant improvements in operational efficiency as well as the advent of a new outcome economy, where companies no longer deliver products and services, but rather measurable results that create value for their customers (O'Halloran and Kvochko, 2015).

The deployment of IIoT solutions is a complex process that, as addressed in Xu et al. (2014) where a functionality–based architecture for IIoT is proposed, impacts on several disciplines, such as communication and computer science.

Communication architectures for IIoT

From a traditional communication aspect, IoT encompasses several heterogeneous systems such as LANs, WSNs, cellular networks, mesh and ad hoc networks, whose interoperability is ensured by the common use of existing Internet protocols, such as IPv6 (Xu et al., 2014).

Moving to the industrial scenario, applications often have stringent QoS requirements, in terms of robustness, reliability, latency, determinism, energy efficiency and security. Therefore, a careful selection of the most appropriate network for a specific application is necessary in order to meet those requirements and provide effective IIoT solutions. Also, according to Mumtaz et al. (2017), the true potential of the IIoT paradigm can be unlocked

only when a wireless communication architecture is envisioned. As a consequence, it is necessary to analyze the suitability of different wireless networks in view of their deployment in IIoT applications. To this regard, several wireless communication systems have been considered for IoT applications (Tayeb et al., 2017).

These wireless networks range from very short–range solutions such as Near Field Communication (NFC), to extremely long–range ones such as Worldwide Interoperability for Microwave Access (WiMAX); from low–power technologies such as BLE, to high–power ones as cellular networks (2G/3G/4G). The several amendments to the IEEE 802.11 standard for WLANs and to the IEEE 802.15.4 standard for WPANs are also highly regarded. Moreover, dedicated wireless networks for industrial applications are considered, such as WirelessHART (WirelessHART) and ISA 100.11a (ISA-100.11a-2009), and are currently being expanded to guarantee IP support under the 6TiSCH family of standards (Dujovne et al., 2014).

In addition to the aforementioned solutions, LPWANs have recently emerged in the IoT scenario (Raza et al., 2017), the most popular being Narrowband Internet of Things (NB-IoT), SigFox, Ingenu, Weightless and LoRaWAN. These networks, that are available on licensed as well as unlicensed bands combine a very long communication range (up to several kms) with extremely long battery life, at the cost of a limited throughput.

LPWANs for indoor industrial monitoring

Nowadays, LPWANs are mostly used for outdoor monitoring applications, such as environmental monitoring (Guibene et al., 2017) and smart metering (Varsier and Schwoerer, 2017). However, their features are appealing for IIoT applications as well, and hence they have been very recently started to be considered also in this scenario (Mumtaz et al., 2017; Sanchez-Iborra and Cano, 2016). Indeed, the significantly high energy efficiency of LPWAN devices can reveal truly interesting for cost–effective IIoT deployments. Moreover, the remarkable communication robustness that allows LPWANs to achieve long–range communications can be useful in industrial applications where the wireless channel is often impaired by multipath and fading (Willig et al., 2005), thus giving them an edge against other low–power wireless technologies.

The above considerations represent the main motivation of the study carried out in this chapter which, basically, addresses the use of LPWANs, and in particular of LoRaWAN, for indoor industrial monitoring applications and supports this claim with a thorough performance assessment that compares LoRaWAN with a reference wireless solution for industrial monitoring, namely the IEEE 802.15.4 standard for WPANs, with

respect to several metrics of interest. To this aim, an accurate simulation model for LoRaWAN networks is developed starting from the work in Magrin et al. (2017), including a realistic channel model for indoor industrial environments.

The possible applications of LoRaWAN to indoor industrial IIoT applications has not been investigated deeply in the scientific literature so far. Indeed, in Haxhibeqiri et al. (2017) a case study is presented that addresses a specific indoor LoRaWAN industrial application, whereas in Margelis et al. (2015) some possible industrial applications of LPWANs are described in general. In Neumann et al. (2016) indoor applications of LoRaWAN are considered, but not for industrial scenarios. In Hernandez et al. (2017), other LPWAN solutions are addressed for IIoT and, finally, Guibene et al. (2017), Magrin et al. (2017) and Petäjäjärvi et al. (2017) deal with the LoRaWAN performance in outdoor scenarios.

Overview of LPWANs

LPWANs are designed to offer affordable connectivity to a high number of low–power devices distributed over large geographical areas. In this section, the most widespread LPWAN solutions are discussed, with a particular focus on LoRaWAN.

LoRa and LoRaWAN

LoRaWAN is an open network standard (LoRaWAN v1.0-2015) developed by the LoRa Alliance, which mainly defines the MAC layer and message formats. It is based on LoRa, a proprietary PHY layer developed by Semtech Corporation and derived by Chirp Spread Spectrum (CSS) modulation. In this technology, each symbol is spread in a fixed bandwidth, B, and the time duration of the symbol is varied according to an index called Spreading Factor (SF) which can range between 7 and 12. Consequently, the duration of a symbol varies from $\frac{1}{B} \times 2^7$ and $\frac{1}{B} \times 2^{12}$. This spreading technique allows to recover data even when the received power is very low (also under the noise level), thus offering very robust communication, at the cost of a reduced data rate (Goursaud and Gorce, 2015), which does not exceed 21.9 Kbps. Moreover, transmissions with different SFs are somewhat orthogonal to each other, increasing network capacity.

LoRaWAN networks are deployed in the unlicensed ISM bands of 863-870 MHz in Europe, and 902-928 MHz in the US. According to the regulations, in these bands the transmitting devices must limit their maximum power to 14 dBm (27 dBm in the 869.4-869.65 MHz sub–band) and adopt either a duty–cycled transmission (0.1, 1 or 10 percent according to the sub–band) or a Listen–Before–Talk/Adaptive–Frequency–Agility (LBT/AFA) behavior.

A LoRaWAN network, includes three types of entities, namely End Devices (EDs), Gateways (GWs) and Network Server (NS). EDs are typical field devices that collect sensor information from the field and, possibly, send commands. They are connected (via wireless links) to one or more GWs that, in turn, are connected (either through a wired or cellular backhaul link) to a single NS, which manages the entire network and originates downlink transmissions (if any). There is no exclusive association between EDs and GW and the same uplink message can be received by several GWs with different signal qualities.

The LoRaWAN specifications define three functional classes, namely Class A, B and C, with the first one being mandatory for all LoRaWAN EDs. Class A EDs access the channel in a random fashion, following an ALOHA–like scheme, and open (at most) two reception windows at predefined slots in time and frequency after each uplink transmission, whereas they remain in sleep mode for the rest of the time. Class B and C devices differentiate mostly for their management of receive windows: class B EDs can open them at scheduled time intervals (they are synchronized with the NS by means of beacon messages broadcasted by the GWs), while class C ones keep them always open, clearly sacrificing energy efficiency for low latency. Finally, authentication and encryption mechanisms at different levels (device, network and application) are envisioned by LoRaWAN specifications to ensure the integrity and security of communications.

Other LPWAN solutions

Besides LoRaWAN, several other LPWAN technologies are available in both licensed and unlicensed bands.

The former includes mainly NB-IoT, which is part of 3GPP Release 13. NB-IoT reuses the LTE design, adopting OFDMA for downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) for uplink transmissions, with resource blocks of 15 kHz $\times$ 0.5 ms and a maximum bandwidth of 180 kHz, offering a peak data rate of 250 Kbps. NB-IoT can be deployed in–band, using LTE resources, in the guard band between two LTE bands, or as stand–alone, by replacing a 200 kHz GSM carrier.

In the unlicensed band, the most widespread LPWANs besides LoRaWAN are SigFox, Ingenu and Weightless. The first one is a proprietary protocol based on Ultranarrow Band (UNB) modulation, with data rate limited to 100 bps in uplink and 600 bps in downlink. SigFox works in the same bands as LoRaWAN and limits the operation of the connected devices to 140 uplink messages and 4 downlink messages per day. Ingenu is also a proprietary technology which works in the 2.4 GHz band, and adopts a patented Random Phase Multiple Access (RPMA) scheme for uplink transmissions,

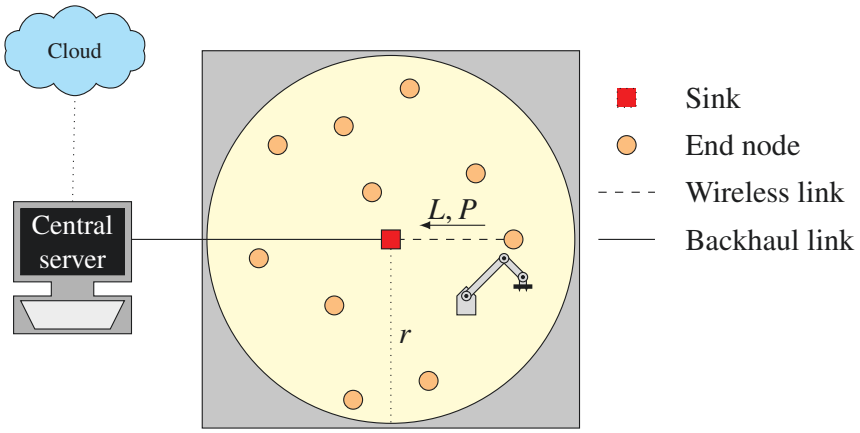


Figure 7.1: A schematic representation of the indoor industrial monitoring scenario considered in this chapter.

with a maximum data rate of 78 Kbps on 40 different 1 MHz wide channels, each of which can host up to 1200 orthogonal signals, thanks to RPMA. Finally, the Weightless SIG proposes three different standards (Weightless-W, Weightless-N and Weightless-P) deployed in different bands (TV white space and sub–GHz ISM bands), offering different data rates (from 200 bps to 10 Mbps) and employing different modulation schemes (UNB, QAM and PSK) and different channel access methods (ALOHA, TDMA and FDMA).

Although all these LPWANs can reveal interesting, the analysis carried out in this chapter focuses exclusively on LoRaWAN, since it is a really promising and popular network with interesting features. Particularly, i) it operates in an unlicensed band, ii) its MAC layer protocol is completely open and iii) it can handle an unlimited number of packets.

The indoor industrial monitoring scenario

The reference scenario considered in this chapter is represented in Fig. 7.1. As can be seen, it refers to a monitoring network composed by N devices (end nodes) deployed in a building where an industrial process is taking place. The devices are distributed within a circular area of radius r and periodically sample different physical quantities that allow to monitor the state of the process. Each node sends the updated sample value as a message of L bytes, with a transmission period of P seconds, to a sink node installed at the center of the building, which will be a GW in a LoRaWAN network or, more generally, a PAN coordinator in a WPAN. The sink will in turn send the data received from the end nodes to a central server, that allows authorized users to access the most updated value sampled by each node, either locally or through cloud interfaces. This configuration resembles that

Table 7.1: Parameters for the indoor industrial monitoring scenario

Parameter	Description	Value
N	Number of nodes	$\{10, \ldots, 1000\}$
r	Coverage radius	200 m
L	Message length	50 bytes
P	Transmission period	$\{60, \ldots, 1800\}$ s

of IWSNs deployed in monitoring systems (Gungor et al., 2014; Lo Bello et al., 2017). Clearly, more complex configurations could be addressed using this type of networks. For example, different applications in which multiple sinks are used could be envisioned, where the nodes can transmit packets of different length with different periods. However, in this preliminar study, a basic application of LoRaWAN to indoor IIoT is considerd, to investigate its effectiveness, leaving the assessment of more complicated scenarios to future works. For the same reason, it is assumed that only uplink transmissions are performed, and that they are not confirmed, meaning that acknowledgement packets are not sent by the sink nodes and, hence, there are no retransmissions.

Tab. 7.1 summarizes the values of the parameters considered in this scenario. It is assumed that the number of nodes in the network ranges from 10 to 1000, while the coverage radius is fixed to 200 meters. Since sensor measurements typically occupy a few bytes, the message length in these applications is generally low, with 50 bytes being a reasonable value considering that also other information (e.g., timestamp, node status, battery life, etc.) may be appended. Finally, the transmission period depends on the dynamics of the sampled quantities and, in the considered scenario, ranges from one minute to 30 minutes. It is worth to notice that the corresponding sampling rates are considerably lower than those encountered in other industrial applications (which can be up to some kHz), since this scenario is targeted only at the on–line monitoring of industrial processes (Iqbal et al., 2017), not aiming at real–time control.

The assessment carried out in this chapter is based on three key performance indicators. The first one, related to the communication reliability, is the Probability of Success (PoS), i.e. the percentage of packets sent by the end nodes which are received correctly by the sink. A second important metric, related to the latency and determinism of the communication, is the Interpacket Time (IPT), namely the interval between two consecutively received packets at the sink pertaining to the same node. More formally, indicating as $r_i(k)$ the time at which the sink receives the k–th packet from end node i, $i = 1, \ldots, N$, the IPT

for node i is the following varying quantity

$$IPT_i(k) = r_i(k+1) - r_i(k) \tag{7.1}$$

Considering an observation period during which the sink has received K_i packets from node i, the Mean Interpacket Time (MIPT) can be computed as

$$MIPT_i(k) = \frac{1}{K_i - 1} \sum_{k=1}^{K_i - 1} IPT_i(k) \tag{7.2}$$

Then, the Global Interpacket Time (GIPT) can be computed averaging the MIPT over all the nodes in the network.

The third and last performance metric is the Average Energy Consumed (AEC) by each end node, which provides insights on the energy efficiency and allows to forecast the battery life of end nodes.

7.2 A realistic LoRaWAN industrial model

While it is possible to find accurate models of LoRaWAN networks in the scientific literature (Magrin et al., 2017), none of them tackles the peculiar features of industrial environments. In this section, a realistic model for LoRaWAN networks deployed in IIoT applications is hence presented.

Channel model

In this chapter, indoor LoRaWAN networks deployed in the 863-870 MHz ISM band will be considered, although the channel model is also applicable to the 902-928 MHz ISM band adopted in US. The model should take into account all the impairments that can be present inside industrial buildings, which can be divided in two categories, namely large–scale effects and small–scale effects.

Large–scale effects include path loss and shadowing, so that the total power loss $L(d)$, is a function of the distance d between transmitter and receiver. It can be expressed (in dB) as

$$L(d)^{dB} = PL(d)^{dB} + \chi_\sigma^{dB} \tag{7.3}$$

where the shadowing term, χ_σ^{dB}, is generally modeled as a zero–mean Gaussian random variable with standard deviation σ. The path loss term, $PL(d)^{dB}$, instead, is often modeled as a fixed term plus a logarithmic function of the distance d multiplied by a coefficient η (path loss exponent), whose value ranges between 2 in ideal conditions (i.e.

line–of–sight and free space) and 4 in NLOS conditions. A common approach is to define a breakpoint distance, referred to as d_1, beyond which the propagation becomes NLOS, yielding the following path loss model

$$PL(d)^{dB} = \begin{cases} 0, & d < d_0 \\ PL_0 + 10\eta_0 \log\left(\frac{d}{d_0}\right), & d_0 \leq d \leq d_1 \\ PL_0 + 10\eta_0 \log\left(\frac{d_1}{d_0}\right) + 10\eta_1 \log\left(\frac{d}{d_1}\right), & d \geq d_1 \end{cases} \tag{7.4}$$

where the parameters PL_0, d_0 and η_0, together with the shadowing standard deviation σ can be extracted from Ai et al. (2015), which reports real–world measurements in the 900 MHz band from different indoor industrial environments. In this scenario, the breakpoint distance d_1 is set to 100 m (half of the coverage radius) and the path loss exponent in NLOS conditions is set to the typical value $\eta_1 = 4$.

The small–scale effects mostly refer to fading, which is typically modeled through a Rayleigh/Rician distribution, whose only parameter is the K–factor K. The work in Ferrer-Coll et al. (2013) reports some K–factor measurements in indoor industrial environments at 868 MHz, whose results have been used for the simulations presented in this chapter.

Link performance model

Taking into account all channel impairments, the received power at the sink in dB can be hence modeled as

$$P_{rx}^{dBm} = P_{tx}^{dBm} - L^{dB} - F^{dB} + G_{tx}^{dB} + G_{rx}^{dB} \tag{7.5}$$

where P_{tx}^{dB} is the transmitted power in dB, G_{tx}^{dB} and G_{rx}^{dB} are the transmit/receive gains in dB respectively, F^{dB} is a margin that accounts for fading and L^{dB} is the total loss due to large–scale effects reported in Eq. (7.3). The SNR can be immediately derived from Eq. (7.5) by subtracting the thermal noise power in dB. According to LoRa chipset specifications (SEMTECH SX1272), for each selected SF, k, a minimum SNR level R_k is requested to achieve a correct demodulation. These values are reported in Tab. 7.2, together with the data rate of each SF and the time required to transmit a 50 bytes message (assuming 125 kHz bandwidth and 4/5 code rate).

Besides path loss, shadowing and fading, the other significant impairment in wireless channels is interference. LoRa CSS modulation is quite robust to external interference from non–LoRa signals (Goursaud and Gorce, 2015), whereas the interference between

Table 7.2: Characteristics of LoRa spreading factors

SF	Required SNR	Data rate	TX time (50 bytes)
7	-7.5 dB	5.47 Kbps	99.58 ms
8	-10 dB	3.13 Kbps	178.69 ms
9	-12.5 dB	1.76 Kbps	336.90 ms
10	-15 dB	0.98 Kbps	632.83 ms
11	-17.5 dB	0.44 Kbps	1183.74 ms
12	-20 dB	0.25 Kbps	2203.6 ms

different LoRa transmissions strongly depend on their SF. In Goursaud and Gorce (2015) it is reported the required Signal–to–Interference Ratio (SIR) ratio in dB, $T_{k,l}$, to allow a correct decoding of a transmission with SF k when an another transmission with SF l is interfering. However, since in typical LoRaWAN networks several concurring transmissions with different SFs are present, evaluating the impact of each interfering SFs may be computationally inefficient. Consequently, an approach based on Equivalent Signal–to–Interference plus Noise Ratio (ESINR) is detailed in the following.

Consider a transmission with SF k and an interfering transmission with SF l, where the SIR is $SIR_{k,l}^{dB}$. A correct decoding is achieved if $SIR_{k,l}^{dB} \geq T_{k,l}$, and also if the SNR is above R_k. It can be hence said that a SIR level of $T_{k,l}$ is "equivalent" to a SNR level of R_k from a system performance perspective. This allows to transform any SIR value with respect to an interfering SF l in an Equivalent Signal–to–Interference Ratio (ESIR) value as

$$ESIR_{k,l}^{dB} = SIR_{k,l}^{dB} + R_k - T_{k,l} = SIR_{k,l}^{dB} + E_{k,l} \tag{7.6}$$

where the matrix E (in dB) is obtained from the values R_k in Tab. 7.2 and the matrix T in Goursaud and Gorce (2015) as

$$\mathbf{E} = \begin{bmatrix} -13.5 & 8.5 & 10.5 & 11.5 & 11.5 & 12.5 \\ 14 & -16 & 10 & 12 & 12 & 12 \\ 14.5 & 14.5 & -18.5 & 10.5 & 12.5 & 12.5 \\ 15 & 15 & 15 & -21 & 11 & 13 \\ 15.5 & 15.5 & 15.5 & 15.5 & -23.5 & 11.5 \\ 16 & 16 & 16 & 16 & 16 & -26 \end{bmatrix} \tag{7.7}$$

In other words, Eq. (7.6) allows to "normalize" the SIR between two interfering transmissions with arbitrary spreading factors, transforming it in a quantity comparable with other SNR values and hence allowing to sum them together. This is achieved through the

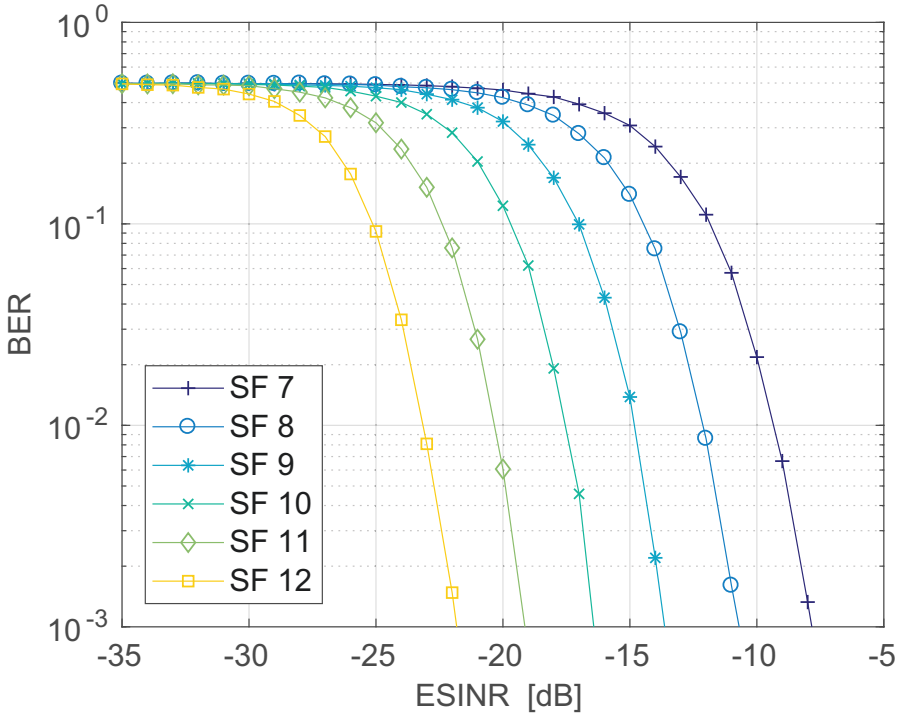


Figure 7.2: BER vs. ESINR (in dB) curves for different LoRa SFs obtained through Matlab simulations.

term $E_{k,l}$, which represents a sort of weight indicating the impact of the interfering SF l on the transmission with SF k. It is worth to observe that this impact is much higher if the two SFs are the same and becomes less relevant as the the two SFs are distant, thus allowing concurrent transmissions to take place.

With the proposed model based on ESINR, the PER of a transmission performed with SF k can be evaluated. First, a single transmission is divided in "chunks", each characterized by a specific set of interferers. The BER during each chunk can be computed through the following operations:

1. The received power P_{rx}^{dBm} is computed according to Eq. (7.5) and the SNR is derived.

2. The SIR for each interfering SF l is derived and the corresponding ESIR is computed according to Eq. (7.6).

3. The ESINR in dB is computed as

$$ESINR_k^{dB} = -10\log_{10}\left(10^{\frac{-SNR_k^{dB}}{10}} + \sum_{l=7}^{12} 10^{\frac{-ESIR_{k,l}^{dB}}{10}}\right) \tag{7.8}$$

4. The BER corresponding to the ESINR is derived according to the curves in Fig. 7.2, which are obtained through Matlab simulations of CSS performance.

Starting from the BER and the number of bits in the chunk, which can be approximated from the chunk duration and the data rate, the Chunk Error Rate (CER) is obtained. Finally, the PER can be derived as

$$PER_k = 1 - \prod_{n=1}^{N_{cks}} (1 - CER_{k,n}) \tag{7.9}$$

where N_{cks} is the number of chunks that make up the transmission.

Strategies for the choice of the spreading factors

One of the degrees of freedom in configuring a LoRaWAN network is to assign to each node its SF. In this chapter some techniques that allow a static assignment of the SFs are considered. It is worth mentioning, however, that the standard also allows for a dynamic strategy, called Adaptive Data Rate (ADR), which is not considered in this scenario since it requires downlink transmissions. Nonetheless, the techniques that are going to be presented can be easily extended to the dynamic case.

The simplest approaches are either to assign the same SF to all the nodes in the network, or to randomly distribute all the available SFs among them. A more refined approach, often adopted in other studies (Magrin et al., 2017), is to assign to each node the lowest SF for which the SNR at the sink is higher than the threshold defined in Tab. 7.2. However, as discussed in Reynders et al. (2017), when the maximum distance between nodes and sink is limited (as in the case of indoor environments), this strategy will always lead to all the nodes being assigned the lowest SF, i.e. 7 (which ensures the highest data rate), thus incurring in fairness problems and not fully exploiting the orthogonality of the different LoRa SFs.

To overcome this issue, an innovative strategy for the selection of SFs is proposed here, based on a constrained optimization procedure. Let $\mathcal{S}$ be the set of available SFs and s_i the SF assigned to node i. This procedure needs to fulfill two constraints:

A) the aforementioned constraint on the SNR, i.e.

$$SNR_i^{dB} \geq R_{s_i}, \quad i = 1, \ldots, N \tag{7.10}$$

B) an additional constraint on the transmit period P, i.e.

$$P \geq \frac{TX(s_i)}{DC_i}, \quad i = 1, \ldots, N \tag{7.11}$$

where $TX(s_i)$ is the transmission time for SF i as reported in Tab. 7.2 and DC_i is the duty cycle limitation of node i, which depends on the operating frequency band (ETSI EN 300 220-1).

After ensuring that these constraints are observed, the SFs are distributed in the most uniform possible way among the nodes, to maximize the orthogonality of transmissions. More formally, let $\mathcal{N}_k$ be the set of nodes for which SF k is assigned and let $\mathcal{N}_{min}$ and $\mathcal{N}_{max}$ be the minimum and maximum cardinality of the sets $\mathcal{N}_k$, i.e.

$$\mathcal{N}_{min} = \min_{k \in \mathcal{S}} |\mathcal{N}_k|, \quad \mathcal{N}_{max} = \max_{k \in \mathcal{S}} |\mathcal{N}_k| \tag{7.12}$$

For each node, a SF is chosen among the ones that respect the constraints of both Eq. (7.10) and Eq. (7.11), so that the difference between the minimum and maximum cardinality is minimized

$$\arg\min |\mathcal{N}_{max} - \mathcal{N}_{min}| \tag{7.13}$$

This strategy, which is indicated as "fair" in the following section, ensures that orthogonality of transmissions is maximized and that the nodes never exceed the duty cycle limitations.

Energy model

A fundamental aspect to address in a LoRaWAN network is the energy consumed by the end nodes. A LoRa ED can be in four possible states (SEMTECH SX1272): sleep, idle, transmitting and receiving. Specifically, a node is always in sleep state except when it transmits a packet and during the automatically opened receive windows, where it can be either idle or receiving (not in the case of the considered scenario, since there are no downlink transmissions).

Tab. 7.3 reports the current consumed by a LoRa ED in the four different states, as reported in SEMTECH SX1272. In order to compute the energy consumption of each ED, the current of each state has been multiplied by the supply voltage (3.3 V) and by the time that the ED passes in that specific state.

Table 7.3: Supply current for LoRa EDs in different states

State	Symbol	Value
Sleep	I_{sleep}	1.5 μA
Idle	I_{idle}	1.4 mA
Transmitting	I_{tx}	28 mA
Receiving	I_{rx}	11.2 mA

7.3 Performance evaluation in an industrial monitoring scenario

The LoRaWAN industrial model has been implemented in the popular ns3 network simulator (ns3) in order to assess the performance of this network in the considered indoor IIoT scenario.

Simulations setup

The starting point for the development of ns3 LoRaWAN simulations was the work in Magrin et al. (2017), in which an original LoRa module for ns3 was presented, allowing accurate simulation of uplink transmission in a LoRaWAN network. Several features have been integrated to this module to allow realistic simulations of LoRaWAN networks employed in IIoT applications. Specifically:

- An accurate channel model for indoor industrial buildings has been introduced, that accounts for path loss, shadowing and fading. Realistic channel model parameters were selected considering the experimental measurements reported in Ai et al. (2015) and Ferrer-Coll et al. (2013).
- The simplified link performance model of Magrin et al. (2017) has been expanded, adding the ESINR–based approach detailed in Sec. 7.2 as well as the Matlab–based BER vs. ESINR curves.
- A different interference model with respect to Magrin et al. (2017) has been used: in the old model, the power of a partially interfering transmission was "equalized" on the entire packet duration, whereas, as discussed in Sec. 7.2, in the adopted model a packet is divided in chunks, each one with a specific set of interferers and corresponding error rate.
- An energy model has been added to Magrin et al. (2017), allowing a precise computation of the energy consumed by the end nodes.

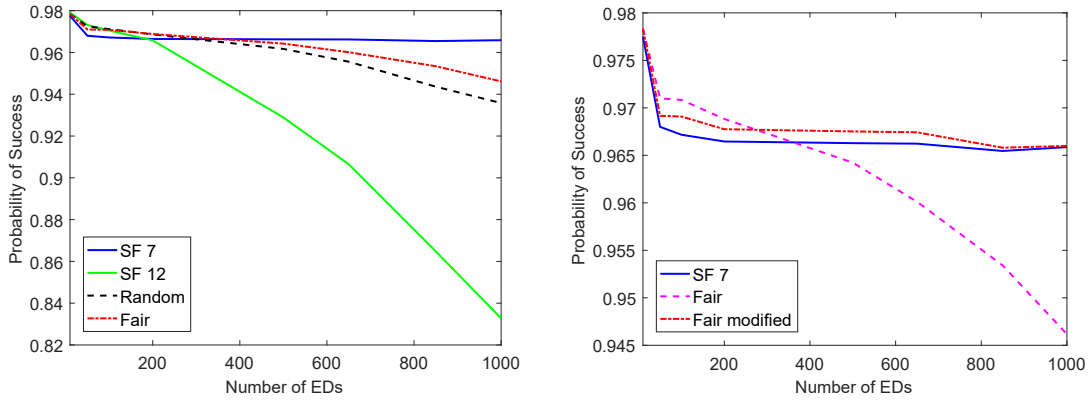
The performance figures of LoRaWAN have been also compared with those of a WPAN, which have been simulated exploiting the *lr-wpan* module included in the standard ns3 distribution, upgraded with some modifications. In particular, the 868 MHz BPSK PHY introduced in the IEEE 802.15.4-2006 standard (IEEE 802.15.4-2015) was considered, in order to provide a fair comparison with LoRaWAN, deployed in the same frequency band. This choice limited the achievable data rate of IEEE 802.15.4 to 20 Kbps. Moreover, the modifications proposed in Rege and Pecorella (2016), mostly concerning the introduction of an energy model, have been considered.

In all the simulations the reference scenario is that reported in Fig. 7.1, with the parameters shown in Tab. 7.1. Specifically, at each simulation, the end nodes are spread over a circle of radius r according to a uniform distribution and they transmit packets to the sink with a period P. Each node randomly selects an initial phase in the interval $[0, P]$ to avoid perfect synchronization of the transmission attempts. Finally, for each choice of parameters, the results were averaged over 10 different runs, each simulating the network performance for 2 hours and characterized by a specific realization of nodes positions and initial phases.

Tuning of a LoRaWAN network

A first assessment is concerned with the comparison of different strategies for the assignment of SFs in a LoRaWAN network used for indoor industrial monitoring applications. Four different strategies, defined in agreement with the analysis carried out in Sec. 7.2 have been considered, namely: two constant strategies in which all the nodes are assigned either the lowest spreading factor (SF 7) or the highest (SF 12); a random strategy in which SFs are randomly assigned to the nodes; the innovative strategy based on equal distribution of SFs presented in Sec. 7.2, called "fair".

Fig. 7.3a reports the PoS, computed over all the transmissions in the network, for the different SF assignment strategies and for different networks sizes, ranging from 10 to 1000 nodes. The transmission period for all nodes is fixed to 10 minutes. It can be observed, first, that the PoS is generally high, confirming the good robustness of LoRa modulation. The worst behavior is obtained when all nodes are assigned SF 12, especially as the network size increases, since the high transmission time with this SF (more than 2 seconds per packet) causes severe interference problems. Conversely, when all nodes are assigned SF 7, the PoS is almost constant to a very high value (about 97%), performing even better, when the number of nodes in the network is high, than the "fair" strategy, which instead yields a slightly higher PoS when there are less than 200 nodes. This result can be explained observing that SF 7 guarantees the lowest transmission time, thus



(a) All SF choice strategies

(b) Only SF 7, "fair" and "fair" restricted to SF 7, 8 and 9

Figure 7.3: PoS vs. number of EDs in a LoRaWAN network employed in IIoT applications for different SF choice strategies. The transmission period is fixed to 10 minutes.

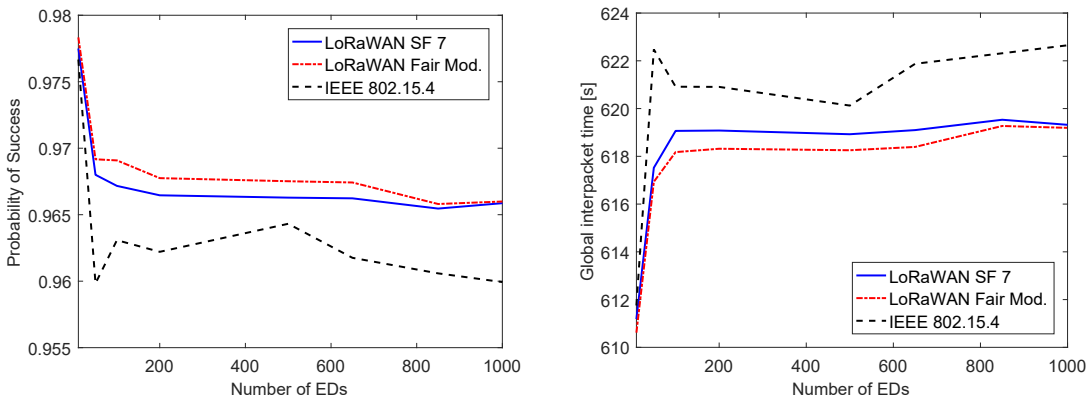
minimizing the interference; conversely, the "fair" strategy leverages the orthogonality of different SFs, but distributes all the available SF values (including the highest ones) among the nodes, increasing the transmission times, and hence ultimately increasing the probability a transmission is subjected interference.

To overcome this issue, a slightly modified version of the fair strategy is proposed, in which the set of available SFs $\mathcal{S}$ is limited to the three lowest ones (7, 8 and 9). The performance of this upgraded scheme can be observed in Fig. 7.3b, always relevant to PoS vs. number of EDs with $P = 10$ minutes. The modified fair strategy now outperforms the strategy in which all nodes are assigned SF 7 both when the number of nodes is low and when it is high, never falling below a PoS of 96.5%. As a result of these considerations, in the following of this evaluation, the only considered schemes for the selection of SFs in LoRaWAN are the "fair modified" strategy and the "constant to SF 7" one.

Comparison of LoRaWAN and IEEE 802.15.4

To support the claim that LoRaWAN can represent a good choice for IIoT monitoring applications, the performance of this network are compared with those of an IEEE 802.15.4–based one.

The configuration of the IEEE 802.15.4 network is equivalent to that of LoRaWAN. Moreover, the non–beacon version of the IEEE 802.15.4 protocol has been considered, in which nodes (that generate new packets with period P) access the channel in a random fashion, following a CSMA algorithm.



(a) PoS vs. number of EDs **(b)** GIPT vs. number of EDs

Figure 7.4: Comparison of LoRaWAN and IEEE 802.15.4 employed in IIoT applications. The transmission period is fixed to 10 minutes.

The first comparison, shown in Fig. 7.4a reports the PoS for different network sizes and a transmission period fixed to 10 minutes. It can be observed that both LoRaWAN strategies ("constant to SF 7" and "fair modified") outperform IEEE 802.15.4. This is due to the better robustness of LoRa modulation with respect to the IEEE 802.15.4 one; indeed, comparing the derivations reported in the standard (IEEE 802.15.4-2015) with the curves in Fig. 7.2, IEEE 802.15.4 needs almost a 10 dB higher SNR to offer the same BER as the less robust LoRa SF. Moreover, in LoRaWAN the collision probability is reduced with respect to IEEE 802.15.4 because of the orthogonality between different SFs.

A second metric considered in the comparison between LoRaWAN and IEEE 802.15.4 is the GIPT, namely the time elapsed between two consecutive correct packets received at the sink averaged over all the nodes in the network. A performance assessment under the same conditions of Fig. 7.4a (N ranging from 10 to 1000 and $P = 10$ minutes) is provided in Fig. 7.4b. It can be observed that all values are quite close to the transmission period, as expected, but IEEE 802.15.4 provides the worst performance for every network size. This is due to both the lower robustness of IEEE 802.15.4 modulation and the higher randomness of its channel access scheme. Among the two LoRaWAN strategies, the "fair modified" strategy confirms to be the best one, offering a slightly lower GIPT.

Finally, the performance figures of LoRaWAN and IEEE 802.15.4 are compared in terms of the AEC by each node in Fig. 7.5. In this case the number of nodes is fixed to 500 and the transmission period is varied from 1 to 30 minutes. First, it can be observed that the AEC by IEEE 802.15.4 nodes is roughly 100 times higher than that

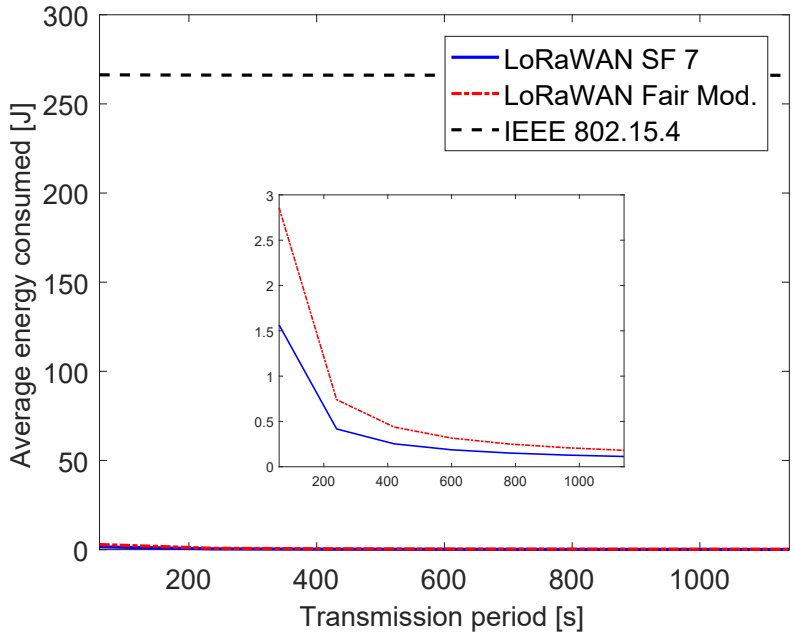


Figure 7.5: AEC in a 2–hours simulation vs. transmission period for LoRaWAN and IEEE 802.15.4 employed in IIoT applications. The number of nodes is fixed to 500. The detailed plot allows to better distinguish the performance of the two SF choice strategies adopted in LoRaWAN.

of LoRaWAN nodes, that stay most of the time in deep sleep state and hence have an extremely low power consumption. To this regard, it is worth observing that the IEEE 802.15.4 protocol has been implemented according to its legacy version, without adopting any specify power saving strategy. Clearly, the introduction of (purposely defined) strategies of this type could definitely lead to better performance in terms of power consumption (El-Hoiydi and Decotignie, 2004).

The detailed plot in Fig. 7.5 allows also to compare the two LoRaWAN strategies for SF choice, "constant to SF 7" and "fair modified", showing that the former scheme allows to consume less energy. This is motivated by the fact that the energy consumed is directly proportional to the amount of time spent in transmitting state, which drains the highest current as reported in Tab. 7.3, and SF 7 offers the lowest possible transmission time. However, it has to be considered that, as shown in Fig. 7.3b, setting the SFs of all nodes to 7 also yields an higher probability of error. Consequently, if retransmissions were activated, the additional attempts necessary to transmit corrupted packets would certainly increase the energy consumption. In any case it is worth observing that, from the simulations carried out, the estimation of the battery lifetime for an ED results about

2 years for a simple 1000 mAh battery.

According to the presented results, it may be concluded that LoRaWAN represented an interesting alternative to IEEE 802.15.4 for non–critical industrial monitoring applications, where sampling rates are not very high and an ultra–low energy consumption is targeted. The latter network still remains a valuable solution, especially if configured in the time–slotted beaconed mode and deployed in the 2.4 GHz band, for more demanding scenarios characterized by higher sampling rates and requiring lower transmission times, which can not be provided by LoRaWAN.

Conclusions

This chapter addressed the adoption of LoRaWAN for indoor industrial monitoring systems which represent an interesting IIoT field of application. After an accurate theoretical analysis, a realistic simulation model of LoRaWAN has been developed that allowed to investigate the behavior of network configurations typically deployed for industrial monitoring. The obtained results showed very good performance in terms of reliability, timeliness and energy consumption. Particularly, a newly introduced technique for the selection of the SF revealed able to outperform other traditionally adopted techniques. A comparison with a configuration of an equivalent WPAN, namely IEEE 802.15.4, has also been carried out and provided encouraging results.

Several future activities can be envisioned as follow–ups of this work. First, the occurrence of downlink transmissions has to be appropriately investigated in order to evaluate the performance of LoRaWAN in presence of retransmissions and adaptive data rate strategies. It would be also important to model Class B LoRa devices, as they support synchronization through beacons, and, hence, may allow to develop a scheduled channel access method (e.g., TDMA) which can improve timeliness. Moreover, in LoRaWAN networks the EDs do not have direct Internet connectivity, so that they can be only remotely accessed through GWs. To this regard, some proposals to integrate IPv6 over LoRaWAN have been already developed (Weber et al., 2016): their feasibility and impact on network performance, however, need to be carefully investigated.

Finally, a further step of analysis is represented by the execution of experimental sessions on real LoRaWAN testbeds that would allow to improve the accuracy of the theoretical and simulation analyses presented in this work.

8

Conclusions

The problem of designing deterministic, timely and reliable wireless networks for industrial applications has been addressed in this thesis. Several solutions have been discussed, acting at different layers of the protocol stack and offering different performance figures.

For what concerns real–time WLANs, outlined in Chap. 4, two main topics have been considered. With reference to the first one, namely the use of IEEE 802.11 n for industrial communications, it has been observed that the best way to configure a MIMO–capable device in industrial applications is to use spatial diversity schemes, such as STBC, in order to increase the reliability of the communication. Combining this feature with other PHY enhancements provided by the standard (e.g., wider 40 MHz channels), data exchange can be faster, more deterministic and more robust with respect to the previous IEEE 802.11g amendment, as confirmed by both theoretical analyses and experimental measurements. The second topic addressed was concerned with industrial RA algorithms and, specifically, a new scheme called RSIN has been proposed. According to this strategy, each node in a WLAN can optimize the number of transmission attempts at MAC layer and the corresponding rates with the goal of maximizing the reliability while ensuring that the data exchange is completed within a given deadline. RSIN leverages on the knowledge of the SNR at the receiver by the transmitter, which can be either explicit (i.e., sent within a reply packet) or based on an estimation procedure. Both approaches have been tested through experimental measurements and numerical simulations, showing

good performance with respect to other state–of–the–art RA schemes.

The possibilities offered by FD–capable wireless devices have been explored in Chap. 5. A new distributed MAC protocol for ad hoc networks, called RCFD, has been proposed, that combines frequency–domain channel access with FD wireless. The protocol is based on three preliminary contention rounds during which nodes advertise their transmission intentions and get access to the channel by transmitting on single OFDM subcarriers while listening to the whole band, thanks to their FD capabilities. RCFD allows to efficiently schedule bidirectional FD communications, as well as to have a fixed and short channel access time, while also solving the hidden terminal issue. The proposed strategy has been compared with other state–of–the–art MAC protocols through both theoretical analysis and numerical simulations, showing higher throughput and fairness and a reduced delay. Future directions in this regard deal on one hand with the actual implementation of this protocol on FD–capable devices, to validate the obtained results, and on the other hand on its extension to industrial wireless communications. The time–bounded and collision–free channel access procedure of RCFD, indeed, can already be interesting for acyclic industrial traffic (e.g., alarms) and must be coupled with a FD–aware TDMA scheduling to handle cyclic traffic.

A new approach towards the realization of industrial wireless networks has been discussed in Chap. 6, where WirelessHP is presented. This vision is based on a complete redesign of the protocol stack, stemming from the observation that currently available wireless standards cannot offer the high performance required by the most critical industrial control applications, especially in terms of latency. As a first step, a low–latency PHY based on OFDM has been proposed and compared with the IEEE 802.11 PHY. Theoretical analyses show that the transmission time of short packets can be reduced significantly with the proposed design, reaching 1 μs in the 2.4/5 GHz band and 200 ns in the mmWave spectrum. The design has been tested in a narrowband demostrator based on SDRs, which validated the transmission time while showing a good reliability. Upgraded versions of this demonstrator are currently being developed, to validate the theoretical analysis. At the same time, a more structured design proposal for the upper layers, starting from the MAC, is on the making, in order to provide the necessary determinism and ensure high communication reliability.

Finally, the use of LoRaWAN to monitor indoor industrial processes in the context of IIoT has been discussed in Chap. 7. This network is typically used in outdoor applications, such as urban monitoring and smart metering, thanks to its long communication range. However, numerical simulations have shown that its long range translates in high reliability when applied to indoor industrial applications and its extremely low power consumption

can also be appealing for these scenarios. The simulations, which were based on an accurate model of a LoRaWAN network deployed in industrial environments, highlighted a prevalence of this network over a IEEE 802.15.4 one, in the considered scenarios. Moreover, an original algorithm for the static selection of transmission rates in a LoRaWAN was proposed and assessed, showing better performance than state–of–the–art ones. In the future, the realization of a LoRaWAN testbed is envisioned, in order to validate these results. Moreover, additional scenarios are going to be analyzed, such as those characterized by the presence of downlink transmissions in additional to uplink ones, as specified by the protocol specifications. Finally, the integration of IPv6 in LoRaWAN end nodes will be further investigated, as it represents a fundamental step towards the actual realization of the IIoT vision.

References

Ai Y., Cheffena M., and Li Q. Radio frequency measurements and capacity analysis for industrial indoor environments. In *2015 9th European Conference on Antennas and Propagation (EuCAP)*, 2015.

Åkerberg J., Gidlund M., and Björkman M. Future research challenges in wireless sensor and actuator networks targeting industrial automation. In *2011 9th IEEE International Conference on Industrial Informatics (INDIN)*, 2011.

Alamouti S. A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, 16(8):1451–1458, 1998.

Aldana C. and others . Joint Proposal Team PHY Simulation Results. IEEE 802.11 TGn document, 06/0067r0, 2006.

Ansari J., Perpinias N., Nähring A., Mähönen P., and Petrova M. Empirical characterization of mm-wave communication links in realistic indoor scenarios. In *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, 2015.

Ashton K. That 'Internet of Things' Thing. URL http://www.rfidjournal.com/articles/view?4986.

Åström K. J. and Wittenmark B. *Computer-controlled systems: theory and design*. Prentice-Hall, 1997.

Au E. Exciting Projects for PHY and MAC Layers of IEEE 802.11 [Standards]. *IEEE Vehicular Technology Magazine*, 11(2):79–81, 2016.

Baldo N., Requena-Esteso M., Núñez Martínez J., Portolès-Comeras M., Nin-Guerrero J., Dini P., and Mangues-Bafalluy J. Validation of the IEEE 802.11 MAC Model in the Ns3 Simulator Using the EXTREME Testbed. In *3rd International Conference on Simulation Tools and Techniques (ICST)*, 2010.

Banelli P., Buzzi S., Colavolpe G., Modenini A., Rusek F., and Ugolini A. Modulation formats and waveforms for 5G networks: Who will be the heir of OFDM?: An overview of alternative modulation schemes for improved spectral efficiency. *IEEE Signal Processing Magazine*, 31(6):80–93, 2014.

Bellalta B. IEEE 802.11 ax: High-efficiency WLANs. *IEEE Wireless Communications*, 23(1):38–46, 2016.

Benvenuto N., Dinis R., Falconer D., and Tomasin S. Single carrier modulation with nonlinear frequency domain equalization: an idea whose time has come–again. *Proceedings of the IEEE*, 98(1):69–96, 2010.

Bharadia D., McMilin E., and Katti S. Full duplex radios. *ACM SIGCOMM Computer Communication Review*, 43(4):375–386, 2013.

Bianchi G. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, 2000.

Bisdikian C. An overview of the Bluetooth wireless technology. *IEEE Communications Magazine*, 39(12):86–94, 2001.

Bluetooth–5.0. Bluetooth core specification version 5.0, December 2016.

Bluetooth–SIG. Bluetooth SIG. URL http://www.bluetooth.com/.

Boccardi F., Heath R. W., Lozano A., Marzetta T. L., and Popovski P. Five disruptive technology directions for 5G. *IEEE Communications Magazine*, 52(2):74–80, 2014.

Brigham E. O. *The Fast Fourier Transform and Its Applications*. Prentice-Hall Englewood Cliffs, NJ, 1974.

Cena G., Seno L., Valenzano A., and Zunino C. On the performance of IEEE 802.11e wireless infrastructures for soft-real-time industrial applications. *IEEE Transactions on Industrial Informatics*, 6(3):425–437, 2010.

Cena G., Valenzano A., and Vitturi S. Hybrid wired/wireless networks for real-time communications. *IEEE Industrial Electronics Magazine*, 2(1):8–20, 2008.

Charfi E., Chaari Fourati L., and Kamoun L. QoS support of voice/video services under IEEE 802.11n WLANs. In *2014 9th International Symposium on Communication Systems, Networks Digital Signal Processing (CSNDSP)*, 2014.

Cheng W., Zhang X., and Zhang H. RTS/FCTS mechanism based full-duplex MAC protocol for wireless networks. In *2013 IEEE Globecom Workshops*, 2013.

Chlebus E. and Divgi G. The Pareto or truncated Pareto distribution? Measurement-based modeling of session traffic for Wi-Fi wireless Internet access. In *2007 IEEE Wireless Communications and Networking Conference (WCNC)*, 2007.

Choi J. I., Hong S., Jain M., Katti S., Levis P., and Mehlman J. Beyond full duplex wireless. In *2012 Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2012.

Choi W., Lim H., and Sabharwal A. Power-controlled medium access control protocol for full-duplex WiFi networks. *IEEE Transactions on Wireless Communications*, 14(7):3601–3613, 2015.

Chow M.-Y. and Tipsuwan Y. Network-based control systems: A tutorial. In *27th Annual Conference of the IEEE Industrial Electronics Society (IECON)*, volume 3, 2001.

Colombo A. W., Karnouskos S., Shi Y., Yin S., and Kaynak O. Industrial cyber–physical systems [scanning the issue]. *Proceedings of the IEEE*, 104(5):899–903, 2016.

Cottet D., van der Merwe W., Agostini F., Riedel G., Oikonomou N., Rueetschi A., Geyer T., Gradinger T., Velthuis R., Wunsch B., and others . Integration technologies for a fully modular and hot-swappable MV multi-level concept converter. In *2015 International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management (PCIM Europe)*, 2015.

Dahlman E., Parkvall S., and Skold J. *4G: LTE/LTE-advanced for mobile broadband*. Academic press, 2013.

Daniels R. C. and Heath Jr R. W. 60 GHz wireless communications: Emerging requirements and design recommendations. *IEEE Vehicular Technology Magazine*, 2 (3):41–50, 2007.

Day B. P., Margetts A. R., Bliss D. W., and Schniter P. Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range. *IEEE Transactions on Signal Processing*, 60(7):3702–3713, 2012.

De Guglielmo D., Brienza S., and Anastasi G. IEEE 802.15.4e: A survey. *Computer Communications*, 88:1–24, 2016.

Decotignie J.-D. Ethernet-based real-time and industrial communications. *Proceedings of the IEEE*, 93(6):1102–1117, 2005a.

Decotignie J.-D. Which Network for Which Application. In **Zurawski R.**, editor, *The Industrial Information Technology Handbook*, chapter 46. CRC Press, 2005b.

Decotignie J.-D. and Pleinevaux P. A survey on industrial communication networks. *Annals of Telecommunications*, 48(9):435–448, 1993.

Dombrowski C. and Gross J. EchoRing: A Low-Latency, Reliable Token-Passing MAC Protocol for Wireless Industrial Networks. In *2015 21st IEEE Conference on European Wireless*, 2015.

Dorf R. C. and Bishop R. H. *Modern control systems*. Pearson, 2011.

Duarte M., Dick C., and Sabharwal A. Experiment-driven characterization of full-duplex wireless systems. *IEEE Transactions on Wireless Communications*, 11(12): 4296–4307, 2012.

Duarte M. and Sabharwal A. Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results. In *2010 Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2010.

Duarte M., Sabharwal A., Aggarwal V., Jana R., Ramakrishnan K., Rice C. W., and Shankaranarayanan N. Design and characterization of a full-duplex multiantenna system for WiFi networks. *IEEE Transactions on Vehicular Technology*, 63(3):1160–1177, 2014.

Dujovne D., Watteyne T., Vilajosana X., and Thubert P. 6TiSCH: deterministic IP-enabled industrial internet (of things). *IEEE Communications Magazine*, 52(12): 36–41, 2014.

Durisi G., Koch T., and Popovski P. Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets. *Proceedings of the IEEE*, 104 (9):1711–1726, 2016.

Dykstra P. Gigabit Ethernet Jumbo Frames, and why you should care. Technical report, ETSI Wireless Factory Starter Group, 1999.

Dzung D., Naedele M., Von Hoff T. P., and Crevatin M. Security for industrial communication systems. *Proceedings of the IEEE*, 93(6):1152–1177, 2005.

El-Hoiydi A. and Decotignie J. D. WiseMAC: an ultra low power MAC protocol for the downlink of infrastructure wireless sensor networks. In *2004 Ninth International Symposium on Computers And Communications (ISCC)*, 2004.

Emami S. *UWB Communication Systems: Conventional and 60 GHz*. Springer, 2013.

Erceg V., Schumacher L., Kyritsi P., Molisch A., Baum D. S., and others . TGn channel models. IEEE 802.11 TGn document, 03/940r4, 2004.

ETSI EN 300 220-1. Technical Report 300 220-1 (V2. 4.1), ETSI, EN, 2012.

Everett E. Full-duplex infrastructure nodes: Achieving long range with half-duplex mobiles. Phd thesis, Rice University, 2012.

Everett E., Duarte M., Dick C., and Sabharwal A. Empowering full-duplex wireless communication by exploiting directional diversity. In *2011 Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2011.

Everett E., Sahai A., and Sabharwal A. Passive self-interference suppression for full-duplex infrastructure nodes. *IEEE Transactions on Wireless Communications*, 13 (2):680–694, 2014.

Farhang-Boroujeny B. and Ding M. Design methods for time-domain equalizers in DMT transceivers. *IEEE Transactions on Communications*, 49(3):554–562, 2001.

Feliciano S., Sarmento H., and de Oliverira J. Z. Field Area Network in a MV/LV substation: A technical and economical analysis. In *2014 IEEE International Conference on Intelligent Energy and Power Systems (IEPS)*, 2014.

Feng X., Zhang J., Zhang Q., and Li B. Use your frequency wisely: Explore frequency domain for channel contention and ACK. In *IEEE International Conference on Computer Communications (INFOCOM)*, 2012.

Ferrer-Coll J., Ängskog P., Elofsson C., Chilo J., and Stenumgaard P. Antenna cross correlation and ricean K-factor measurements in indoor industrial environments at 433 and 868 MHz. *Wireless Personal Communications*, 73(3):587–593, 2013.

Fettweis G. P. The tactile internet: Applications and challenges. *IEEE Vehicular Technology Magazine*, 9(1):64–70, 2014.

Ford R., Zhang M., Mezzavilla M., Dutta S., Rangan S., and Zorzi M. Achieving ultra-low latency in 5G millimeter wave cellular networks. *IEEE Communications Magazine*, 55(3):196–203, 2017.

Friedman J., Hastie T., and Tibshirani R. *The elements of statistical learning.* Springer, 2001.

Gamba G., Tramarin F., and Willig A. Retransmission Strategies for Cyclic Polling Over Wireless Channels in the Presence of Interference. *IEEE Transactions on Industrial Informatics*, 6(3):405–415, August 2010.

Gerlach-Erhardt H. Real time requirements in industrial automation. Technical report, ETSI Wireless Factory Starter Group, 2009.

Gomez C., Oller J., and Paradells J. Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology. *Sensors*, 12(9):11734–11753, 2012.

Goursaud C. and Gorce J.-M. Dedicated networks for IoT : PHY / MAC state of the art and challenges. *EAI endorsed transactions on Internet of Things*, 2015.

Goyal S., Liu P., Gurbuz O., Erkip E., and Panwar S. A distributed MAC protocol for full duplex radio. In *2013 Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2013.

Guibene W., Nowack J., Chalikias N., Fitzgibbon K., Kelly M., and Prendergast D. Evaluation of LPWAN Technologies for Smart Cities: River Monitoring Use-Case. In *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2017.

Gungor V. C., Hancke-sr G. P., and Hancke-jr. G. P. Guest Editorial Special Section on Industrial Wireless Sensor Networks. *IEEE Transactions on Industrial Informatics*, 10(1):762–765, Feb. 2014.

Gungor V. C. and Hancke G. P. Industrial wireless sensor networks: Challenges, design principles, and technical approaches. *IEEE Transactions on Industrial Electronics*, 56(10):4258–4265, 2009.

Haxhibeqiri J., Karaağaç A., Van den Abeele F., Joseph W., Moerman I., and Hoebeke J. LoRa indoor coverage and performance in an industrial environment: case study. In *2017 IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*, 2017.

Heiskala J. and Terry J. *OFDM Wireless LANs: A Theoretical and Practical Guide*. Sams, 2001.

Hernandez D. M., Peralta G., Manero L., Gomez R., Bilbao J., and Zubia C. Energy and coverage study of LPWAN schemes for Industry 4.0. In *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*, 2017.

Hirai J., Kim T.-W., and Kawamura A. Practical study on wireless transmission of power and information for autonomous decentralized manufacturing system. *IEEE Transactions on Industrial Electronics*, 46(2):349–359, 1999.

Holland G., Vaidya N. H., and Bahl P. A Rate-Adaptive MAC Protocol for Multi-Hop Wireless Networks. In *Proceedings of the ACM SIGMOBILE*, 2001.

IEC 61158-2003. IEC 61158: Digital data communications for measurement and control - Fieldbus for use in industrial control systems. Standard, International Electrotechnical Commission, May 2003.

IEC 61784-2 - 2007. IEC 61784: Digital data communications for measurement and control – part 2: Additional profiles for ISO/IEC 8802–3 based communication networks in real–time applications. Standard, International Electrotechnical Commission, November 2007.

IEC 62601-2015. IEC PAS 62601: Industrial networks - Wireless communication network and communication profiles - WIA-PA. Standard, International Electrotechnical Commission, December 2015.

IEC 62948-2017. IEC PAS 62948: Industrial networks - Wireless communication network and communication profiles - WIA-FA. Standard, International Electrotechnical Commission, July 2017.

IEEE 802.11-2016. IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Standard, Institute of Electrical and Electronic Engineers, December 2016.

IEEE 802.15.1-2005. IEEE Standard for Local and Metropolitan Area Networks–Part 15.1a: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for Wireless Personal Area Networks (WPAN). Standard, Institute of Electrical and Electronic Engineers, June 2005.

IEEE 802.15.3-2016. IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPAN). Standard, Institute of Electrical and Electronic Engineers, July 2016.

IEEE 802.15.4-2015. IEEE Standard for Local and Metropolitan Area Networks–Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs). Standard, Institute of Electrical and Electronic Engineers, April 2016.

IEEE 802.3-2015. IEEE Standard for Ethernet. Standard, Institute of Electrical and Electronic Engineers, March 2016.

IETF-RFC 4919. IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals. RFC 4919, August 2007.

Industrie 4.0. Industrie 4.0. URL https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html.

Iqbal Z., Kim K., and Lee H. N. A cooperative wireless sensor network for indoor industrial monitoring. *IEEE Transactions on Industrial Informatics*, 13(2):482–491, 2017.

ISA-100.11a-2009. ISA-100.11a Wireless Systems for Industrial Automation: Process Control and Related Applications, 2009.

Jain M., Choi J. I., Kim T., Bharadia D., Seth S., Srinivasan K., Levis P., Katti S., and Sinha P. Practical, real-time, full duplex wireless. In *17th annual international conference on Mobile computing and networking*, 2011.

Jain R., Chiu D.-M., and Hawe W. R. A quantitative measure of fairness and discrimination for resource allocation in shared computer system. Technical report, Eastern Research Laboratory, Digital Equipment Corporation, 1984.

Jin Y. and Dai F. F. Impact of Transceiver RFIC Impairments on MIMO System Performance. *IEEE Transactions on Industrial Electronics*, 59(1):538–549, Jan 2012.

Johansson N. A., Wang Y. P. E., Eriksson E., and Hessler M. Radio access for ultra-reliable and low-latency 5G communications. In *2015 IEEE International Conference on Communications (ICC)*, 2015.

Jung H., Kwon T. T., Cho K., and Choi Y. REACT: Rate Adaptation using Coherence Time in 802.11 WLANs. *Computer Communications*, 34(11):1316 – 1327, 2011.

Kansal A., Hsu J., Zahedi S., and Srivastava M. B. Power management in energy harvesting sensor networks. *ACM Transactions on Embedded Computing Systems (TECS)*, 6(4):32, 2007.

Karl H. and Willig A. *Protocols and architectures for wireless sensor networks*. John Wiley & Sons, 2007.

Kim D., Lee H., and Hong D. A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers. *IEEE Communications Surveys & Tutorials*, 17(4):2017–2046, 2015.

Kim J. Y., Mashayekhi O., Qu H., Kazandjieva M., and Levis P. Janus: A novel MAC protocol for full duplex radio. Computer Science Technical Reports (CSTR), Stanford University, 2013.

Kulkarni P. and Quadri S. Simple and Practical Rate Adaptation Algorithms for Wireless Networks. In *IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks Workshops (WoWMoM)*, June 2009.

Lee J.-S., Su Y.-W., and Shen C.-C. A comparative study of wireless protocols: Bluetooth, UWB, ZigBee, and Wi-Fi. In *33rd Annual Conference of the IEEE Industrial Electronics Society (IECON)*, 2007.

Lee J. and Kwak Y. 5G Standard Development: Technology and Roadmap. In **Luo F.-L. and Zhang C.**, editors, *Signal Processing for 5G*, chapter 23, pages 561–576. John Wiley & Sons, 2016.

Lee S., Lee K. C., Lee M. H., and Harashima F. Integration of mobile vehicles for automated material handling using Profibus and IEEE 802.11 networks. *IEEE Transactions on Industrial Electronics*, 49(3):693–701, 2002.

Lennvall T., Åkerberg J., Hansen E., and Yu K. A New Wireless Sensor Network TDMA Timing Synchronization Protocol. In *2016 14th IEEE International Conference on Industrial Informatics (INDIN)*, 2016.

Lo Bello L., Akerberg J., Gidlund M., and Uhlemann E. Guest Editorial Special Section on New Perspectives on Wireless Communications in Automation: from Industrial Monitoring and Control to Cyber-Physical Systems. *IEEE Transactions on Industrial Informatics*, 13(3):1393–1397, 2017.

LoRaWAN v1.0-2015. LoRa Alliance LoRaWAN specification. Standard, LoRa Alliance, 2015.

Luvisotto M., Pang Z., and Dzung D. Ultra High Performance Wireless Control for Critical Applications: Challenges and Directions. *IEEE Transactions on Industrial Informatics*, 13(3):1448–1459, 2017a.

Luvisotto M., Pang Z., Dzung D., Zhan M., and Jiang X. Physical Layer Design of High Performance Wireless Transmission for Critical Control Applications. *IEEE Transactions on Industrial Informatics*, to appear 2017b.

Luvisotto M., Sadeghi A., Lahouti F., Vitturi S., and Zorzi M. RCFD: A frequency-based channel access scheme for full-duplex wireless networks. In *2016 IEEE International Conference on Communications (ICC)*, 2016a.

Luvisotto M., Sadeghi A., Lahouti F., Vitturi S., and Zorzi M. RCFD: A Novel Channel Access Scheme for Full-Duplex Wireless Networks Based on Contention in Time and Frequency Domains. *arXiv preprint arXiv:1606.01038*, 2016b. *[submitted to IEEE Transactions on Mobile Computing]*.

Luvisotto M., Tagliapietra A., Romagnolo S., Tramarin F., and Vitturi S. Real–Time Wireless Extensions of Industrial Ethernet Networks. In *2017 15th IEEE International Conference on Industrial Informatics (INDIN)*, 2017c.

Luvisotto M., Tramarin F., and Vitturi S. A learning algorithm for rate selection in real-time wireless LANs. *Computer Networks*, 126:114–124, 2017d.

Magrin D., Centenaro M., and Vangelista L. Performance evaluation of LoRa networks in a smart city scenario. In *2017 IEEE International Conference on Communications (ICC)*, 2017.

Maqhat B., Baba M., and Rahman R. A-MSDU real time traffic scheduler for IEEE802.11n WLANs. In *2012 IEEE Symposium on Wireless Technology and Applications (ISWTA)*, 2012.

Margelis G., Piechocki R., Kaleshi D., and Thomas P. Low Throughput Networks for the IoT: Lessons learned from industrial implementations. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, 2015.

Marti P., Villà R., Fuertes J. M., and Fohler G. Networked Control Systems Overview. In **Zurawski R.**, editor, *The Industrial Information Technology Handbook*, chapter 47. CRC press, 2005.

Miao G., Zander J., Sung K. W., and Slimane S. B. *Fundamentals of Mobile Data Networks*. Cambridge University Press, 2016.

Minstrel. The minstrel rate control algorithm for mac80211. URL https://wireless.wiki.kernel.org/en/developers/documentation/mac80211/ratecontrol/minstrel.

Mishra A. K., Nigam Y. K., and Singh D. R. Controlled blasting in a limestone mine using electronic detonators: A case study. *Journal of the Geological Society of India*, 89(1):87–90, 2017.

Miura K. and Bandai M. Node architecture and MAC protocol for full duplex wireless and directional antennas. In *2012 23rd IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, 2012.

Moraes R., Vasques F., Portugal P., and Fonseca J. A. VTP-CSMA: A virtual token passing approach for real-time communication in IEEE 802.11 wireless networks. *IEEE Transactions on Industrial Informatics*, 3(3):215–224, 2007.

Mumtaz S., Alsohaily A., Pang Z., Rayes A., Tsang K. F., and Rodriguez J. Massive Internet of Things for Industrial Applications: Addressing Wireless IIoT Connectivity Challenges and Ecosystem Fragmentation. *IEEE Industrial Electronics Magazine*, 11(1):28–33, 2017.

Neumann P., Montavont J., and Noël T. Indoor deployment of low-power wide area networks (LPWAN): A LoRaWAN case study. In *2016 12th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2016.

Ngo H.-D. and Yang H.-S. Latency and Traffic Reduction for Process-Level Network in Smart Substation Based on High-Availability Seamless Redundancy. *IEEE Transactions on Industrial Electronics*, 63(4):2181–2189, 2016.

Niu Y., Li Y., Jin D., Su L., and Vasilakos A. V. A Survey of Millimeter Wave Communications (mmWave) for 5G: Opportunities and Challenges. *ACM Journal of Wireless Networks*, 21(8):2657–2676, 2015.

ns3. The Network Simulator 3. URL http://www.nsnam.org/.

Ogata K. *Discrete-time control systems*. Prentice-Hall, 1995.

O'Halloran D. and Kvochko E. Industrial Internet of Things: Unleashing the Potential of Connected Products and Services. Technical report, World Economic Forum, 2015. URL http://reports.weforum.org/industrial-internet-of-things/.

OMRON WD30-2002. *WD30-ME/-SE/-ME01/-SE01 DeviceNet Wireless Units*. OMRON, 2002.

Ong S. K. and Nee A. Y. C. *Virtual and augmented reality applications in manufacturing*. Springer Science & Business Media, 2013.

Orfanus D., Indergaard R., Prytz G., and Wien T. EtherCAT-based platform for distributed control in high-performance industrial applications. In *2013 18th IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*, 2013.

Pang Z., Luvisotto M., and Dzung D. Wireless High-Performance Communications: The Challenges and Opportunities of a New Target. *IEEE Industrial Electronics Magazine*, 11(3):20–25, 2017.

Parikh P. P., Sidhu T. S., and Shami A. A Comprehensive Investigation of Wireless LAN for IEC 61850-Based Smart Distribution Substation Applications. *IEEE Transactions on Industrial Informatics*, 9(3):1466–1476, 2013.

Park C. and Rappaport T. S. Short-range wireless communications for next-generation networks: UWB, 60 GHz millimeter-wave WPAN, and ZigBee. *IEEE Wireless Communications*, 14(4):70–78, 2007.

Pei G. and Henderson T. R. Validation of OFDM error rate model in ns-3. *Boeing Research Technology*, 2010.

Perahia E. and Stacey R. *Next generation wireless LANs: 802.11 n and 802.11 ac*. Cambridge university press, 2013.

Perera C., Liu C. H., and Jayawardena S. The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey. *IEEE Transactions on Emerging Topics in Computing*, 3(4):585–598, 2015.

Petäjäjärvi J., Mikhaylov K., Yasmin R., Hämäläinen M., and Iinatti J. Evaluation of LoRa LPWAN Technology for Indoor Remote Health and Wellbeing Monitoring. *International Journal of Wireless Information Networks*, 24(2):153–165, 2017.

Petersen S. and Carlsen S. WirelessHART versus ISA100. 11a: The format war hits the factory floor. *IEEE Industrial Electronics Magazine*, 5(4):23–34, 2011.

Rappaport T. S. *Wireless communications: principles and practice*. Prentice-Hall, 1996.

Rappaport T. S., Murdock J. N., and Gutierrez F. State of the art in 60-GHz integrated circuits and systems for wireless communications. *Proceedings of the IEEE*, 99(8):1390–1436, 2011.

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

Ray A. Introduction to networking for integrated control systems. *IEEE Control Systems Magazine*, 9(1):76–79, 1989.

Ray P. P. and Agarwal S. Bluetooth 5 and Internet of Things: Potential and architecture. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1461–1465, 2016.

Raza U., Kulkarni P., and Sooriyabandara M. Low Power Wide Area Networks: An Overview. *IEEE Communications Surveys & Tutorials*, 19(2):855–873, 2017.

Rege V. and Pecorella T. A Realistic MAC and Energy Model for 802.15.4. In *ACM Workshop on Ns-3*, 2016.

Rentschler M. and Laukemann P. Performance analysis of parallel redundant WLAN. In *2012 17th IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*, 2012.

Reynders B., Meert W., and Pollin S. Power and spreading factor control in low power wide area networks. In *2017 IEEE International Conference on Communications (ICC)*, 2017.

Sabharwal A., Schniter P., Guo D., Bliss D. W., Rangarajan S., and Wichman R. In-band full-duplex wireless: Challenges and opportunities. *IEEE Journal on Selected Areas in Communications*, 32(9):1637–1652, 2014.

Sahai A., Patel G., and Sabharwal A. Pushing the limits of full-duplex: Design and real-time implementation. *arXiv preprint arXiv:1107.0607*, 2011.

Saif A., Othman M., Subramaniam S., and AbdulHamid N. Impact of aggregation headers on aggregating small MSDUs in 802.11n WLANs. In *2010 International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, Dec 2010.

Salman N., Rasool I., and Kemp A. H. Overview of the IEEE 802.15.4 standards family for low rate wireless personal area networks. In *2010 7th IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2010.

Sanchez-Iborra R. and Cano M.-D. State of the Art in LP-WAN Solutions for Industrial IoT Services. *Sensors*, 16(5), 2016.

Santandrea G. A Profinet IO application implemented on Wireless LAN. In *2006 IEEE International Workshop on Factory Communication Systems (WFCS)*, 2006.

Santonja-Climent S., Todoli-Ferrandis D., Albero-Albero T., Sempere-Paya V., Silvestre-Blanes J., and Alcober J. Analysis of control and multimedia real-time traffic over SIP and RTP on 802.11n wireless links for utilities networks. In *2010 15th IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*, Sept 2010.

Sauter T. The continuing evolution of integration in manufacturing automation. *IEEE Industrial Electronics Magazine*, 1(1):10–19, 2007.

Sauter T. The three generations of field-level networks—evolution and compatibility issues. *IEEE Transactions on Industrial Electronics*, 57(11):3585–3595, 2010.

Schaich F., Wild T., and Chen Y. Waveform Contenders for 5G - Suitability for Short Packet and Low Latency Transmissions. In *2014 79th IEEE Vehicular Technology Conference (VTC Spring)*, 2014.

Scheible G., Dzung D., Endresen J., and Frey J. E. Unplugged but connected – Design and implementation of a truly wireless real-time sensor/actuator interface. *IEEE Industrial Electronics Magazine*, 1(2):25–34, 2007.

Schmidl T. M. and Cox D. C. Low-overhead, low-complexity [burst] synchronization for OFDM. In *1996 IEEE International Conference on Communications (ICC)*, 1996.

Schulz P., Matthe M., Klessig H., Simsek M., Fettweis G., Ansari J., Ashraf S. A., Almeroth B., Voigt J., Riedel I., and others . Latency critical IoT applications in 5g: Perspective on the design of radio interface and network architecture. *IEEE Communications Magazine*, 55(2):70–78, 2017.

SEMTECH SX1272. *SX1272/73: 860 MHz to 1020 MHz Low Power Long Range Transceiver*. Semtech, March 2017. URL http://www.semtech.com/images/datasheet/sx1272.pdf.

Sen S., Choudhury R. R., and Nelakuditi S. Listen (on the frequency domain) before you talk. In *9th ACM SIGCOMM Workshop on Hot Topics in Networks*, 2010.

Sen S., Roy Choudhury R., and Nelakuditi S. No time to countdown: Migrating backoff to the frequency domain. In *17th annual international conference on Mobile computing and networking*, 2011.

Senaratne D. and Tellambura C. Beamforming for space division duplexing. In *2011 IEEE International Conference on Communications (ICC)*, 2011.

Silva B., Pang Z., Åkerberg J., Neander J., and Hancke G. Experimental study of UWB-based high precision localization for industrial applications. In *2014 IEEE International Conference on Ultra-WideBand (ICUWB)*, 2014.

Silvestre-Blanes J., Berenguer-Sebastiá J., Sempere-Paya V., and Ferrandis D. T. 802.11n Performance analysis for a real multimedia industrial application. *Computers in Industry*, 66(0):31 – 40, 2015.

Singh N., Gunawardena D., Proutiere A., Radunovi B., Balan H. V., and Key P. Efficient and fair MAC for wireless networks with self-interference cancellation. In *2011 International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2011.

Skeie T., Johannessen S., and Holmeide O. Timeliness of real-time IP communication in switched industrial ethernet networks. *IEEE Transactions on Industrial Informatics*, 2(1):25–39, 2006.

Swamy V. N., Rigge P., Ranade G., Sahai A., and Nikolic B. Network coding for high-reliability low-latency wireless control. In *2016 IEEE Wireless Communications and Networking Conference (WCNC)*, 2016.

Swamy V. N., Suri S., Rigge P., Weiner M., Ranade G., Sahai A., and Nikolic B. Cooperative communication for high-reliability low-latency wireless control. In *2015 IEEE International Conference on Communications (ICC)*, 2015.

Tamaki K., Raptino H. A., Sugiyama Y., Bandai M., Saruwatari S., and Watanabe T. Full duplex media access control for wireless multi-hop networks. In *2013 77th IEEE Vehicular Technology Conference (VTC Spring)*, 2013.

Tanghe E., Joseph W., Verloock L., Martens L., Capoen H., Herwegen K. V., and Vantomme W. The industrial indoor channel: large-scale and temporal fading at 900, 2400, and 5200 MHz. *IEEE Transactions on Wireless Communications*, 7(7): 2740–2751, 2008.

Tayeb S., Latifi S., and Kim Y. A survey on IoT communication and computation frameworks: An industrial perspective. In *2017 7th IEEE Annual Computing and Communication Workshop and Conference (CCWC)*, 2017.

Thomesse J.-P. Fieldbus technology in industrial automation. *Proceedings of the IEEE*, 93(6):1073–1101, 2005.

Thread. Thread Group. URL http://threadgroup.org/.

Tobagi F. and Kleinrock L. Packet switching in radio channels: part ii–the hidden terminal problem in carrier sense multiple-access and the busy-tone solution. *IEEE Transactions on Communications*, 23(12):1417–1433, 1975.

Toh C. and Norum L. A high speed control network synchronization jitter evaluation for embedded monitoring and control in modular multilevel converter. In *2013 IEEE PowerTech (POWERTECH)*, 2013.

Tramarin F. Industrial Wireless Sensor Networks – Simulation and measurement in an interfering environment. Phd thesis, University of Padova, 2012.

Tramarin F. and Vitturi S. Strategies and services for energy efficiency in real-time Ethernet networks. *IEEE Transactions on Industrial Informatics*, 11(3):841–852, 2015.

Tramarin F., Vitturi S., and Luvisotto M. Improved rate adaptation strategies for real-time industrial IEEE 802.11n WLANs. In *2015 20th IEEE Conference on Emerging Technologies and Factory Automation (ETFA)*, 2015.

Tramarin F., Vitturi S., and Luvisotto M. Performance analysis of IEEE 802.11 Rate Selection for Industrial Networks. In *42nd Annual Conference of the IEEE Industrial Electronics Society (IECON)*, 2016a.

Tramarin F., Vitturi S., and Luvisotto M. A Dynamic Rate Selection Algorithm for IEEE 802.11 Industrial Wireless LAN. *IEEE Transactions on Industrial Informatics*, 13(2):846–855, 2017.

Tramarin F., Vitturi S., Luvisotto M., and Zanella A. On the Use of IEEE 802.11 n for Industrial Communications. *IEEE Transactions on Industrial Informatics*, 12(5): 1877–1886, 2016b.

Tse D. and Viswanath P. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

Varsier N. and Schwoerer J. Capacity limits of LoRaWAN technology for smart metering applications. In *2017 IEEE International Conference on Communications (ICC)*, 2017.

Vitturi S., Peretti L., Seno L., Zigliotto M., and Zunino C. Real-time Ethernet networks for motion control. *Computer standards & interfaces*, 33(5):465–476, 2011.

Vitturi S., Tramarin F., and Seno L. Industrial wireless networks: The significance of timeliness in communication systems. *IEEE Industrial Electronics Magazine*, 7(2): 40–51, 2013.

Wan J., Tang S., Shu Z., Li D., Wang S., Imran M., and Vasilakos A. V. Software-Defined Industrial Internet of Things in the Context of Industry 4.0. *IEEE Sensors Journal*, 16(20):7373–7380, 2016.

Wang L., Wu K., and Hamdi M. Combating hidden and exposed terminal problems in wireless networks. *IEEE Transactions on Wireless Communications*, 11(11):4204–4213, 2012.

Weber P., Jäckle D., Rahusen D., and Sikora A. IPv6 over LoRaWAN. In *2016 3rd International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, 2016.

Wei Y.-H., Leng Q., Han S., Mok A. K., Zhang W., and Tomizuka M. RT-WiFi: Real-time high-speed communication protocol for wireless cyber-physical control applications. In *2013 34th IEEE Real-Time Systems Symposium (RTSS)*, 2013.

Willig A., Kubisch M., Hoene C., and Wolisz A. Measurements of a Wireless Link in an Industrial Environment using an IEEE 802.11-Compliant Physical Layer. *IEEE Transactions on Industrial Electronics*, 49(6):1265–1282, 2002.

Willig A. Recent and emerging topics in wireless industrial communications: A selection. *IEEE Transactions on Industrial Informatics*, 4(2):102–124, 2008.

Willig A., Matheus K., and Wolisz A. Wireless technology in industrial networks. *Proceedings of the IEEE*, 93(6):1130–1151, 2005.

WirelessHART. HART Field Communication Protocol Specification, rev. 7.5, 2007. URL https://fieldcommgroup.org.

Wittenmark B., Nilsson J., and Torngren M. Timing problems in real-time control systems. In *1995 American Control Conference (ACC)*, 1995.

Wolf W. Cyber-physical systems. *Computer*, 42(3):88–89, 2009.

Wollschlaeger M., Sauter T., and Jasperneite J. The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0. *IEEE Industrial Electronics Magazine*, 11(1):17–27, 2017.

Xia D., Hart J., and Fu Q. Evaluation of the Minstrel rate adaptation algorithm in IEEE 802.11g WLANs. In *2013 IEEE International Conference on Communications (ICC)*, 2013.

Xiong J. and Lam J. Stabilization of linear systems over networks with bounded packet loss. *Automatica*, 43(1):80–87, 2007.

Xu L. D., He W., and Li S. Internet of Things in Industries: A Survey. *IEEE Transactions on Industrial Informatics*, 10(4):2233–2243, 2014.

Xu X. From cloud computing to cloud manufacturing. *Robotics and Computer-Integrated Manufacturing*, 28(1):75 – 86, 2012.

Yamamoto K., Ichihara F., Hasegawa K., Tukuda M., and Omura I. 60 GHz wireless signal transmitting gate driver for IGBT. In *2015 27th IEEE International Symposium on Power Semiconductor Devices IC's (ISPSD)*, 2015.

Yang S.-H., Yang H.-S., Ahn Y.-H., and Kim Y.-H. Performance analysis of IEC 61850 based substation. In *2012 14th International Conference on Advanced Communication Technology (ICACT)*, 2012.

Yih C. H. Analysis and Compensation of DC Offset in OFDM Systems Over Frequency-Selective Rayleigh Fading Channels. *IEEE Transactions on Vehicular Technology*, 58 (7):3436–3446, 2009.

Yilmaz O. N., Wang Y.-P. E., Johansson N. A., Brahmi N., Ashraf S. A., and Sachs J. Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, 2015.

Yoo D. S., Stark W. E., Yar K. P., and Oh S. J. Coding and Modulation for Short Packet Transmission. *IEEE Transactions on Vehicular Technology*, 59(4):2104–2109, 2010.

Yu K., Pang Z., Gidlund M., Åkerberg J., and Björkman M. Realflow: Reliable real-time flooding-based routing protocol for industrial wireless sensor networks. *International Journal of Distributed Sensor Networks*, 10(7), 2014.

Zheng W., Ma H., and He X. Modeling, analysis, and implementation of real time network controlled parallel multi-inverter systems. In *2012 7th International Power Electronics and Motion Control Conference (IPEMC)*, 2012.

Zhong Q.-C. *Robust control of time-delay systems*. Springer Science & Business Media, 2006.

Zhou W. and Srinivasan K. SIM+: A simulator for full duplex communications. In *IEEE International Conference on Signal Processing and Communications (SPCOM)*, 2014.

Zhou W., Srinivasan K., and Sinha P. RCTC: Rapid concurrent transmission coordination in Full Duplex Wireless networks. In *2013 21st IEEE International Conference on Network Protocols (ICNP)*, 2013.

Zhu H., Pang Z., Xie B., and Bag G. IETF IoT based wireless communication for latency-sensitive use cases in building automation. In *2016 25th IEEE International Symposium on Industrial Electronics (ISIE)*, 2016.

ZigBee. The ZigBee Alliance. URL http://www.zigbee.org/.