



Standards
Certification
Education & Training
Publishing
Conferences & Exhibits

Setting the Standard for Automation™

TECHNICAL REPORT

ISA-TR100.00.01-2006

The Automation Engineer's Guide to Wireless Technology Part 1 – The Physics of Radio, a Tutorial

Approved 29 December 2006

NOTICE OF COPYRIGHT

This is a copyright document and may not be copied or distributed in any form or manner without the permission of ISA. This copy of the document was made for the sole use of the person to whom ISA provided it and is subject to the restrictions stated in ISA's license to that person.

It may not be provided to any other person in print, electronic, or any other form. Violations of ISA's copyright will be prosecuted to the fullest extent of the law and may result in substantial civil and criminal penalties.

ISA-TR100.00.01-2006

The Automation Engineer's Guide to Wireless Technology Part 1 : The Physics of Radio, a Tutorial

ISBN: 978-1-934394-09-0

Copyright © 2006 by ISA. All rights reserved. Not for resale. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of the Publisher.

ISA
67 Alexander Drive
P. O. Box 12277
Research Triangle Park, North Carolina 27709
USA

Preface

This preface, as well as all footnotes and annexes, is included for information purposes and is not part of ISA-TR100.00.01-2006.

This document has been prepared as part of the service of ISA towards a goal of uniformity in the field of instrumentation. To be of real value, this document should not be static but should be subject to periodic review. Toward this end, the Society welcomes all comments and criticisms and asks that they be addressed to the Secretary, Standards and Practices Board; ISA; 67 Alexander Drive; P. O. Box 12277; Research Triangle Park, NC 27709; Telephone (919) 549-8411; Fax (919) 549-8288; E-mail: standards@isa.org.

The ISA Standards and Practices Department is aware of the growing need for attention to the metric system of units in general, and the International System of Units (SI) in particular, in the preparation of instrumentation standards. The Department is further aware of the benefits to USA users of ISA standards of incorporating suitable references to the SI (and the metric system) in their business and professional dealings with other countries. Toward this end, this Department will endeavor to introduce SI-acceptable metric units in all new and revised standards, recommended practices, and technical reports to the greatest extent possible. *Standard for Use of the International System of Units (SI): The Modern Metric System*, published by the American Society for Testing & Materials as IEEE/ASTM SI 10-97, and future revisions, will be the reference guide for definitions, symbols, abbreviations, and conversion factors.

It is the policy of ISA to encourage and welcome the participation of all concerned individuals and interests in the development of ISA standards, recommended practices, and technical reports. Participation in the ISA standards-making process by an individual in no way constitutes endorsement by the employer of that individual, of ISA, or of any of the standards, recommended practices, and technical reports that ISA develops.

CAUTION — ISA ADHERES TO THE POLICY OF THE AMERICAN NATIONAL STANDARDS INSTITUTE WITH REGARD TO PATENTS. IF ISA IS INFORMED OF AN EXISTING PATENT THAT IS REQUIRED FOR USE OF THE DOCUMENT, IT WILL REQUIRE THE OWNER OF THE PATENT TO EITHER GRANT A ROYALTY-FREE LICENSE FOR USE OF THE PATENT BY USERS COMPLYING WITH THE DOCUMENT OR A LICENSE ON REASONABLE TERMS AND CONDITIONS THAT ARE FREE FROM UNFAIR DISCRIMINATION.

EVEN IF ISA IS UNAWARE OF ANY PATENT COVERING THIS DOCUMENT, THE USER IS CAUTIONED THAT IMPLEMENTATION OF THE DOCUMENT MAY REQUIRE USE OF TECHNIQUES, PROCESSES, OR MATERIALS COVERED BY PATENT RIGHTS. ISA TAKES NO POSITION ON THE EXISTENCE OR VALIDITY OF ANY PATENT RIGHTS THAT MAY BE INVOLVED IN IMPLEMENTING THE DOCUMENT. ISA IS NOT RESPONSIBLE FOR IDENTIFYING ALL PATENTS THAT MAY REQUIRE A LICENSE BEFORE IMPLEMENTATION OF THE DOCUMENT OR FOR INVESTIGATING THE VALIDITY OR SCOPE OF ANY PATENTS BROUGHT TO ITS ATTENTION. THE USER SHOULD CAREFULLY INVESTIGATE RELEVANT PATENTS BEFORE USING THE DOCUMENT FOR THE USER'S INTENDED APPLICATION.

HOWEVER, ISA ASKS THAT ANYONE REVIEWING THIS DOCUMENT WHO IS AWARE OF ANY PATENTS THAT MAY IMPACT IMPLEMENTATION OF THE DOCUMENT NOTIFY THE ISA STANDARDS AND PRACTICES DEPARTMENT OF THE PATENT AND ITS OWNER. ADDITIONALLY, THE USE OF THIS DOCUMENT MAY INVOLVE HAZARDOUS MATERIALS, OPERATIONS OR EQUIPMENT. THE DOCUMENT CANNOT ANTICIPATE ALL POSSIBLE APPLICATIONS OR ADDRESS ALL POSSIBLE SAFETY ISSUES ASSOCIATED WITH USE IN HAZARDOUS CONDITIONS. THE USER OF THIS DOCUMENT MUST EXERCISE SOUND PROFESSIONAL JUDGMENT CONCERNING ITS USE AND APPLICABILITY UNDER THE USER'S

PARTICULAR CIRCUMSTANCES. THE USER MUST ALSO CONSIDER THE APPLICABILITY OF ANY GOVERNMENTAL REGULATORY LIMITATIONS AND ESTABLISHED SAFETY AND HEALTH PRACTICES BEFORE IMPLEMENTING THIS DOCUMENT.

THE USER OF THIS DOCUMENT SHOULD BE AWARE THAT THIS DOCUMENT MAY BE IMPACTED BY ELECTRONIC SECURITY ISSUES. THE COMMITTEE HAS NOT YET ADDRESSED THE POTENTIAL ISSUES IN THIS VERSION.

The following served as voting members on the SP100 committee:

NAME	COMPANY
W. Manges, Co-Chairman	Oak Ridge National Lab
R. Sanders, Co-Chairman	Exxonmobil Research & Engineering
A. Severinson, Working Group Chairman	Omnex Control Systems Inc.
M. Zielinski, Managing Director*	Emerson Process Management
B. Armstrong	Eaton Corp.
A. Bastiaans	TNO
R. Bhojani	Bayer Technology Services
M. Buhr*	Bayer Material Science
R. Caro	CMC Associates
K. Demachi	Yokogawa Electric Corporation
B. Drake	Frontline Test Equipment Inc.
M. Dultra	SMAR International Corp.
S. Dwars	Shell Global Solutions International BV
A. Espejo	Syncrude Canada Ltd.
P. Flanders	Saudi Aramco
P. Fuhr	Aprion Inc.
E. Hand	Sara Lee Food & Beverage
T. Hasegawa*	Yokogawa Electric Corporation
J. Hult	Oceana Sensor
J. Jamison	OPTI Canada Inc.
H. Kagan	Invensys I/A
D. Kaufman*	Honeywell
P. Kinney	Kinney Consulting LLC
D. Lafferty	BP America Inc.
G. LaFramboise	Chevron Energy Technology Co.
R. Lally*	Oceana Sensor Technologies
E. Lambert	Wunderlich-Malec Engineering
C. McIntyre	Endress + Hauser
N. McLeod	ARKEMA
M. McNerney	Boeing
M. Mets	Savannah River National Labs.
G. Murphy	University of Alabama
K. Orth	Emerson Process Management
T. Phinney	Honeywell ACS Adv. Tech. Lab
M. Regan*	Adaptive Instruments
N. Sands	El du Pont
T. Schnaare*	Emerson Process Management
D. Sexton	GE Global Research
H. Storey	Shell Global Solutions
R. Velarde	Advanced Industrial Networks
I. Verhappen*	Syncrude Canada Ltd.
J. Woodward*	Exxonmobil Research & Engineering
G. Yon	Adaptive Instruments Corp.

The following served as members of the Standards and Practices Board and approved the document on 29 December 2006.

NAME	AFFILIATION
I. Verhappen, Vice President	Syncrude Canada, Ltd.
F. Amir	E I Du Pont Co.
D. Bishop	Consultant
M. Coppler	Ametek Inc.
B. Dumortier	Schneider Electric
W. Holland	Consultant
E. Icaan	ACES Inc.
A. Iverson	Ivy Optiks
R. Jones	Consultant
K. P. Lindner	Endress + Hauser Process Solutions
V. Maggioli	Feltronics Corp.
T. McAviney	Jacobs Engineering Group
A. McCauley	Chagrin Valley Controls Inc.
G. McFarland	Emerson Process Management
R. Reimer	Rockwell Automation
J. Rennie	Consultant
N. Sands	E I Du Pont Co.
H. Sasajima	Yamatake Corp.
T. Schnaare	Rosemount Inc.
J. Tatera	Tatera & Associates
R. Webb	Consultant
W. Weidman	Parsons Energy and Chemicals
J. Weiss	KEMA Inc.
M. Widmeyer	Stanford Linear Accelerator Center
C. Williams	Eastman Kodak Co.
M. Zielinski	Emerson Process Management

This page intentionally left blank.

Contents

1	Scope and Purpose	9
2	Free space communications basics	10
2.1	Elements of a radio link	10
2.2	Noise and interference in radio systems	19
2.3	The effect of noise on radio link performance	20
2.4	System gain and link budget	22
2.5	The effect of bit rate on radio link performance	22
2.6	Spread spectrum technologies	23
2.7	Orthogonal frequency division multiplexing	27
2.8	Signal coding to mitigate bit errors	28
2.9	Diversity, equalisation and other advanced techniques	30
3	Radio propagation	30
3.1	RF power	31
3.2	Receiver sensitivity	31
3.3	Frequency	32
3.4	The radio path	32
3.5	Antennas	37
3.6	Coaxial cables and connectors	39
3.7	The effect of gain and loss	40
4	Spectrum sharing in the ISM bands	41
4.1	Introduction	41
4.2	Coexistence, interoperability and interworking	41
4.3	Spectrum coordination – “legislated coexistence”	42
4.4	Services in the regulated spectrum bands	43
4.5	Assessing the threat of interference	45
4.6	The near-far problem in the ISM bands	49
4.7	Interference in direct sequence spread spectrum radios	52
4.8	Interference in frequency-hopping systems	59
4.9	Summary and conclusions	62
	Annex A - Glossary	63
	Annex B - Bibliography and References	65

This page intentionally left blank.

The Automation Engineer's Guide to Wireless Technology Part 1 - The Physics of Radio, a Tutorial

1 Scope and Purpose

In this technical report, “radio” means “the wireless transmission of information between two or more locations using electromagnetic waves”. The name “radio” (derived from “radiation”) was officially adopted at the 1906 Berlin Radiotelegraph Convention, although the alternative “wireless” remained in parallel use and is now gaining precedence over its competitor. This change in nomenclature is particularly relevant where radio technology is replacing wires. For example the telephone companies’ copper pair subscriber loops are rapidly losing ground to cellular phones, Category 5 (CAT5) cable is giving way to 802.11 wireless local area networks (LANs), and even fiber optic cables can be replaced over short distances by line-of-sight free-space laser links.

The industrial instrumentation and automation field has traditionally relied on wired connections, but the need for greater flexibility and lower costs favors growing applications of wireless technology. In response to this need, the ISA-SP100 Committee was formed to address wireless manufacturing and control systems applications such as the following:

- field sensors used for monitoring, control, alarm, and shutdown;
- wireless real time field-to-business systems (e.g., wireless equipment interfacing work order systems, control LAN, business LAN, voice).

This covers all industries including fluid processing, material processing, and discrete parts manufacturing environments.

Equipment cost will be an important factor. It is easy to imagine new and useful monitor and control applications which, although not feasible today because of the cost of the wired connections, will become practical when the cost of wireless networks has been reduced sufficiently.

The broad scope of these applications suggests that there is unlikely to be an optimum “one size fits all” wireless solution. The bandwidth, throughput, signal latency (transmission delay), bit error ratio, availability and reliability requirements of a wireless local area network carrying large volumes of file transfer or voice traffic or a link used for real-time control of critical temperature or pressure parameters in a process tank are very different from those of a link used to monitor the water level in a remote reservoir, where only a few bytes may need to be transmitted three or four times a day. This technical report focuses on radios using the unlicensed industrial, scientific and medical (ISM) and Unlicensed National Information Infrastructure (UNII) bands. In this technical report where there are references to the ISM bands the UNII bands should also be considered to be included.

These bands have certain aspects in common:

- They are required to comply with regulations set by government agencies, such as the Federal Communications Commission (FCC) in the United States. For example, both transmitter output power and radiated power density are capped, to limit the potential for interference. In general, the regulations are very flexible and provide ample scope for developing innovative solutions to wireless communications problems.
- They are exposed to the constant threat of radio frequency interference from other equipment operating in the same bands.

This technical report presents a general tutorial on the basic principles of radio communications, and then discusses the performance of spread spectrum radio systems in the presence of interference from other ISM band users and non-ideal propagation conditions, such as may be expected in the industrial plant environment. Its objective is to give readers a realistic understanding of how radio links can complement and/or replace wired connections, the factors influencing link range, and the pitfalls for the unwary. Industrial applications for wireless communications range from the relatively non-critical, such as asset monitoring to support preventive maintenance programs, all the way to the highly-critical, involving plant, personnel and public safety. The technical report will show that well-designed radio systems can satisfy these varying needs, but that the associated tradeoffs of performance vs. security and reliability result in different system solutions for different applications.

In general, the radio links analyzed in the technical report are assumed to operate in the ISM bands, and to carry information in the form of digital bit streams. Discussions are limited to the physical layer and error-control portions of the data-link layer of the seven layer open systems interconnection (OSI) model. Wireless networking concepts are touched on only superficially, in the context of spectrum sharing and spectrum management; further details may be incorporated in a future publication. Where examples are given, they are based on the 2.4 GHz ISM band and the rules in effect in North America.

Section 2 of the technical report deals first with the performance of a single, one-way radio link in ideal “free space” conditions, assuming unobstructed line-of-sight propagation of the carrier waves and no sources of interference (such as other transmitters, electrical machinery or electronic equipment), the only sources of impairment being the thermal noise generated in the receiver and the reduction in signal strength due to path losses.

Section 3 of the technical report covers various aspects of radio wave propagation, including reductions in signal strength due to increasing distance from the transmitter and transmission through “lossy” (partially-conducting) media, reflection of electromagnetic waves at conducting, non-conducting or partially-conducting surfaces, multipath signal fading caused by reflections and techniques for mitigating these effects, and the design, selection and deployment of antennas.

Section 4 of the technical report deals with the “real world” issue of how multiple radio systems can “co-exist” without suffering unacceptable levels of radio frequency interference from other users sharing the same frequency bands. Topics discussed include the traditional regulatory approach to spectrum management, the constraints applicable to the unlicensed frequency bands, sources of interference and the effects of interference on radio performance, and techniques to mitigate the effects of interference.

The final section of the technical report addresses the issue of wireless communication standards and how they might be rationalized to facilitate the secure and reliable deployment of radio links in the industrial environment for applications in all categories from non-critical to highly critical.

2 Free space communications basics

2.1 Elements of a radio link

Radio (or wireless) links make use of electromagnetic radiation covering a band of frequencies made up of channels. The width of a channel used is determined by the amount of information to be transmitted in a given time and the modulation and filtering design of the radio. A higher data rate generally requires a wider channel when using a particular modulation format. Modulation is the method by which the data is impressed onto the radio signal. Modulation techniques which impress more data onto a given band are generally more complicated to implement and less immune to propagation impairments and interference. Regulations set the acceptable channel widths and in some cases even the modulation types that may be used in a given band. Narrow band radios normally stay on a single center frequency and occupy only the bandwidth required to accommodate

their data transmission requirements. In order to more effectively share the frequencies, radios operating in unlicensed bands typically use additional techniques to avoid generating interference and improve their resistance to interference. Two of the more popular methods of operation within the ISM bands are frequency hopping spread spectrum (FHSS) and direct sequence spread spectrum (DSSS). Frequency hopping radios move a narrow band around to different frequencies. Direct sequence radios modulate a narrowband signal with an additional wide band random noise-like signal that has the effect of spreading it over a wider band of frequencies.

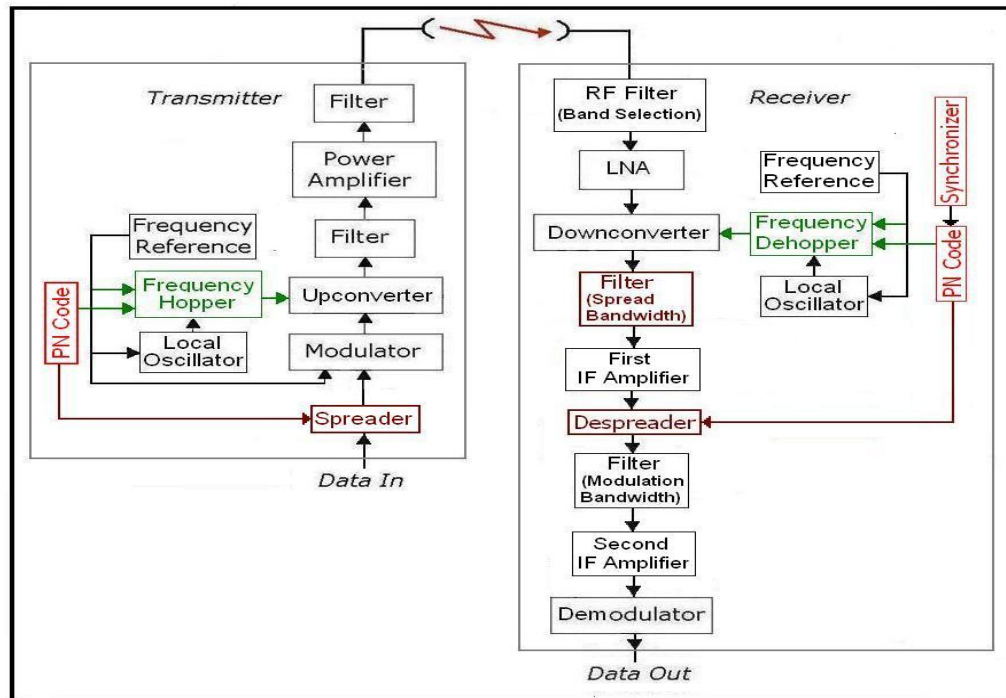


Figure 2.1 – Block Diagram of a Generic Radio Link

There are newer techniques such as Ultra Wide Band and Chirp transmission that are emerging and in the future there will undoubtedly be continuing development and innovation. Another popular technique is Orthogonal Frequency Division Multiplexing (OFDM) which generates a large number of narrow signals adjacent to each other sharing the overall data payload.

Figure 2.1 shows the block diagram of a generic radio link, which we will assume to be operating in one of the ISM bands. The modules shown as non-shaded will typically be found in both narrow-band and spread spectrum radios. By inserting the shaded modules “PN Code” and “Spreader” on the transmit side and “Synchronizer”, “PN Code” and “Despreader” on the receiver side, we can convert this narrow-band radio into a DSSS radio. If we instead insert the shaded modules “PN Code” and “Frequency Hopper” on the transmitter side and “Synchronizer”, “PN Code” and “Frequency Dehopper” on the receiver side, we convert it to a frequency-hopping spread spectrum (FHSS) radio. Sections 2.1 through 2.5 deal primarily with the general case of a radio link carrying digital data. Direct sequence and frequency-hopping spread spectrum technologies will be discussed in section 2.6.

2.1.1 A note about filters

Filters will be a recurring topic in this technical report. They are extensively used in radio communications, and generally perform one of two basic functions:

- to allow certain frequencies to pass (the “pass band”) while rejecting (or attenuating) other frequencies (the “stop” band or “attenuation band”);
- to control the shape of the transmitted signal spectrum.

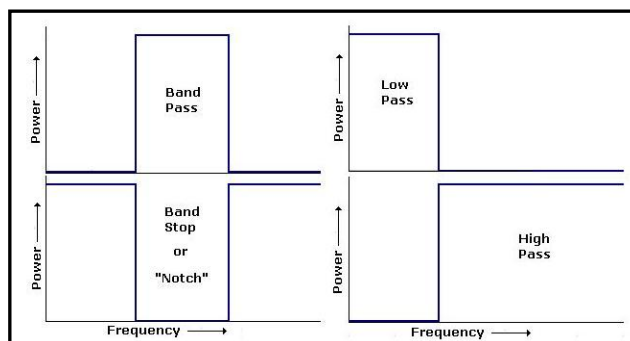


Figure 2.2 – Common Filter Types

Filters in the first of these categories are usually described in terms of their characteristics, as shown in Figure 2.2: “bandpass”, “lowpass”, “highpass”, and “bandstop” or “notch”. The five filters shown in Figure 2.1 usually will be implemented as bandpass filters. We will discuss these filters in more detail later in the technical report.

For many radio applications, the rectangular characteristics shown in Figure 2.2 would be ideal, but cannot be realized in practice, although complex filters are capable of achieving quite close approximations. Typically, the “cut-off characteristic will be a smooth “roll-off”, the stop band attenuation will be finite, and both the pass band and the attenuation band will exhibit ripples. See Figure 2.3.

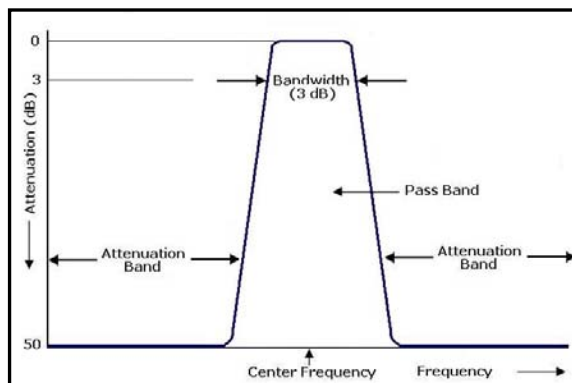


Figure 2.3 – Filter Terminology

An important characteristic of a bandpass filter is its “quality factor” or “Q”, which is defined as the center frequency of a filter divided by the bandwidth, where the bandwidth is defined as the frequency of the upper 3 dB* roll-off point minus the frequency of the lower 3 dB roll-off point. For example, a

filter with a center frequency of 2.400 GHz and a 3 dB bandwidth of 100 kHz has a Q of $(2,400/0.1) = 24,000$. Such a high-Q filter would be extremely difficult to design and implement. When we come to discuss the receiver design shown in Figure 2.1, we will see how radio designers have traditionally dealt with this problem.

(* For an explanation of dB (decibel) power measurement, please see the power amplifier section below.)

We will discuss spectrum shaping filters in the “Modulation and Spectrum Bandwidth” section below.

2.1.2 Transmitter

Although it is not mandatory under the regulations, most transmission standards used in the ISM bands divide the total frequency range of the band into several narrower, equally-spaced “channels”. The radio transmitter generates a radio frequency (RF) electromagnetic wave (the “carrier”) centered on one of these channel frequencies and superimposes the data signal on the carrier by a process known as “modulation”. The transmitter must ensure that the carrier’s transmitted power, frequency and spectrum bandwidth comply with the specifications set by the regulatory agencies.

2.1.2.1 Data signal coding

In Figure 2.1, in the absence of the DSSS spreader, the input data signal is assumed to be applied directly to the modulator. In reality, it is usually subjected to some preliminary “coding” to add “overhead” data bits for receiver synchronization, error correction, and error checking. In this technical report, we will generally ignore such overhead bits. We will use the symbol “R” to represent the data signal bit rate in bit/s.

2.1.2.2 Pseudo noise (PN) code generator

This module generates a pseudorandom digital sequence used to spread the bandwidth in DSSS systems and to generate the hop frequencies in FHSS systems. Detailed discussion of this module is deferred to section 2.6.

2.1.2.3 Frequency reference

Although the regulators do not require ISM band radios to comply with specific frequency plans, the center frequency of the transmitted carriers must be accurate enough to allow the transmitters and receivers to “find” one another. Requirements are more stringent for radios specified over a wide temperature range and for systems operating at lower bit rates. In such cases, frequency accuracy and stability are generally achieved by locking the various frequencies used to generate the carrier to a crystal-controlled frequency reference.

2.1.2.4 Modulation and spectrum bandwidth

Radios transfer information by changing the value of some parameter of the carrier signal. This process is generally known as “modulation”, but the expression “keying” (from the early days of wireless telegraphy) is also used. The carrier parameter varied may be its amplitude (amplitude modulation, AM, or amplitude-shift keying, ASK), frequency (frequency modulation, FM, or frequency shift keying, FSK), phase (phase modulation, PM, or phase shift keying, PSK) or a combination of these, such as quadrature amplitude modulation, QAM, which transfers data by independently changing the amplitude of two carrier waves which are 90° out of phase with each other. In this technical report, we will limit our discussions to the single-level modulation schemes like FSK, binary phase shift keying (BPSK) or Gaussian minimum shift keying (GMSK) typically used in ISM band systems. Frequency shift keying switches the frequency transmitted between different values to represent the data signal. Gaussian FSK is a variation of this that controls the rate of shift between

the frequencies. BPSK switches the phase of the RF carrier between 2 points which are typically 180° apart to represent data 1s and 0s. GMSK is a filtered version of minimum shift keying (MSK) that in itself is a form of phase shift keying. GMSK first filters the digital waveform first using a special lowpass filter with a smooth “Gaussian” shaped roll-off, and then frequency modulates the carrier with the filtered waveform.

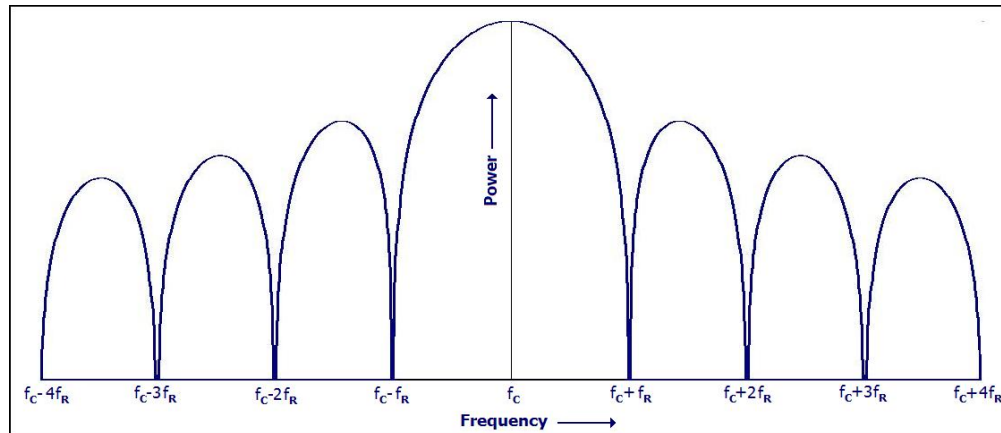


Figure 2.4 – Typical Spectrum of Digitally-Modulated Carrier

Modulating a carrier with a digital bit stream produces a signal centered at the carrier frequency f_C with a power vs. frequency characteristic like that shown in Figure 2.4. This is called the frequency spectrum of the modulated carrier signal. The spectrum shape of a modulated carrier wave is determined principally by the data rate, the choice of modulation scheme and the filtering before or after the modulator. With BPSK modulation, the bandwidth of the main lobe of this spectrum from null to null equals $(2 \times f_R)$, where f_R means a frequency in hertz equal to the data rate in bit/s.

(The sidebands on each side of the main lobe are not actually cut off as shown in Figure 2.4, but continue outwards with ever-decreasing amplitudes.)

It is not necessary to transmit the sidebands, or even the full main lobe, in order to recover the data at the receiver. Bell Labs researcher Harry Nyquist showed that all the information is contained within half the width of the main lobe, as shown in light blue in Figure 2.5. This is known as the Nyquist bandwidth. Any bandwidth wider than half the main lobe is called “excess Nyquist bandwidth”, and is given the symbol α (alpha). Practical receivers typically look at something between 75% and 100% of the main lobe (50% to 100% excess Nyquist bandwidth, shown in yellow in Figure 2.5). In this technical report, we shall call this the “modulation bandwidth”.

Since the sidebands are superfluous to the recovery of the data and the energy contained in them can cause interference in other receivers, transmitting them amounts in effect to “polluting” the RF environment. For this reason, most standards put limits on the sideband amplitudes that can be transmitted. For example, the IEEE 802.11b standard requires that the signal be at least 30 dB down

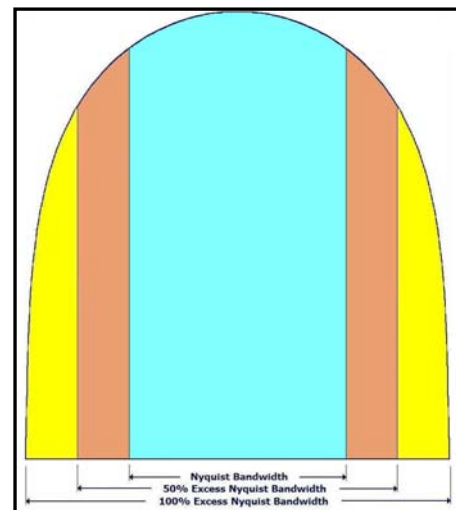


Figure 2.5 – Nyquist

from its peak energy at ± 11 MHz from the center frequency and at least 50 dB down from its peak energy at ± 22 MHz from the center frequency.

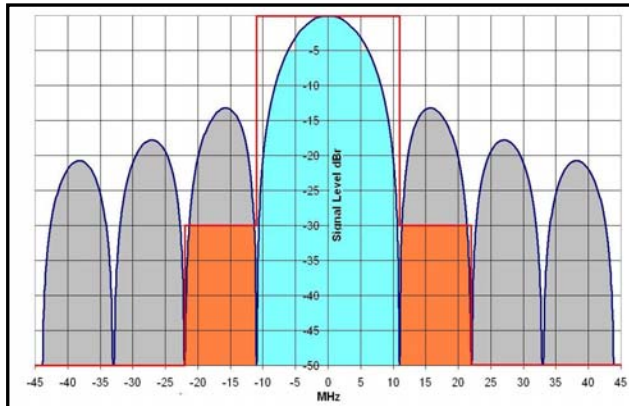


Figure 2.6 – IEEE 802.11b Spectral Mask

This requirement, referred to as the “spectral mask”, is shown by the red line in Figure 2.6. The blue line shows the unfiltered spectrum. The gray-shaded areas represent the spectrum energy that must be removed to comply with the mask. This is generally achieved by filtering the digital bit stream at the input to the modulator with a special type of low-pass filter. The filter is designed to remove the higher frequency components of the square wave data pulses so that the carrier satisfies the legal requirements of the regulations and the data signal can still be recovered from the carrier’s main lobe.

When we discuss spectrum sharing and coexistence later, we will show that the orange areas in the first sidelobes have the potential to cause significant interference.

2.1.2.5 Upconverter and filter

In Figure 2.1, the modulator impresses the data signal on an intermediate frequency (IF) carrier signal, which is then mixed with a stable local oscillator (LO) signal (“upconverted”) to generate the modulated RF carrier signal. The IF and LO signals are both derived from the stable frequency reference. The mixer generates both the sum and the difference frequencies, so a filter must be used between the modulator and the power amplifier to filter out both the LO frequency and the unwanted “image frequency”. Since the filter must be wide enough to cover the full ISM band to permit frequency agility, the IF and LO frequencies must be chosen so that they are both outside the ISM bandwidth. The spectral purity of the oscillator also affects the amount of unwanted energy emitted into adjacent channels. Practical oscillators have some level of “phase noise” that has an effect outside of the control of the IF filter.

In some transmitter designs, the modulation may be applied directly to the RF carrier, in which case the local oscillator, upconverter and image rejection filter are not required.

2.1.2.6 Power amplifier

ISM band equipment must receive type approval certification from an accredited test lab before it is offered for sale, and must comply with all the specifications set by regulators. The certification documents will specify the maximum transmitter RF output power permitted for the approved device. The power is usually relatively low with the regulations varying depending on the jurisdiction. The allowed power in most instances is no more than 1 W or 100 mW with specific restrictions based on the modulation and bandwidth. These restrictions are intended to limit the possibility of interference with other users’ equipment operating in the same band. If the output power from the modulator or

upconverter is low, the radio may include an internal power amplifier to increase the transmitter output power up to the type approval level.

Operational range may sometimes be extended by using high gain antennas, but regulatory restrictions generally apply to this option as well.

Transmitter RF power output is generally specified in milliwatts (mW) or watts (W). In most operational radio systems, the power levels measured at the transmitter and receiver antenna ports differ by many orders of magnitude, so it is desirable to use the logarithmic decibel (dB) scale to measure and compare power levels. For example, the ratio of two power levels, P_1 and P_2 , is calculated by the following equation:

$$(P_1/P_2) \text{ dB} = 10 \log_{10}(P_1/P_2)$$

For the purposes of quick comparison it is useful to note that a 3 dB difference in power indicates a factor of 2 in power (3 dB more is double and 3 dB less is $\frac{1}{2}$). Also, a 10 dB difference is an order of magnitude (10 dB more is 10 times and 10 dB less is $\frac{1}{10}$). Further, a number of tens indicates the number of orders of magnitude (30 dB is a factor of 1000).

If P_2 is chosen to be some fixed reference power, such as 1 mW or 1 W, the decibel scale yields absolute power values:

$$P_{\text{dBm}} = 10 \log_{10} P_{\text{mW}}$$

where P_{dBm} is the power level expressed in dBm (dB referenced to 1 mW), and P_{mW} is the power level expressed in mW. The corresponding equation for a reference power of 1 W is:

$$P_{\text{dBW}} = 10 \log_{10} P_{\text{W}}$$

where P_{dBW} is the power level expressed in dBW (dB referenced to 1 W), and P_{W} is the power level expressed in W.

Some useful values for low-power ISM band radios are:

$$1 \text{ mW} = 0 \text{ dBm} = -30 \text{ dBW}$$

$$10 \text{ mW} = 10 \text{ dBm} = -20 \text{ dBW}$$

$$50 \text{ mW} = 17 \text{ dBm} = -13 \text{ dBW}$$

$$100 \text{ mW} = 20 \text{ dBm} = -10 \text{ dBW}$$

$$200 \text{ mW} = 23 \text{ dBm} = -7 \text{ dBW}$$

$$1000 \text{ mW} = 1 \text{ W} = 30 \text{ dBm} = 0 \text{ dBW}$$

(Another variant of the decibel scale is “dBi”. This unit is a measure of antenna gain, not power, and will be explained below.)

2.1.2.7 Transmit filter

The primary function of the transmit RF filter is to suppress harmonics of the carrier frequency at the output of the transmitter. To allow for frequency agility, the transmit filter is generally wide enough to pass the full range of frequencies in the ISM band.

2.1.2.8 Transmit antenna, antenna gain and spreading loss

The transmit antenna may be located internally as a printed circuit board trace or a lumped element component on the board, externally on the transmitter housing, or remotely, connected by a coaxial cable. The transmit antenna acts as the coupling mechanism between the transmission line and free space, allowing the carrier wave to radiate outwards from the transmitter location. It determines the polarization of the wave, and may also focus and direct it in a desired direction. The ratio of the power emitted by a given antenna, as measured at a point along the line of maximum power, to the power which would be measured at the same point, if transmitted by a hypothetical (spherically omnidirectional) “isotropic radiator” which radiates uniformly in all directions, is called the “antenna gain”, expressed in dBi (decibels relative to an isotropic source).

In general, a larger transmit antenna will have a higher gain and will focus the energy into a narrower beam, increasing the power density in the beam. In most cases a narrower beam reduces the threat of interference to receivers outside the beam; its higher power density means it is more likely to cause interference to receivers located in the beam. In the ISM band, the FCC has dealt with this by decreeing that the maximum Effective Isotropic Radiated Power (EIRP) cannot exceed 6 dBW (4 W). EIRP is calculated by multiplying the transmitter power at the transmit antenna input port by the antenna gain. A system with the maximum legal transmitter power of 1 watt at the antenna input is limited to a 6 dBi antenna gain. The European Telecommunications Standards Institute (ETSI), the regulator for Europe, permits only a much lower limit. The FCC has also allowed a specific exception for point to point links in the 2.4 GHz band, where higher EIRP using directional antennas is permitted.

In free space, the power density in the transmitted beam decreases in inverse proportion to the square of the distance from the transmitter. This computed reduction should be considered to be the minimum for unobstructed paths. In practice there are many other propagation effects that may cause a higher loss. Specific locations close to a transmitter may have little or no signal strength. In urban environments, crowded with obstructions, a more realistic average loss can vary with the fourth power of the distance. On the other hand, waveguiding in corridors can reduce the loss exponent to less than two. This power density reduction, referred to as “spreading loss” or as “path loss”, is the most significant loss factor in determining the maximum range capability for any given radio system. To transmit over a longer range, use as high-gain an antenna as necessary, provided that the EIRP limit specified by regulation is not exceeded.

Arrays of antennas may be used to achieve a better link through diversity. Newer techniques making use of multiple transmitters and receivers may be used to enhance a link through time and space diversity. The benefits of such techniques are balanced against a requirement for more hardware and more complicated algorithms which currently cost more to implement.

2.1.3 Receiver

2.1.3.1 Receive antenna

At the other end of the link, a fraction of the energy in the original wave is “picked up” by the receive antenna, which, like the transmit antenna, may be located internally as a printed circuit board trace or a lumped element component on the board, externally on the receiver housing, or remotely, connected by a coaxial cable. Depending on the specific application and constraints on physical size, the receive antenna design may be “directional”, to maximize the signal power received from a fixed direction, or “omni-directional”, to allow reception from any direction. In general, larger antennas will be more directional and will capture more energy from the received wave (i.e. have higher gain). In one-way links, where antenna gain is not constrained by the limits on transmitted EIRP, high-gain receive antennas can be used to achieve very long range reception.

2.1.3.2 Receiver RF filter, LNA, downconverter and IF amplifiers

The receiver shown in Figure 2.1 is an example of the “classical” configuration called “superheterodyne”, in which the received RF signal is filtered, amplified and mixed down to a lower IF in a series of modules generally referred to as the receiver’s “front-end”. In section 4.7, we will show that this apparently complex design solves two fundamental problems in the physics of radio – the need to achieve very high degrees of amplification without overloading the amplifiers, and the need to reject all interfering signals outside a very narrow portion of the RF spectrum without having to build filters with unrealistically-high Q factors.

The RF filter minimizes the risk of interference from other transmitters by rejecting signals outside the receiver’s operating band. In ISM band radios, the RF filter is usually wide enough to accept the whole ISM band. The low noise amplifier (LNA) has a relatively low gain, to ensure linear operation, well below saturation. The amplified signal is then down-converted (mixed with a stable local oscillator frequency derived from the frequency reference subsystem) to shift it to a lower center frequency, after which it must be filtered to remove the local oscillator and image frequencies and amplified further. The spectral purity of the oscillator used to perform the first downconversion is also a factor in ability of the receiver to reject strong signals in adjacent channels. An effect known as reciprocal mixing causes adjacent signals to spread and it is not correctable by the IF filter. Ideally the oscillator will be a single pure tone but practical signal sources usually have some level of “phase noise” which appears as a form of modulation. In a DSSS radio, the bandwidth of the second filter is chosen to pass the main lobe of the spread spectrum signal and reject all energy outside it. The third filter has a bandwidth equal to the modulation bandwidth, after which the remaining signal is amplified a third time to the level required by the demodulator. (In narrow-band and FHSS radios, the second filter is not required.)

Overall, the purpose of the filters is to reject out-of-band interferers. High quality receivers may have substantial amounts of filtering, while simple receivers may have next to none.

2.1.3.3 Demodulator

The demodulator, also referred to as the detector, is the “heart” of a radio system. Along with the modulation bandwidth filter, it is largely responsible for determining the performance quality of the radio link. Its prime function is to decide for every data bit whether the transmitter sent a “one” or a “zero”. In the ideal noise-free case, assuming that the full spectrum is transmitted (as shown in Figure 2.4, but including the additional “further-out” sidebands), the demodulated waveform would be made up of perfectly square pulses, so the bit values could be estimated with 100% accuracy. In practice, the filtering applied in the transmitter to force the transmitted signal into compliance with the spectral mask (see Figure 2.6) “rounds off” the square edges of the received data signal. One way to compensate for this effect is to sample the pulses in the middle of the symbol period.

In the presence of noise, the waveform is “fuzzy” to a greater or lesser degree, and the task of making the correct “one” or “zero” decision becomes more and more difficult as the signal power is reduced and/or the noise power is increased. The task can be simplified, if the receiver can establish a good model of the remote transmitter, specifically if an accurate copy of the transmitter data timing clock can be recovered from the received data signal to permit synchronous data detection, and a good frequency and phase reference can be derived to permit coherent detection. These reference signals are usually recovered by phase-locking local clock and carrier reference oscillators in the receiver to the received signal, using fairly long time constants to derive the best possible averages. A synchronous, coherent receiver will have a better clock reference, allowing it to sample the received signal at the optimum time instant for accurate detection, and a better carrier reference signal, permitting more accurate estimates of the information contained in the received signal. These characteristics are generally found in conjunction with other higher-quality design features, such as more tightly-matched filters, lower-threshold noise detectors, etc. As a result, “good” receivers cost

more than ordinary receivers, but deliver significantly better performance, in terms of operating range and interference tolerance. The receiver designer has many design variables to play with, and tradeoffs must be made between cost, complexity and performance.

2.1.3.4 PN code generator and synchronizer

The PN code generator is used to despread the received signal in DSSS systems and to de-hop it in FHSS systems. Both these systems require that the PN code generated in the receiver is precisely synchronized with that used in the transmitter. This function is implemented by the synchronizer module. Detailed discussion is deferred to section 2.6.

2.2 Noise and interference in radio systems

Noise and interference in radio systems can be broadly defined as “unwanted disturbances superimposed upon a useful signal that tend to obscure its information content”. The term “noise” is generally used to denote disturbances which result from basic physical properties of the devices used in building the radio equipment. When we refer to noise in this technical report, we mean thermal noise [which is due to the fact that every object with a temperature higher than “absolute zero” (0° kelvin, or -273° C) radiates electromagnetic energy, the power and frequency of which is a function of its temperature] and receiver noise, which includes the effects of random non-thermal movement of electrons in the amplifier and detector devices used in the receiver. Interference can be a problem even at high received signal levels, but noise is usually only a concern when signal levels are low.

The thermal noise power N in a bandwidth B hertz produced by a resistor at temperature T kelvin is computed using Boltzmann's equation:

$$N = kTB$$

where k is Boltzmann's constant ($1.380658 \times 10^{-23} \text{ J K}^{-1}$, where J K^{-1} is joules per degree kelvin).

Example: Assuming the standard reference temperature of 290° K (17° C), the thermal noise power, N , in a 1 hertz bandwidth is given by:

$$N = 1.380658 \times 10^{-23} \times 290 \text{ W/Hz} = 4.045328 \times 10^{-18} \text{ mW/Hz} = -174 \text{ dBm/Hz}$$

Receiver noise specifications are usually quoted as a noise figure, which is defined as the ratio of the total noise power per unit bandwidth at the detector input to the noise power at the detector input due to the thermal noise at the receiver input, when the input termination is held at the reference temperature of 290° K. Noise figure is expressed in dB and computed as follows, where all measurements are made at frequency f :

$$(\text{noise figure})_f = 10 \log_{10} [P_{\text{noise out}} / (G \cdot P_{\text{noise in}})] \text{ dB}$$

where $(\text{noise figure})_f$ is the noise figure at frequency f , G is the system power gain ($10 \log_{10} (\text{power out/power in})$), $P_{\text{noise out}}$ is the total noise power in watts per unit bandwidth at the input to the detector, and $P_{\text{noise in}}$ is the thermal noise power in watts at the receiver input.

Example: Assuming unity gain and a noise figure of 6 dB, the effective receiver noise power per hertz for an input noise temperature of 290° K will be -168 dBm/Hz.

The thermal noise power density at the detector input is denoted by the symbol N_0 .

The other source of “unwanted disturbances” that radio systems have to deal with is interference. Whereas thermal noise and receiver noise are implied by the laws of physics and the radio engineer can only strive to minimize their effects by good system design, interference is predominantly man-

made in origin. Human applications of electrical energy result in the intentional emission of electromagnetic radiation, such as signals from broadcast radio, television, telecommunications, radar and navigational system transmitters, and in unintentional emissions, including energy radiated from electrical machinery, medical diagnostic systems, microwave ovens, automobile ignition systems, and diverse electronic equipment.

In the unlicensed frequency bands, the most serious sources of interference are likely to be other users operating locally in the same band. It is convenient to define an interference power density I_0 , analogous to the thermal noise power density N_0 , where I_0 is the total in-band interference power divided by the receiver bandwidth. The receiver performance depends on the total noise and interference power density at the detector input, $(N_0 + I_0)$, where the two quantities are added on a power sum basis.

2.3 The effect of noise on radio link performance

The ultimate task of the receiver is to determine for every data bit if the transmitter has sent a “one” or a “zero”. If every such decision is made correctly, the bit error ratio is zero, and the original data signal has been transmitted over the link perfectly. Imperfect detection due to noise results in transmission errors, generally quantified as the bit error ratio (BER, often referred to incorrectly as the “bit error rate”), which is defined as the number of erroneous bits received during some stipulated time period divided by the total number of bits received in the same period. Other measures that are used are FER, frame error rate, and PER, packet error rate, which measure errors within a defined frame or packet of data bits. Where a re-transmission protocol requires a repeat of a frame or packet of data containing an error, this latter can be a useful measurement.

In analog radio systems, one of the most important measures of performance is the signal to noise ratio, S/N , where S is the power level of the desired signal, and N is the total noise power level in a defined bandwidth. The equivalent single-parameter performance metric commonly used in analyzing digital radio systems is E_b/N_0 - the ratio of the energy per bit, E_b , to the noise power per hertz, N_0 , measured at the receiver detector input.

If P_R is the signal power in watts at the input to the receiver, and R is the data rate in bits per second, E_b is given by:

$$E_b \text{ joules per bit} = (P_R \text{ joules per second}) / (R \text{ bits per second}).$$

If there is no interference, the BER of a given link is a function of the E_b/N_0 at the detector input. N_0 is determined by the noise figure of the receiver, so can be taken as fixed.

As the E_b drops due to reduced signal strength, the statistical probability that the detector will make an incorrect decision increases progressively, and the BER will increase, gradually at first, and then faster and faster, until the receive signal strength crosses the receiver “threshold” value, below which the performance is deemed to be unacceptable.

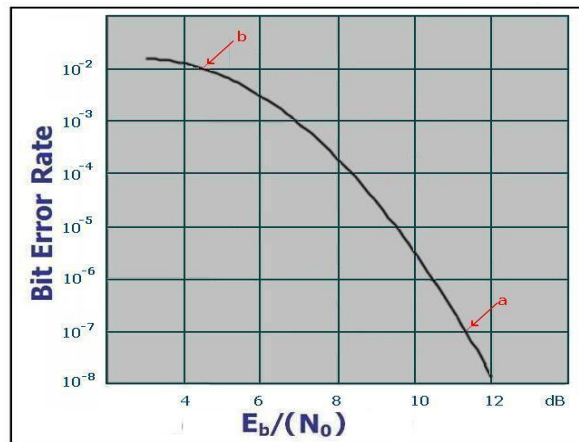


Figure 2.7 – Waterfall Curve

Graphs showing this effect for a given system are generally known as “waterfall curves”. A typical example is shown in Figure 2.7. (The numerical values for BER and E_b/N_0 in the figure are not intended to represent any specific receiver system. They have been included to simplify the following discussion.) Generally speaking, the shape of the waterfall curve is determined by the modulation scheme used in the radio, while its position along the x-axis is a function of the quality of the receiver design.

If we assume that the curve is for a moderate quality synchronous, non-coherent receiver, a higher quality synchronous, coherent receiver with a lower noise figure would exhibit a similar curve to the left of that shown in the figure (i.e., a lower error rate for the same E_b/N_0), while the curve for a simple asynchronous non-coherent receiver with a higher noise figure would again be a similar curve to the right of the given curve (a higher error rate for the same E_b/N_0).

If interference is negligible in an operating environment, the maximum link range a radio can handle for error-free performance will be noise-limited, and the waterfall curve will be an important tool when engineering the link. First, we must decide what metric we will use to define our threshold. This is generally application-dependent. For digitized voice or video, threshold is generally specified in terms of bit error ratio. For constant-length packet data, the packet error ratio may be a better measure of acceptable or unacceptable performance. If the data protocol in use resends errored packets, signal latency (transmission delay) will be increased and throughput will be reduced due to the retransmission, so a maximum latency value or a minimum throughput might be used to define receiver threshold. For simplicity in this discussion, we will stick to the bit error ratio. The BER value we choose to define threshold will depend on the sensitivity of the end application to errors. A link used to monitor the water level in a mountain reservoir should be able to live with a higher error ratio at threshold than one carrying critical sensor data for a process control system.

Let's assume that we have engineered our link so that under ideal conditions it will be operating at point “a” on the curve in Figure 2.7, with an E_b/N_0 of around 11.3 dB and a nominal BER of 10^{-7} (normally a very acceptable error ratio). If our signal level dropped by 7 dB, our E_b/N_0 would be shifted to point “b” on the curve, and the link performance would be seriously degraded, from one error in every 10 million bits to one error in every 100 bits ($BER = 10^{-2}$). In reality, most receivers of the types we are discussing in this technical report would lose clock or carrier synchronization at such a high error ratio, causing the link to go down. A more realistic link design might set the threshold BER at 10^{-5} , giving us a threshold E_b/N_0 of about 9.5 dB. It is standard practice to choose the link's nominal operating point to provide a substantial reserve of power above the threshold level as a