

WILEY

无线通信

(第二版)

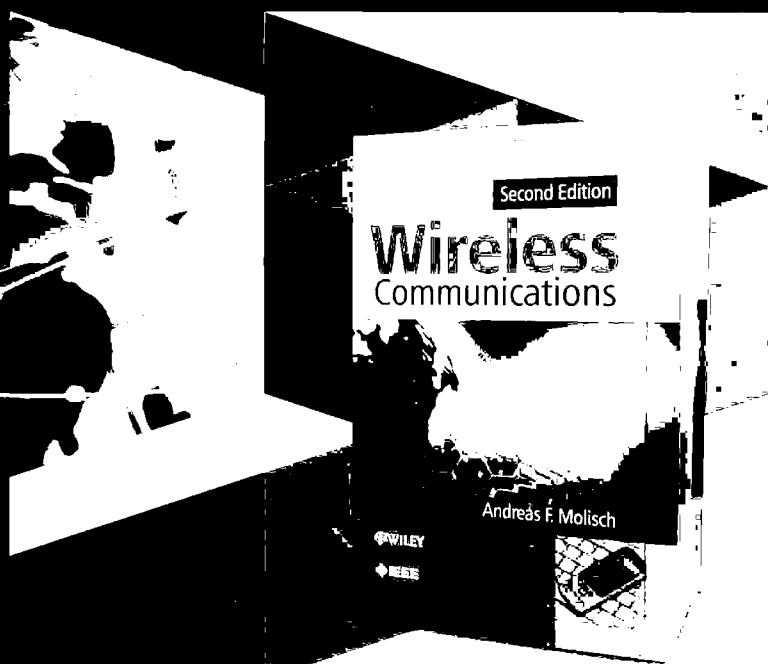
Wireless Communications
Second Edition

【美】Andreas F. Molisch 著

田 斌 帖 翊 任光亮 译

信息与通信技术

Wireless Communications
Second Edition



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>



无线通信 (第二版)

Wireless Communications Second Edition



移动通信技术原理与应用的集大成之作

无线通信在过去20多年获得蓬勃发展,无线通信权威专家Andreas F. Molisch在本书中讨论了无线通信领域的最新研究和发展。

本书深入分析了无线通信领域的当前技术,全书分五大部分,共计29章。既包括传统主题,如瑞利衰落、平坦衰落信道下的误比特率及均衡,又包含近来新兴的主题,如CDMA系统中的多用户检测,MIMO系统和认知无线电。书中也涵盖了主流的技术标准,包括蜂窝系统、无绳电话和无线局域网。

本书特色

- 讨论无线传播信道,收发信机和信号处理,多址和高级收发信机方案,标准无线通信系统;
- 将数学描述与直观的物理解释相结合,使读者对概念的理解一目了然。
- 新增认知无线电、协作通信与中继、视频编码、3GPP长时演进和WiMax等章,并新增多用户MIMO、802.11n和信息论等重要小节。
- 配套网站(www.wiley.com/go/molisch)包含有关DECT的辅助材料、习题参考答案、教学PPT、补充附录、缩略词表和其他有用资源。

作者简介 Andreas F. Molisch博士

美国南加州大学电气工程系教授,无线设备与系统(WiDeS)团队领导者。IEEE会士、IET会士、IEEE杰出讲师。曾在美国贝尔实验室、瑞典隆德大学、三菱电子研究实验室、维也纳技术大学任资深主任研究员、教授等职。出版及合著4部著作,参加撰写11部书,发表120篇以上的期刊论文及大量会议论文,拥有70多项专利和60个标准。是国际无线电科学联合会委员会主席,IEEE 802.16m、IEEE 802.15.3a、IEEE 802.15.4及COST273a信道模型小组的主席,IEEE 802.11n,多带OFDM联盟可测量性小组主席。曾任大量期刊和特刊编辑、多个国际会议的大会主席及技术委员会主席。

其他相关图书
请扫二维码↓



WILEY

Copies of this book sold without a Wiley sticker
on the cover are unauthorized and illegal

ISBN 978-7-121-24989-1



9 787121 249891 >



策划编辑: 马 岚
责任编辑: 马 岚
封面设计: 一克米工作室

定价: 109.00元

经典译丛·信息与通信技术

无线通信

(第二版)

Wireless Communications
Second Edition

[美] Andreas F. Molisch 著

田斌 帖翊 任光亮 译

电子工业出版社
Publishing House of Electronics Industry

内 容 简 介

本书系统讲解无线通信原理、技术和系统设计所涉及的各个方面。全书内容共分五部分(合计 29 章),包括无线传播信道的机制、特性、建模与探测,通信收发信机的调制、分集、信道编码、语音编码和均衡技术,多址与蜂窝、OFDM、扩频技术、多天线技术、认知无线电、中继与协作通信、视频编码,以及当前主流和最新出现的无线标准系统。本书既包含无线通信的通用处理技术,又包含当前热门的研究专题。书中的讲解凸显数学描述和直观物理解释相结合,深入浅出,给出的大量例题和习题(为第 30 章)取自当前主流无线通信系统和标准的实际案例。

本书适合作为通信工程和电子信息类相关专业高年级本科生、研究生和实践工程师的教材,更可作为无线通信工程师和科研人员案头必备的参考书。

Wireless Communications, Second Edition, 9780470741870, Andreas F. Molisch

Copyright © 2011, John Wiley & Sons, Ltd.

All rights reserved. This translation published under license.

AUTHORIZED TRANSLATION OF THE EDITION PUBLISHED BY JOHN WILEY & SONS, Ltd.

No part of this book may be reproduced in any form without the written permission of John Wiley & Sons, Ltd.

本书简体中文字版专有翻译出版权由英国 John Wiley & Sons, Ltd. 授予电子工业出版社:未经许可,不得以任何手段和形式复制或抄袭本书内容。

版权贸易合同登记号 图字:01-2011-3070

图书在版编目(CIP)数据

无线通信:第2版/(美)莫利斯(Molisch, A. F.)著;田斌,帖翊,任光亮译.

北京:电子工业出版社,2015.1

(经典译丛·信息与通信技术)

书名原文:Wireless Communications, Second Edition

ISBN 978-7-121-24989-1

I. ①无… II. ①莫… ②田… ③帖… ④任… III. ①无线通信-高等学校-教材 IV. ①TN92

中国版本图书馆 CIP 数据核字(2014)第 278987 号

策划编辑:马 岚

责任编辑:马 岚

印 刷:三河市鑫金马印装有限公司

装 订:三河市鑫金马印装有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:787×1092 1/16 印张:45.5 字数:1165 千字

版 次:2008 年 1 月第 1 版

, 2015 年 1 月第 2 版

印 次:2015 年 1 月第 1 次印刷

定 价:109.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zltz@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

译者序

无线通信近十余年来获得蓬勃的发展,已成为社会信息化最重要的推动力量,也是当前信息产业的重要支柱。无线通信技术极大地影响着人们的工作方式、生活方式和娱乐方式,个人通信的梦想指日可待。无线通信已经成为通信工程相关专业的一门重要课程,相关研究文献、论著、教材不断问世。无线通信领域的顶级专家、美国南加州大学 Molisch 教授的这本《无线通信》无疑是这方面的上乘之作。

该书 2005 年底出版了第一版,当时已堪称最全面清晰的集大成之作,中译本也于 2008 年初出版。该书分五个部分,第一部分讲述了无线业务的应用和需求,无线通信的技术挑战,以及噪声受限和干扰受限系统;第二部分论述了传播机制、信道统计描述、宽带和方向性信道、信道建模、信道探测以及天线,体现了无线通信信道传播的所有方面;第三部分则描述了构建一个无线系统的收发信机设计,包括基本方案、调制解调和信道编译码、语音编码和信道均衡;第四部分则是多址和高级收发信机方案,讲述了多址与蜂窝、扩频、OFDM 及多天线系统等最新专题;第五部分则描述了当前主流的无线通信系统和标准。

第二版紧跟技术发展,新增了认知无线电、协作通信与中继、视频编码、3GPP 长期演进和 WiMax 等章,以及多用户 MIMO、802.11n 和信息论等重要小节,并结合当前的技术发展对全书其他部分进行了进一步修订和补充。有些章节内容尽管变化不大,却补充了大量习题,对于理解内容及实际中的应用有很大帮助。同时,从方便阅读的角度进行了不少修改,比如不少缩略词用全称代替(有些我们在第一版翻译时已经做过类似的处理),专业术语较多的几章,直接将专业术语的缩略词表附在该章后面。除此之外,作者花了较大心力进行推敲和润色,有的地方看似仅仅改动了一个虚词、换了一个近义词、增加了一个修饰语,或者增删了一点说明,文字的语气和表达的含义则有明显的变化。我们在第二版翻译时也都逐字逐句核对了这些细微变化,反复推敲译文文字,以求精确表达作者的本来含义。

第二版保持并强化了第一版内容系统完整、概念清晰、理论严谨而又直观形象的特色,突出了技术发展的最新成果,涵盖了无线通信的最新研究内容与前沿领域,既是通信工程领域的本科和研究生的绝好教材,更可作为无线通信领域研究人员和工程技术人员案头必备的参考书。

本书第 4 章至第 7 章、第 14 章、第 15 章、第 21 章、第 23 章和第 26 章由田斌翻译,第 1 章至第 3 章、第 8 章、第 9 章、第 17 章、第 18 章、第 22 章、第 24 章和第 25 章由帖翊翻译,第 10 章至第 13 章、第 16 章、第 19 章、第 20 章、第 27 章至第 29 章由任光亮翻译,第 30 章为前面各章的习题,其译文由相应各章的译者分别完成,田斌负责统筹全书。原书有误的地方大都在译者注中标出,以便读者可能阅读原著时参考,个别出现在公式中的错误则直接更正过来。本书作者 Molisch 教授非常重视本书的翻译工作,多次和译者进行邮件交流。本书所有的中文译稿都寄给过他,他邀请的中国学生和学者阅读后再和他交流,提出了不少有益的建议。

西安电子科技大学研究生王奇伟、成波、丛犁、任晓娜、沈东阳、彭任斌、闵泉云、李奕洋、李慕菡、吴迪、龚阳雪、李林甫、冯妙琨、白渊玲等协助做了大量工作。借此机会，一并表示诚挚的谢意。尽管译者尽了最大努力，但由于译者水平和时间所限，肯定存在疏漏和不当之处，敬请读者不吝指正，我们发现错误会及时在 <http://web.xidian.edu.cn/btian> 上提供中译本勘误表。本书原著的教辅材料和勘误表可以参考 www.wiley.com/go/molisch。

译 者

于西安电子科技大学综合业务网国家重点实验室

前言和致谢

从 2005 年本书推出第一版至今,无线通信的研究与技术持续处于势不可挡的进步之中。这一发展状况以及本书第一版所获得的积极回应促成了第二版的问世,这一版将涵盖近年来出现的一些新的课题。这样,我们就可以通过现在的版本使本书与现代化无线通信工程所涉及主题的宽广程度仍然保持一致。

新版本中有超过 150 页的新内容,涵盖了以下方面:

- 认知无线电(新增的第 21 章);
- 协作通信、中继和 ad hoc 网络(新增的第 22 章);
- 视频编码(新增的第 23 章);
- 3GPP 长期演进(新增的第 27 章);
- WiMAX(新增的第 28 章)。

此外,新版本对以下内容作了明显的扩充和增补:

- MIMO(在第 20 章中),尤其是关于多用户 MIMO 新增了一节的内容(20.3 节)。
- IEEE 802.11n(高吞吐量 WiFi),在 29.3 节。
- 编码(比特交织编码调制),在 14.5 节。
- 在 14.1 节、14.9 节及附录 17(见配套网站)中介绍了信息论。
- 信道模型:在附录 7(见配套网站)中有对标准化信道模型的更新内容。
- 许多处小改动和重新阐释,部分基于本书的读者和教师的反馈。

这些扩充对学生(及研究者)学习“新潮”技术至关重要。绝大多数增补内容对于有关高级无线概念和技术的研究生课程应当是最适合不过的了。此外,关于 LTE(或 WiMAX)的内容,在更为基础的课程中,也非常适合作为标准化系统的示例(例如,可以取代关于 GSM 或者 WCDMA 系统的讨论)。

就第一版而言,采用本书作为其课程教材的教师是可以获得演示幻灯片和试题解答手册的。这些材料同样可以从出版商配套网站 www.wiley.com/go/molisch 获得。这个站点包含针对本书所有读者的重要资源:包括勘误表、内容更新、额外的参考文献和其他类似的材料。

作者为新增内容的撰写付出了相当大的努力,同时也得到了南加州大学电气工程系主任 Sandy Sawchuk 的鼎力支持。特别感谢 Anthony Vetro,他撰写了关于视频编码的新增章节(第 23 章)。我也要感激那些热心地同意对新增内容进行评阅的专家们,他们是: Honggang Zhang 和 Natasha Devroye(第 21 章), Gerhard Kramer、Mike Neely、Bhaskar Krishnamachari(第 22 章), Erik Dahlman(第 27 章), Yang-Seok Choi Hujun Jin 和 V. Shashidar(第 28 章), Guiseppe Caire(第 14 章和第 17 章的新增内容), Robert Heath 和 Claude Oestges(第 20 章的新增内容),以及 Eldad Perahia(29.3 节)。像往常一样,任何残存错误的责任均在于我。

我也非常感谢南加州大学(USC)参加我多门课程的学生们,同样感谢世界各地的读者和学生们,他们曾提出过有助于内容更正和改进的各种建议。新增章节的习题由 Junyang Shen、Hao Fang 和 Christian Mehluehrer 编写。感谢 Neelesh B. Mehta 提供了关于 LTE 的多幅图片给我。

就第一版而言,来自 John Wiley & Sons 有限公司的 Mark Hammond 担任组稿编辑, Sarah Tilley 担任制作编辑。特别感谢 Laserwords 的 Dhanya Ramesh,她相当专业地完成了书稿的录入排版工作。

第一版前言

1994年,当我以讲义的形式为一门无线课程撰写本书的初稿时,其前言是以论证开设这样一门课程的必要性开始的。我详尽地解释了理解无线系统(尤其是数字蜂窝系统)对于通信工程师的重要程度。现在,10多年过去了,这样的论证看上去稍微有些离奇和过时了。无线产业已经成为电信产业中增长最快的部分,并且在世界上几乎不存在什么人不是某种形式的无线技术的使用者。从普遍存在的蜂窝电话到无线局域网,到正在快速增长的无线传感器,人们被无线通信设备包围着。

学习无线通信的关键性挑战之一就是影响这一领域的令人惊奇的宽广主题。传统上,通信工程师的关注点主要集中在像数字调制和编码理论这样一些内容上,而天线与电波传播的研究领域是与之完全分离的,人们甚至以为“这是两条永不相交的平行线”。然而,这种各行其道的方式对于无线通信的学习是不适用的。我们需要理解影响系统性能的所有方面,并能够使整个系统得以运作。本书就试图提供这样一个概括——集中于无线通信物理层的运作上。

另一个挑战在于,不仅实际无线系统,而且作为无线技术基础的科学也处于持续不断的变化之中。常常有人认为,无线系统快速地变化,而无线通信的科学基础保持不变,因而工程师可以依靠以前获取的知识来应付许多轮的系统革新,仅仅需要对他们的一套技能作出微小调整就可以了。这种想法听上去很好,可遗憾的是,它是错误的。例如,10年以前,像多天线系统、OFDM、Turbo码和LDPC码及多用户检测这样一些主题通常只是学术好奇心之所在,最多也只能成为博士阶段的课程;今天,它们不仅支配着主流的研究和系统发展,而且代表着学生和从业工程师们所必须掌握的至关重要的基础知识。我所希望的就是,通过对新的方面和更多“经典”主题的兼顾处理,本书可以向目前的学生和研究者提供将来仍然会有价值的知识和工具。

本书为高年级本科生和研究生,以及从业工程师和研究者撰写。尽管在本书相应章节的开始处会简短地回顾有关领域的通信基础知识,我们仍然假定读者是对通信基本理论(如调制、解调)和电磁场理论的基本内容有所了解的。本书的核心材料试图将学生提升到可以阅读更高级的专题文章甚至是研究论文的程度;对于所有想要更加深入的读者,大多数章节都包括“深入阅读”小节,其中列举了最重要的参考文献。本教材既包括数学公式又给出直观的解释。我坚信这种“双管齐下”的方式可以引导学生对材料作最深入的理解。除作为教材外,也希望本书能成为研究者和从业者的参考工具书。为达此目的,我尝试着使得每个孤立章节更容易阅读。所有的首字母缩写词在每一章头一次出现时都会给予解释(而不仅仅是在书中头一次出现时);符号表解释了公式中用到的符号的含义。频繁地交互引用也将有助于实现这一目的。

纲要

本书分为五个部分。

第一部分(引言)对无线通信给出了高层次的综述。第1章首先对不同无线业务进行了分类,并描述了各种不同应用在数据速率、覆盖范围、电能消耗等方面所形成的种种要求。这一章也包

括简要的历史回顾以及对无线通信的经济和社会因素的讨论。第2章描述了无线通信的基本挑战,如,多径传播和受限的频谱资源。第3章讨论了噪声和干扰如何限制无线系统的通信能力,以及链路预算如何作为简单的系统规划工具来提供关于可达覆盖范围和性能的初步设计依据。

第二部分描述无线传播信道和天线的各个方面。由于传播信道是通信得以发生的媒介,理解它对本书其余内容的理解是至关重要的。第4章描述了基本传播过程:自由空间传播、反射、绕射、散射和波导效应。人们发现信号从发射机到接收机可能有不同的传播路径,它们涉及到一种或多种上述传播过程,从而产生了许多多径分量。对多径传播效应采用统计描述的方式往往是比较适宜的。第5章针对窄带系统给出了一个统计表示,解释了小尺度(瑞利)衰落和大尺度衰落。第6章讨论了宽带系统的一些表示方法,以及在发射机和接收机处可以辨别多径成分方向的系统。第7章介绍了不同传播环境中的传播信道的特定模型,涵盖了路径损耗模型、宽带模型和方向性模型。因为所有的实际信道模型都是基于测量(或者为测量所证实)的,第8章总结了用于测量信道脉冲响应的多种技术。最后,第9章简要地讨论了无线应用中使用的天线,特别谈到在基站和移动台处的不同限制。

本书第三部分讨论无线收发信机的结构和理论。在第10章对射频收发信机的组件进行了简短的概括之后,第11章描述无线应用中用到的不同调制制式。对调制制式的讨论不仅包括数学公式和信号空间表示,也包括了对各种应用目的下其优、缺点的评估。第12章的主题是所有这些调制解调器在平坦衰落信道和频率选择性信道中的性能。在此,我们得到的一个关键性结论是:衰落会导致错误概率急剧上升,且增加发射功率并不是改善性能的适宜方式。随后两章以改善这种情况下的性能为动机,分别讨论了分集和信道编码,这两种措施对降低衰落信道下的错误概率非常有效。有关编码的章节也包括了对近年来获得广泛关注的近香农限的编码(Turbo码和LDPC码)的讨论。由于语音通信仍然是蜂窝电话和类似设备的最为重要的应用,第15章讨论了语音数字化的不同方式,通过压缩信息可以使语音在无线信道上高效传输。最后,第16章讨论了均衡器,它被用来降低宽带频率选择性无线信道的有害效应。本部分所有章节都讨论单个链路,即一个发射机与一个接收机之间的链路。

然后,第四部分考虑在给定区域内同时运作多条无线链路的客观需求。这个称为多址(multiple-access)的问题有着许多不同的解决方案。第17章讨论频率域的多址(FDMA)和时间域的多址(TDMA),以及分组无线电,后者对数据传输的重要性不断增长。该章也讨论了蜂窝原理和频率重用概念,这一概念构成了蜂窝系统以及许多其他大容量无线系统的技术基础。第18章描述了扩展频谱技术,尤其是采用不同扩频序列来区分不同用户的码分多址(CDMA)。该章也讨论了多用户检测,它是一种可以大大降低多址干扰影响的非常先进的接收机方案。第四部分的另一个主题是“高级收发信机技术”。第19章描述了OFDM(正交频分复用),这种调制方式可以在大时延扩展的信道中支持非常高的数据速率。最后,第20章讨论了多天线技术:其中,“智能天线”通常被设置在基站处,是一种具备复杂信号处理能力的多天线单元,可以用来降低干扰(作为其众多优点之一),因而也就增加了蜂窝系统的容量。而MIMO(多输入多输出)系统就更进了一步,它允许通过发射机的多天线单元来传输并行数据流,然后由接收机处的多天线单元接收和解调。这些系统相对于单个链路来说可以获得巨大的容量提升。

本书最后一部分描述标准化的无线系统。标准化至关重要,因为这样可以使不同厂商的设备能够协作,并可以支持用户跨越国界时系统的无缝隙运作。本书在第21章至第23章分别描述了最为成功的蜂窝无线标准,即GSM(全球移动通信系统),IS-95及其高级形式CDMA

2000 和宽带 CDMA(也称为 UMTS)。进而,第 24 章描述了最重要的无线局域网标准,即 IEEE 802.11。

本书的配套网站(www.wiley.com/go/molisch)含有一些我认为有用但放入本书会使之显得过于庞大的材料。特别地,各章的附录和关于 DECT(数字增强型无绳电话)系统这一最重要的无绳电话标准的补充材料,可以在该网址找到。

课程实施建议

本书涵盖了无线通信的全部内容,从非常基本的主题到相当高级的主题,所提供的素材远远超过了一个学期课程中所能讲解的份量。教师因此可以根据学生的不同水平和兴趣来自由地订制教学内容。书中在正文里包含了精选的例子,并且在书末(第 25 章)给出了大量的课后练习题,练习题的答案以及讲课用的幻灯片,教师可以从本书的配套网站上下载得到。

以下包括几种不同要求的课程的内容组织建议:

- 入门性课程:

- 引言(第 1 章至第 3 章);
- 无线信道基础(4.1 节至 4.3 节、5.1 节至 5.4 节、6.1 节、6.2 节、7.1 节至 7.3 节);
- 初级信号处理(第 10 章、第 11 章和 12.1 节、12.2.1 节、12.3.1 节、13.1 节、13.2 节、13.4 节、14.1 节至 14.3 节、16.1 节和 16.2 节);
- 多址和系统设计(第 17 章、第 22 章和 18.2 节、18.3 节、21.1 节至 21.7 节)。

- 无线传播

- 引言(第 2 章);
- 基本传播效应(第 4 章);
- 统计信道描述(第 5 章和第 6 章);
- 信道建模和测量(第 7 章和第 8 章);
- 天线(第 9 章)。

此课程也可以结合进更多的电磁场理论和天线的基本素材。

- 无线通信的高级主题

- 引言和复习:教师应根据听众的具体情况进行内容选择;
- CDMA 和多用户检测(18.2 节至 18.4 节);
- OFDM(第 19 章);
- 超宽带通信(18.5 节和 6.6 节);
- 多天线系统(6.7 节、7.4 节、8.5 节、13.5 节、13.6 节和第 20 章);
- 高级编码(14.5 节和 14.6 节)。

- 当前的无线系统

- 基于 TDMA 的蜂窝系统(第 21 章);
- 基于 CDMA 的蜂窝系统(第 22 章和第 23 章);
- 无绳系统(配套网站上的补充材料);
- 无线局域网(第 24 章);
- 以及依据听众的知识水平从之前章节中选出的有关基本理论的素材。

第一版致谢

这本书是我在无线通信领域多年教学和研究的结晶。这许多年来,我曾在两所大学(奥地利的维也纳理工大学(TUV)和瑞典的隆德大学(LU))任教,还曾在三个工业研究实验室(奥地利维也纳的FTW电信研究中心,美国新泽西州米德尔顿的AT&T(BELL)研究实验室,美国麻省剑桥的三菱电气研究实验室(MERL))工作过,并在这些地方与我的同事们共事,同时还与在欧洲、美国和日本的其他机构工作的众多研究者进行过合作。所有这些都深深影响了我对无线通信的认识,其影响所及都体现在这本书里。我对他们都深怀感激之情。首先,我第一个想要感谢的是奥地利无线通信的先驱和前辈 Ernst Bonek,是他开启了这项计划。我曾与他进行过不计其数的探讨,这些讨论涉及到本书以及先于本书的讲稿(这些讲稿曾用于在维也纳理工大学我们共同任教的课程)在理论内容和教学方法上的方方面面。没有他的建议和鼓励,这本书是永远不可能付梓的。我还要对我在维也纳理工大学时的同事和学生,特别是 Paulina Erätuuili、Josef Fuhl、Alexander Kuchar、Juha Laurila、Gottfried Magerl、Markus Mayer、Thomas Neubauer、Heinz Novak、Bernhard P. Oehry、Mario Paier、Helmut Rauscha、Alexander Schneider、Gerhard Schultes 和 Martin Steinbauer 所给予的帮助深表谢意。我在隆德大学的同事和学生同样为本书作出了非常大的贡献,他们是: Peter Almers、Ove Edfors、Fredrik Floren、Anders Johanson、Johan Karedal、Vincent Lau、Andre Stranne、Fredrik Tufvesson 和 Shurjeel Wyne。他们的贡献不光在于就素材如何组织表达方面给出过许多颇具建设性的建议,而且书中用到的图示和例子都出自他们之手,尤其是书中大多数练习题及其解答都是由他们完成的,而 19.5 节的内容正是基于 Ove Edfors 的思想。特别感谢来自格拉茨理工大学(GUT)的 Gernot Kubin,他完成了关于语音编码的第 15 章。我在 FTW、AT&T 和 MERL 的同事和上司: Markus Kommenda、Christoph Mecklenbraeuer、Helmut Hofstetter、Jack Winters、Len Cimini、Moe Win、Martin Clark、Yang-Seok Choi、Justin Chuang、Jin Zhang、Kent Wittenburg、Richard Waters、Neelesh Mehta、Phil Orlik、Zafer Sahinoglu、Daqin Gu (他对第 24 章做出了很大的贡献)、Giovanni Vanucci、Jonathan Yedidia、Yves-Paul Nakache 和 Hongyuan Zhang 同样深深地影响了本书。除了许多具体的帮助和建议,更是为着他对本书所表现出的积极关注和所给出的弥足珍贵的意见,我向 Larry Greenstein 致以特别的谢意和感激之情。

同样特别感谢本书的诸位评阅人。原稿曾被出版社选定的匿名专家们和我的几个在不同研究机构任职的朋友和同事严谨地审读过: John B. Anderson (第 11 章至第 13 章)、Anders Derneryd (第 9 章)、Larry Greenstein (第 1 章至第 3 章、第 7 章、第 17 章至第 19 章); Steve Howard (第 22 章)、Thomas Kaiser (第 20 章)、Achilles Kogantis (第 23 章)、Gerhard Kristensson (第 4 章)、Thomas Kuerner (第 21 章)、Gerald Matz (第 5 章和第 6 章)、Neelesh B. Mehta (第 20 章)、Bob O'Hara (第 24 章)、Phil Orlik (17.4 节)、John Proakis (第 16 章)、Said Tatesh (第 23 章)、Reiner Thomae (第 8 章)、Chintha Tellambura (第 11 章至第 13 章)、Giorgia Vitetta (第 16 章)、Jonathan Yedidia (第 14 章)。我向所有这些人都致以至深的感激之情。当然,我应对仍可能存在的任何错漏担负全责。

出版人 Mark Hammond、项目编辑 Sarah Hinton 以及助理编辑 Olivia Underhill 都来自于 John Wiley & Sons 有限公司,他们曾以专业水平的建议和极大的耐心指导着本书的写作。Manuela Heigl 和 Katalin Stibli 充满激情而又小心谨慎地完成了大量录入和制图工作,Originator 公司非常专业地对文稿进行了排版。

目 录

第一部分 引 言

第 1 章 无线业务的应用和需求	2	1.3.4 能量消耗	16
1.1 历史	2	1.3.5 频谱的使用	17
1.1.1 一切是如何开始的	2	1.3.6 传输方向	18
1.1.2 第一个系统	3	1.3.7 服务质量	18
1.1.3 模拟蜂窝系统	4	1.4 经济和社会因素	19
1.1.4 GSM 及世界范围的蜂窝 革命	4	1.4.1 构建无线通信系统的经济 条件	19
1.1.5 新无线系统及电信泡沫 的爆裂	5	1.4.2 无线通信市场	20
1.1.6 无线通信的复兴	6	1.4.3 行为影响力	21
1.2 业务类型	7	第 2 章 无线通信的技术挑战	22
1.2.1 广播	7	2.1 多径传播	23
1.2.2 寻呼	8	2.1.1 衰落	23
1.2.3 蜂窝电话	9	2.1.2 符号间干扰	25
1.2.4 集群无线电	10	2.2 频谱限制	26
1.2.5 无绳电话	10	2.2.1 频率的分配	26
1.2.6 无线局域网	11	2.2.2 受管制频谱的频率重用	28
1.2.7 个域网	12	2.2.3 非管制频谱的频率重用	28
1.2.8 固定无线接入	12	2.3 能量限制	29
1.2.9 ad hoc 网络和传感器网络	12	2.4 用户移动性	30
1.2.10 卫星蜂窝通信	13	第 3 章 噪声受限和干扰受限系统	31
1.3 业务需求	13	3.1 引言	31
1.3.1 数据速率	14	3.2 噪声受限系统	31
1.3.2 覆盖范围和用户数目	14	3.2.1 链路预算	33
1.3.3 移动性	15	3.3 干扰受限系统	36

第二部分 无线传播信道

第 4 章 传播机制	40	4.2.1 斯涅尔(Snell)定律	41
4.1 自由空间衰减	40	4.2.2 分层电介质结构的反射和 透射	43
4.2 反射和透射	41		

4.2.3	d^{-4} 功率定律	45	6.2.2	一般的情况	84
4.3	绕射	46	6.3	无线信道的系统理论描述	86
4.3.1	单屏或楔形绕射	46	6.3.1	确定性线性时变系统的特性	86
4.3.2	多屏绕射	50	6.3.2	随机系统函数	87
4.4	粗糙表面的散射	54	6.4	WSSUS 模型	88
4.4.1	基尔霍夫理论	54	6.4.1	广义平稳	88
4.4.2	微扰理论	55	6.4.2	非相关散射	89
4.5	波导	56	6.4.3	WSSUS 假设	89
4.6	附录	56	6.4.4	抽头延迟线模型	90
	深入阅读	56	6.5	精简参数	91
第 5 章	无线信道的统计描述	57	6.5.1	相关函数的积分	91
5.1	概述	57	6.5.2	功率时延谱的矩	91
5.2	时不变两径模型	58	6.5.3	多普勒谱的矩	92
5.3	时变两径模型	60	6.5.4	相干带宽和相干时间	93
5.4	不含主导分量的小尺度衰落	61	6.5.5	窗参数	95
5.4.1	一个计算机实验	61	6.6	超宽带信道	96
5.4.2	振幅与相位统计量的数学推导	64	6.6.1	具有大相对带宽的超宽带信号	96
5.4.3	瑞利分布的性质	65	6.6.2	具有大绝对带宽的超宽带信号	97
5.4.4	瑞利分布场强的衰落余量	67	6.7	方向性描述	97
5.5	含有主导分量的小尺度衰落	68	6.8	附录	100
5.5.1	一个计算机实验	68		深入阅读	100
5.5.2	振幅和相位分布的推导	69	第 7 章	信道模型	101
5.5.3	Nakagami 分布	71	7.1	概述	101
5.6	多普勒谱	72	7.2	窄带模型	102
5.6.1	移动中的移动台的时变性	72	7.2.1	小尺度和大尺度衰落模型	102
5.6.2	固定无线系统中的时变性	74	7.2.2	路径损耗模型	102
5.7	衰落的时间依赖性	75	7.3	宽带模型	103
5.7.1	电平通过率	75	7.3.1	抽头延迟线模型	103
5.7.2	平均衰落持续时间	76	7.3.2	功率时延谱模型	104
5.7.3	随机频率调制	77	7.3.3	射线和簇到达时间模型	105
5.8	大尺度衰落	78	7.3.4	标准化的信道模型	106
5.9	附录	81	7.4	方向性模型	106
	深入阅读	81	7.4.1	一般模型结构及因子分解	106
第 6 章	宽带和方向性信道的特性	82	7.4.2	基站处的角度色散	107
6.1	概述	82	7.4.3	移动台处的角度色散	107
6.2	时延色散的成因	82			
6.2.1	两径模型	82			

7.4.4 极化	108	8.5 可分辨方向测量	127
7.4.5 模型实现	108	8.5.1 接收阵列的数据模型	128
7.4.6 标准化方向性模型	110	8.5.2 波束成形	129
7.4.7 多输入多输出矩阵模型	111	8.5.3 高分辨率算法	130
7.5 确定性信道建模方法	111	8.5.4 多输入多输出测量	132
7.5.1 射线发射	112	8.6 附录	133
7.5.2 射线跟踪	113	深入阅读	133
7.5.3 有效性考虑	114	第9章 天线	134
7.5.4 地理数据库	115	9.1 引言	134
7.6 附录	115	9.1.1 天线的系统集成	134
深入阅读	115	9.1.2 天线参量特征	134
第8章 信道探测	117	9.2 移动台天线	137
8.1 引言	117	9.2.1 单极子和偶极子天线	137
8.1.1 信道探测的必要性	117	9.2.2 螺旋天线	138
8.1.2 通用探测器结构	117	9.2.3 微带天线	139
8.1.3 无线信道的可辨识性	119	9.2.4 平面倒F形天线	140
8.1.4 对测量数据的影响	120	9.2.5 辐射耦合双L形天线	141
8.2 时域测量	121	9.2.6 多波段天线	141
8.2.1 冲激探测器	121	9.2.7 移动台天线的安置	141
8.2.2 相关探测器	122	9.3 基站天线	142
8.3 频域分析	123	9.3.1 天线类型	142
8.4 改进的测量方法	123	9.3.2 阵列天线	142
8.4.1 扫描时延互相关器	123	9.3.3 改进天线方向图	143
8.4.2 逆滤波	124	9.3.4 环境对天线方向图的影响	143
8.4.3 平均	125	深入阅读	145
8.4.4 同步	126		
8.4.5 矢量网络分析仪测量	126		

第三部分 收发信机和信号处理

第10章 无线通信链路结构	149	11.2.3 功率谱	158
10.1 收发信机结构	149	11.2.4 信号空间图	159
10.2 简化模型	153	11.3 几种重要的调制	161
深入阅读	153	11.3.1 二进制相移键控	161
第11章 调制	154	11.3.2 正交相移键控 QPSK	163
11.1 概述	154	11.3.3 $\pi/4$ -DQPSK	165
11.2 基本概念	155	11.3.4 偏移正交相移键控	167
11.2.1 脉冲幅度调制	155	11.3.5 高阶调制	168
11.2.2 多脉冲调制和连续相位调制	157	11.3.6 二进制频移键控	169
		11.3.7 最小频移键控	171

11.3.8	最小频移键控的解调	173	13.3	宏分集和同播	207
11.3.9	高斯最小频移键控	174	13.4	信号的合并	208
11.3.10	脉冲位置调制	175	13.4.1	选择式分集	208
11.3.11	频谱效率小结	177	13.4.2	开关分集	210
11.4	附录	177	13.4.3	合并分集	211
深入阅读		177	13.5	在衰落信道中分集接收的 差错率	215
第12章	解调	178	13.5.1	平坦衰落信道中的 差错率	215
12.1	加性高斯白噪声信道中的 解调器结构和差错率	178	13.5.2	频率选择性衰落信道中的 差错率	217
12.1.1	信道模型和噪声	178	13.6	发射分集	219
12.1.2	信号空间图及最佳 接收机	179	13.6.1	具有信道状态信息的 发射机分集	220
12.1.3	差错率的计算方法	181	13.6.2	无信道状态信息的 发射机分集	220
12.2	平坦衰落信道中的差错率	186	13.7	附录	221
12.2.1	平均误比特率(经典的 计算方法)	186	深入阅读		221
12.2.2	计算平均差错率的 另一种方法	188	第14章	信道编码	222
12.2.3	中断概率与平均错误 概率	191	14.1	编码与信息论基础	222
12.3	时延及频率色散衰落信道中 的差错概率	192	14.1.1	编码历史与动机	222
12.3.1	差错基底的物理原因	192	14.1.2	信息论的基础概念	222
12.3.2	采用群时延法计算 差错基底	196	14.1.3	信道容量	224
12.3.3	一般衰落信道:二次高斯 变量法	197	14.1.4	功率与带宽的关系	225
12.3.4	误比特率	199	14.1.5	与实际系统的关系	226
深入阅读		199	14.1.6	实际编码的分类	226
第13章	分集	200	14.2	分组编码	227
13.1	引言	200	14.2.1	引言	227
13.1.1	分集的原理	200	14.2.2	编码	228
13.1.2	相关系数的定义	201	14.2.3	译码	229
13.2	微分集	201	14.2.4	差错识别和纠正	230
13.2.1	空间分集	202	14.2.5	级联码	231
13.2.2	时分集	204	14.3	卷积码	231
13.2.3	频率分集	205	14.3.1	卷积码原理	231
13.2.4	角度分集	206	14.3.2	维特比译码 (经典表示)	233
13.2.5	极化分集	207	14.3.3	维特比算法的改进	235
			14.4	网格编码调制	236
			14.4.1	基本原理	236
			14.4.2	集分割	238

14.5	比特交织编码调制	240	15.3.1	短时平稳模型	261
14.6	Turbo 码	241	15.3.2	线性预测声码器(LPC) ...	262
14.6.1	引言	241	15.3.3	正弦模型	264
14.6.2	编码器	241	15.3.4	谐波 + 噪声模型	264
14.6.3	Turbo 译码器	242	15.3.5	循环平稳模型	264
14.7	低密度奇偶校验码	243	15.4	量化和编码	265
14.7.1	低密度奇偶校验码的 定义	244	15.4.1	标量量化	265
14.7.2	低密度奇偶校验码的 编码	244	15.4.2	矢量量化	266
14.7.3	低密度奇偶校验码的 译码	245	15.4.3	预编码中的噪声成形 ...	268
14.7.4	性能改进	247	15.4.4	合成分析	270
14.8	衰落信道的编码	248	15.4.5	信源信道联合编码	271
14.8.1	交织	248	15.5	从语音传输到声音的 远程呈现	272
14.8.2	分组码	250	15.5.1	语音激活检测	272
14.8.3	卷积码	250	15.5.2	接收端增强	273
14.8.4	级联码	251	15.5.3	回声和噪声	273
14.8.5	衰落信道下的网格编 码调制	251	15.5.4	远程呈现的业务增强 ...	273
14.9	衰落信道的信息论性能限 ...	252	深入阅读		274
14.9.1	各态历经容量与中断 容量	252	第 16 章	均衡器	275
14.9.2	仅接收端已知信道状 态信息的容量	253	16.1	引言	275
14.9.3	CSIT 和 CSIR 的容量 ——注水法	253	16.1.1	时域和频域均衡	275
14.10	附录	254	16.1.2	信道和均衡器的模型 ...	276
深入阅读		254	16.1.3	信道估计	277
第 15 章	语音编码	255	16.2	线性均衡器	278
15.1	概述	255	16.2.1	迫零均衡器	279
15.1.1	语音电话作为对话多 媒体业务	255	16.2.2	最小均方误差准则	279
15.1.2	信源编码基础	255	16.2.3	均方误差均衡器的 自适应算法	281
15.1.3	语音编码器设计	256	16.2.4	其他线性结构	283
15.2	语音	257	16.3	判决反馈均衡器	283
15.2.1	语音的产生	257	16.3.1	最小均方误差判决 反馈均衡器	284
15.2.2	语言声学	258	16.3.2	迫零判决反馈均衡器 ...	285
15.2.3	语音感知	260	16.4	最大似然序列估计: 维特比 检测器	285
15.3	语音信号的随机模型	261	16.5	均衡器结构的比较	287
			16.6	分数间隔均衡器	288
			16.7	盲均衡	288
			16.7.1	简介	288

16.7.2 恒模算法	289	16.7.5 评价	290
16.7.3 盲最大似然估计	289	16.8 附录	290
16.7.4 使用二阶或高阶统计特性的算法	290	深入阅读	290

第四部分 多址和高级收发信机方案

第 17 章 多址和蜂窝原理	293	18.2.5 同步	320
17.1 引言	293	18.2.6 码族	320
17.2 频分多址	293	18.3 蜂窝码分多址系统	324
17.2.1 频率划分实现多址	293	18.3.1 码分多址原理回顾	324
17.2.2 中继增益	295	18.3.2 功率控制	325
17.3 时分多址	298	18.3.3 提升容量的方法	327
17.4 分组无线电	299	18.3.4 与其他多址方式的结合	328
17.4.1 ALOHA	300	18.4 多用户检测	328
17.4.2 载波侦听多址	301	18.4.1 简介	328
17.4.3 分组预约多址	302	18.4.2 线性多用户检测	329
17.4.4 各方式之间的比较	302	18.4.3 非线性多用户检测器	330
17.4.5 分组无线电的路由问题	303	18.5 跳时脉冲无线电	332
17.5 双工	304	18.5.1 简易脉冲无线电	332
17.6 蜂窝网络原理	305	18.5.2 跳时脉冲无线电	333
17.6.1 重用距离	305	18.5.3 时间色散信道下的脉冲无线电	335
17.6.2 小区形状	305	深入阅读	337
17.6.3 六边形小区的小区规划	306	第 19 章 正交频分复用	338
17.6.4 增加容量的方法	308	19.1 引言	338
17.7 附录	310	19.2 正交频分复用的原理	338
深入阅读	310	19.3 收发信机实现	339
第 18 章 扩展频谱系统	311	19.4 频率选择性信道	340
18.1 跳频多址	311	19.4.1 循环前缀	341
18.1.1 跳频原理	311	19.4.2 频率选择性信道中的性能	342
18.1.2 跳频多址(FHMA)	312	19.4.3 编码正交频分复用	344
18.2 码分多址	313	19.5 信道估计	345
18.2.1 直接序列扩频的基本原理	314	19.5.1 基于导频符号的方法	345
18.2.2 多址	315	19.5.2 基于离散导频的方法	345
18.2.3 数学表示	316	19.5.3 基于特征值分解的算法	346
18.2.4 码分多址情况下的多径传播效应	317	19.6 峰平比	347

19.6.1	峰平比问题的起因	347	20.3.1	性能限制	396
19.6.2	降低峰平比的方法	348	20.3.2	调度	397
19.7	载波间干扰	349	20.3.3	上行链路线性预编码	397
19.8	自适应调制和容量	353	20.3.4	下行链路线性预编码	398
19.8.1	信道质量估计	353	20.3.5	闭环系统和量化反馈	401
19.8.2	参数自适应	353	20.3.6	基站协作	402
19.8.3	选定参数信号发射	355	深入阅读		403
19.9	正交频分多址接入	355	第21章	认知无线电	404
19.10	多载波码分多址	357	21.1	问题描述	404
19.11	采用频域均衡的单载波调制	358	21.2	认知收发信机结构	406
深入阅读		359	21.3	交织原理	407
第20章	多天线系统	360	21.4	频谱感知	408
20.1	智能天线	360	21.4.1	分级系统中的频谱感知	408
20.1.1	什么是智能天线	360	21.4.2	探测器的类型	409
20.1.2	目的	360	21.4.3	多节点检测	410
20.1.3	增大容量	361	21.4.4	认知导频	411
20.1.4	接收机结构	363	21.5	频谱管理	412
20.1.5	天线权重调整算法	366	21.5.1	频谱机会跟踪	412
20.1.6	上行链路与下行链路	371	21.6	频谱共享	412
20.1.7	下行链路中的天线权重自适应算法	372	21.6.1	引言	412
20.1.8	网络方面	374	21.6.2	非合作博弈	412
20.1.9	多用户分集和随机波束成形	374	21.6.3	部分合作博弈	413
20.2	多输入多输出系统	376	21.6.4	集中式方案	414
20.2.1	简介	376	21.7	填充式	415
20.2.2	空分复用如何发挥作用	376	21.8	下垫式分级接入：超宽带系统通信	416
20.2.3	系统模型	377	深入阅读		420
20.2.4	信道状态信息	378	第22章	中继、多跳和协作通信	421
20.2.5	非衰落信道的容量	378	22.1	引言与缘起	421
20.2.6	平坦衰落信道的容量	380	22.1.1	中继原理	421
20.2.7	信道的影响	383	22.2	中继基础	422
20.2.8	分层空时结构	387	22.2.1	基本协议	422
20.2.9	分集	389	22.2.2	译码转发	424
20.2.10	分集、波束成形增益和空分复用的权衡	391	22.2.3	放大转发	426
20.2.11	MIMO 反馈	392	22.2.4	压缩转发	427
20.3	多用户 MIMO	395	22.3	多节点并行中继	427
			22.3.1	中继选择	429
			22.3.2	分布式波束成形	430

22.3.3	正交信道传输	431	22.7	网络编码	455
22.3.4	分布式空时码	432	22.7.1	双向中继	455
22.3.5	编码协作	432	22.7.2	网络编码基础	456
22.3.6	喷泉码	435	22.7.3	网络编码应用于无线 网络	457
22.4	多跳网络中的路由和 资源分配	435	22.7.4	干扰对准	457
22.4.1	数学基础	436	深入阅读		458
22.4.2	路由协议的目标和 分类	437	第23章	视频编码	460
22.4.3	源路由	438	23.1	引言	460
22.4.4	基于链路状态路由	439	23.1.1	数字视频表示和格式	460
22.4.5	距离向量路由	440	23.1.2	视频编码结构	461
22.4.6	基于地理的路由	441	23.2	变换和量化	462
22.4.7	分级路由	442	23.2.1	离散余弦变换	462
22.4.8	节点移动性的影响	442	23.2.2	标量量化	463
22.4.9	数据驱动路由	443	23.3	预测	464
22.4.10	功率分配策略	444	23.3.1	帧内预测	464
22.4.11	多消息路由——随机 网络优化	444	23.3.2	帧间预测	465
22.4.12	比例法则	447	23.4	熵编码	466
22.5	协作网络中的路由和资源 分配	448	23.4.1	哈夫曼编码	466
22.5.1	分离边路由和任意 路径路由	448	23.4.2	算术编码	467
22.5.2	带有能量累积的路由	449	23.5	视频编码标准	468
22.5.3	喷泉码	450	23.6	分层视频编码	469
22.5.4	其他协作路由问题	451	23.6.1	可伸缩视频编码	469
22.5.5	比例法则	452	23.6.2	多视角视频编码 (MVC)	471
22.6	各种应用	453	23.7	差错控制	471
22.6.1	专用中继	454	23.7.1	传输层机制	471
22.6.2	无线自组织网络中的 中继和用户协作	454	23.7.2	视频容错编码	472
			23.7.3	解码端错误掩盖	474
			23.8	视频流	474
			深入阅读		475

第五部分 标准的无线系统

第24章	全球移动通信系统	479	24.2.3	运行支持子系统	482
24.1	历史回顾	479	24.3	空中接口	482
24.2	系统概述	480	24.4	逻辑与物理信道	486
24.2.1	基站子系统	480	24.4.1	逻辑信道	486
24.2.2	网络和交换子系统	481	24.4.2	逻辑信道与物理信道 之间的映射	488

24.5	同步	490	25.5.3	上行链路数据突发随 机化和门控	517
24.5.1	频率同步	490	25.5.4	下行链路扩频和 调制	518
24.5.2	时间同步	491	25.5.5	讨论	520
24.5.3	定时提前	491	25.6	逻辑和物理信道	520
24.5.4	突发结构小结	492	25.6.1	业务信道	520
24.6	编码	492	25.6.2	接入信道	521
24.6.1	语音编码	492	25.6.3	导频信道	521
24.6.2	信道编码	493	25.6.4	同步信道	521
24.6.3	加密	495	25.6.5	寻呼信道	521
24.6.4	跳频	496	25.6.6	功率控制子信道	522
24.7	均衡器	496	25.6.7	将逻辑信道映射为 物理信道	522
24.8	电路交换数据传输	498	25.7	越区切换	522
24.9	建立连接和切换	498	25.8	附录	523
24.9.1	识别号码	498	深入阅读	523	
24.9.2	移动用户识别	499	第26章	WCDMA/UMTS	524
24.9.3	建立连接的例子	500	26.1	历史回顾	524
24.9.4	不同类型切换举例	501	26.2	系统概述	525
24.10	业务和计费	504	26.2.1	物理层概述	525
24.10.1	现有业务	504	26.2.2	网络结构	525
24.10.2	计费	506	26.2.3	层次蜂窝结构	527
24.11	GSM 术语	507	26.2.4	数据速率和服务分类	527
24.12	附录	510	26.3	空中接口	528
深入阅读	510		26.3.1	频带和双工	528
第25章	IS-95 和 CDMA 2000	511	26.3.2	时域双工和频域双工的 模式	528
25.1	历史回顾	511	26.3.3	无线频率相关的方面	528
25.2	系统概述	511	26.4	物理和逻辑信道	530
25.3	空中接口	512	26.4.1	逻辑信道	530
25.3.1	工作频段和双工方式	512	26.4.2	物理信道	531
25.3.2	扩频和调制	512	26.5	语音编码、复用和信道 编码	533
25.3.3	功率控制	512	26.5.1	语音编码	533
25.3.4	导频信号	513	26.5.2	复用和交织	534
25.4	编码	513	26.6	扩谱和调制	537
25.4.1	语音编码器	513	26.6.1	帧结构、扩频码和 沃尔氏-哈达曼 (Walsh- Hadamard) 码	537
25.4.2	纠错编码	514			
25.5	扩频和调制	515			
25.5.1	长、短扩频码和 Walsh 码	515			
25.5.2	上行链路扩频与调制	516			

26.6.2	上行链路	537	27.4.7	物理下行控制信道	568
26.6.3	下行链路	541	27.4.8	物理随机接入信道	570
26.7	物理层过程	543	27.4.9	与物理上行共享信道 相关的控制信号的一 般性质	571
26.7.1	小区搜索和同步	543	27.4.10	物理上行控制信道	571
26.7.2	建立连接	544	27.4.11	物理上行共享信道	572
26.7.3	功率控制	544	27.5	物理层过程	573
26.7.4	切换	545	27.5.1	建立连接	573
26.7.5	过载控制	546	27.5.2	重传和可靠性	574
26.8	WCDMA 术语	547	27.5.3	调度	575
深入阅读		548	27.5.4	功率控制	575
27.5.5	切换	575	27.6	LTE 术语	576
第 27 章	3GPP 长期演进	549	深入阅读		577
27.1	引言	549	第 28 章	WiMAX/IEEE 802.16	578
27.1.1	历史	549	28.1	引言	578
27.1.2	目标	549	28.1.1	历史	578
27.2	系统概述	550	28.1.2	WiMAX 与现有的蜂窝 系统	579
27.2.1	频带和频谱灵活性	550	28.2	系统概述	579
27.2.2	网络结构	552	28.2.1	物理层概述	579
27.2.3	协议结构	552	28.2.2	频带	580
27.2.4	物理层和 MAC 层 概述	553	28.2.3	媒体接入控制层概述	580
27.3	物理层	554	28.2.4	网络结构	581
27.3.1	帧、时隙和符号	554	28.3	调制与编码	582
27.3.2	调制	556	28.3.1	调制	582
27.3.3	映射到物理资源—— 下行链路	557	28.3.2	编码	583
27.3.4	映射到物理资源—— 上行链路	558	28.4	逻辑和物理信道	585
27.3.5	导频或参考信号	558	28.4.1	帧和区域	585
27.3.6	编码	562	28.4.2	帧结构的细节	585
27.3.7	多天技术	563	28.4.3	数据到(逻辑)子信道 的映射	587
27.4	逻辑和物理信道	565	28.4.4	前导和导频的原则	587
27.4.1	数据映射到(逻辑) 子信道	565	28.4.5	PUSC	588
27.4.2	同步信号	566	28.4.6	TUSC	593
27.4.3	广播信道	567	28.4.7	FUSC	593
27.4.4	与下行共享信道有关 的控制信道综述	568	28.4.8	AMC	595
27.4.5	物理控制格式指示 信道	568	28.4.9	信道探测	595
27.4.6	物理 HARQ 指示信道	568	28.5	多天技术	596

28.5.1	空时编码和空间复用	596	29.2.4	同步和信道估计	610
28.5.2	MIMO 预编码	598	29.3	IEEE 802.11n	610
28.5.3	MIMO 中的 SDMA 和 软切换	599	29.3.1	概述	610
28.6	链路控制	599	29.3.2	调制和编码	611
28.6.1	建立连接	599	29.3.3	多天线技术	612
28.6.2	调度和资源请求	600	29.3.4	20 MHz 或 40 MHz 信道	613
28.6.3	服务质量	601	29.3.5	帧头和前导	614
28.6.4	功率控制	601	29.3.6	信道估计	615
28.6.5	切换和移动性支持	601	29.4	802.11 无线局域网中的 分组传输	616
28.6.6	节约功率	602	29.4.1	通用 MAC 结构	616
28.7	WiMAX 术语	603	29.4.2	帧格式	617
深入阅读		603	29.4.3	分组无线电多址	617
第 29 章	无线局域网	604	29.5	无线局域网的其他选择和 未来的发展	619
29.1	引言	604	29.6	WLAN 术语	619
29.1.1	历史	604	深入阅读		620
29.1.2	应用	605	第 30 章	习题	621
29.1.3	媒体接入控制层与 物理层的关系	606	参考文献		655
29.2	802.11a/g——基于正交频分 复用的局域网	606	缩略词表		679
29.2.1	频段	607	符号表		701
29.2.2	调制与编码	608			
29.2.3	头部	609			

第一部分 引言

- | |
|-----------------|
| 第1章 无线业务的应用和需求 |
| 第2章 无线通信的技术挑战 |
| 第3章 噪声受限和干扰受限系统 |

本书第一部分介绍无线通信的基本应用，以及这种通信形式的内在技术问题。第1章在简要介绍了无线通信发展史之后，对不同类型的无线业务进行了描述，并指出了它们的主要差别。1.3节将从另一个不同角度来审视同一问题：实际系统中应采用怎样的数据速率、覆盖范围等等，尤其是需要将这些性能措施进行怎样的搭配（例如，短距离时应采用怎样的数据速率发送，长距离时需要怎样的数据速率）？第2章则描述了无线通信所面临的技术挑战，特别强调了衰落和同道干扰。第3章描述了设计一个无线系统的最基本问题，即在噪声受限或者干扰受限的系统中进行链路预算。

学习了本书这一部分以后，读者将会对不同类型的无线业务有一个概括性的了解，并能理解这些业务中的每一种所涉及的技术挑战。应对这些挑战的措施将在本书后面的各部分中予以描述。

第1章 无线业务的应用和需求

无线通信是近25年间在工程领域里获得巨大成功的传奇之一——不仅从科学的角度讲，其进展有目共睹，而且在市场规模和社会影响方面也是如此。25年前一些默默无闻的公司，由于它们的无线产品，其公司名现在已在世界范围内家喻户晓。并且，在某些国家里，无线产业正支配着整个国民经济。工作习惯甚至人与人之间更为一般的交流方式，都随着可以在“任何地点、任何时间”进行通话的可能性而发生了巨大的变化。

长期以来，人们都将无线通信与蜂窝电话关联起来，因为它具有最大的市场份额，并且对日常生活具有最大的影响力。近来，无线计算机网络也已在人们的工作习惯和用户移动性方面带来了显著的改变，比如在咖啡店里回复电子邮件已成为家常便饭。但是，除了这些被广泛宣传的应用以外，还有大量人们无法清晰地意识到的应用也已开发出来，并且正在开始改变我们的生活。用无线传感器网络监测工厂，以无线链路代替计算机和键盘之间的线缆，用无线定位系统监视卡车（装载有通过无线射频标签来识别的货物）的位置。对无线工程师们来说，这些各式各样的新应用所形成的技术挑战与日俱增，本书旨在对目前及未来所面临技术挑战的应对措施予以全面论述。

开发新的技术解决方案一般有两条途径：工程驱动和市场驱动。在第一条途径下，工程师们提出一个卓越的科学见解，但并未考虑到直接的应用。随着时间的推进，市场需求可以帮助发现由这一见解所指引并使之成为可能的应用^①。在另一条途径下，市场需要某种特定的产品，由工程师们试着开发出满足这种需求的技术解决方案。本章将描述这些市场需求。我们从无线通信简史开始，旨在向人们传递关于过去100年里科学和市场如何发展的感受。然后描述一些构成当前无线市场主体的业务类型。这些业务类型中的每一种都在数据速率、覆盖范围、用户数目、能量消耗和移动性等方面满足着特定的应用需求。1.3节将讨论所有这些方面。这一节末尾将描述无线设备工程与（由其所引发的）社会人群行为变化的互动关系。

1.1 历史

1.1.1 一切是如何开始的

当我们回首通信的发展历史时，会发现无线通信实际上是最古老的通信形式，因为当吼叫声和丛林鼓声发挥传递信息的作用时不需要任何电线或线缆。甚至最古老的“电磁”（光）通信也是无线的，因为通过烟雾传递信息的方式正是基于光信号沿视距（line of sight）线路传播的原理。然而，我们所知的无线通信却肇始于麦克斯韦（Maxwell）和赫兹（Hertz）的努力，他们为我们理解电磁波的传输奠定了基础。在他们开创性的工作之后不久，特斯拉（Tesla）就向人们

^① 随后，第2章给出了对无线通信中主要技术挑战的总结——工程驱动的解决方案的理论基础。第3章至第23章讨论了这些挑战的技术细节和科学基础，而第24章至第29章详细说明了近些年来已开发的一些特定的系统。

演示了借助于电波的信息传输,即实质上的第一个无线通信系统。1898年,马可尼(Marconi)进行了从一条船上到英吉利海峡怀特(Wight)岛的为世人瞩目的无线通信演示。值得注意的是,虽然特斯拉是在这一至关重要的尝试中第一个取得成功的,马可尼却具有更好的公众影响力,从而被广泛地引证为无线通信的发明者,并因此获得了1909年的诺贝尔奖^①。

在随后的年月里,广播的应用(及后来的电视)逐渐遍及全世界。虽然在目前“通行的”意义上我们经常不把广播或电视理解为“无线通信”,但从科学意义上,它们的确应归入此范畴,因为同样是借助于电磁波将信息从一个地方传送到另一个地方。如汽车电台所证实的,甚至可以将它们视为“移动通信”。大量的基础性研究,尤其是关于无线传播信道的研究,最初都是为了娱乐广播的有效传播而开展的。到了20世纪30年代后期,一个无线信息传输的广大网络(尽管是单向的)已然就位。

1.1.2 第一个系统

与此同时,双向移动通信的需求开始显现。在警察局和军队中,这种双向通信有着显而易见的应用前景,并且它们也是最先将无线系统用于封闭的用户群的机构。第二次世界大战期间及其结束后不久,军事应用带动了许多研究。这也是通信的许多理论基础基本奠定的时期。克劳德·香农(Claude Shannon)的开创性著作*A Mathematical Theory of Communications*(通信的数学理论)就出现于这一时期,它明确指出了在数据速率和信噪比受限情况下进行无差错传输的可能性。该著作中的某些提议,比如在频率选择信道中采用最优功率分配,时至今日才刚刚被引入无线系统中。

无线通信在20世纪40年代和20世纪50年代有几个重要进展。公民波段(CB)电台的应用得到普及,从而建立起公路上汽车与汽车之间新的通信方式。采用这些系统进行的通信,对于传送至关重要的交通信息,以及在拥有这些设备的司机们所组成的封闭群体内部传递有关情况,都是非常有用的,但却缺少与公用电话系统的接口,通信距离也被限制在大约100 km,这取决于(移动)发射机的发射功率。1946年,第一个移动电话系统在美国圣路易斯投入使用。这一系统具有与公共电话交换网络(PSTN)这一陆上有线电话系统之间的接口,尽管这一接口不是自动的,还需要人工电话接线员。然而,由于整座城市总共只有6个语音信道,这个系统很快就遭遇到用户容量的限制。这就激发了人们去研究即使所分配的频谱受限,也能提升可同时服务的用户数目的方法。美国电话电报公司(AT&T)公司贝尔实验室的研究人员找到了答案:蜂窝原理,将地理区域划分为许多小区,不同小区内可以使用相同频率。时至今日,这一原理构成了绝大多数无线通信的基础。

尽管理论上取得了突破,蜂窝电话在20世纪60年代并没有经历显著的增长。然而,在另一条不同的阵线上,人们取得了鼓舞人心的进展:1957年,苏联发射了第一颗人造卫星(Sputnik),美国紧随其后。这一进展预示着卫星通信这一新的研究领域的出现^②。许多基本问题亟待解决,包括大气传播效应、太阳风暴的影响、为卫星设计太阳能板及其他长效电源等问题。如今,卫星通信已成为无线通信的一个重要领域(尽管不是本书要专门讨论的问题)。其最为普及的应用是卫星电视的传送。

① 实际上,马可尼的专利在20世纪40年代被撤消了。

② 卫星通信,特别是通过静止轨道卫星进行的通信,早在20世纪40年代就已经被科幻小说家科拉克(Arthur C. Clark)提到过。

1.1.3 模拟蜂窝系统

20 世纪 70 年代,人们重新拾起对蜂窝通信的兴趣。在科学研究方面,这些年中,针对路径损耗、多普勒谱、衰落统计量,以及决定模拟电话系统性能的其他参量建立了模型,并实现了公式化。这方面工作的集大成者是杰克斯(Jakes)所著的 *Microwave Mobile Radio* 一书[Jakes 1974],它总结了当时这一领域内的技术发展水平。20 世纪 60 年代和 20 世纪 70 年代也出现了许多最初专门为陆上有线通信进行的基础研究,后来亦被证明同样有助于无线通信。例如,自适应均衡器的基本理论以及当时逐渐形成的多载波通信的概念。

就无线电话走向实用而言,设备小型化方面的进展使人们对“便携式”设备的想象日趋现实。像摩托罗拉(Motorola)和 AT&T 这样的公司为争夺该领域的领先地位而相互竞争,并且为促进其发展作出了重大贡献。日本电话和电报(NTT)公司 1979 年在东京建成了商用蜂窝电话系统。然而,却是由一家瑞典公司建立起了第一个具有大的覆盖范围和自动交换功能的系统:直到这时,爱立信 AB(Ericsson AB)公司主要以其电话交换设备为人们所熟悉,人们原以为该公司对无线通信不感兴趣。可是,正是在交换技术方面的专业经验和采用数字交换技术的决定(在那时是比较大胆的),使其能将一个大区域内的不同小区整合到单个网络中,并建立起北欧移动电话(NMT)系统[Meurling and Jeans 1994]。要注意的是,虽然交换技术是数字的,但射频传输技术仍然是模拟的,因而仍称为模拟系统。后来,其他一些国家开发出了自己的模拟电话标准,例如美国的系统称为高级移动电话系统(AMPS)。

NMT 系统的调查还确立了一种估算市场规模的有趣方法:在瑞典,商业顾问们将可能的移动电话用户数目和梅赛德斯 600(Mercedes 600,那时的顶级豪华轿车)的数目等同起来。移动电话永远不可能成为一个规模巨大的市场,这似乎是显而易见的。同样的想法也出现在蜂窝电话的发明者(AT&T)的管理层。根据一家咨询公司的建议,他们认定移动电话永远不可能吸引数量巨大的使用者,并停止了蜂窝通信的商业活动^①。

模拟系统为无线革命铺平了道路。20 世纪 80 年代期间,它们以激动人心的步伐得到了发展,在欧洲,其市场占有率已高达 10%,尽管在美国其影响力要稍差一些。在 20 世纪 80 年代初,电话是“便携式的”,而绝非手持的。在大多数语言里,它们仅仅被称为“车载电话(carphones)”,因为电池和发射机装载于汽车的行李箱里,由于过于沉重而不能随身携带。但到了 20 世纪 80 年代末,具有良好的通话质量和令人十分满意的电池使用时间的手机已经比比皆是。模拟手机质量之优异以至于数字电话在一些市场领域难以立足,似乎完全没有进一步发展的必要性。

1.1.4 GSM 及世界范围的蜂窝革命

尽管公众并未意识到从模拟向数字转变的必要性,网络的运营者却更清楚这一点。模拟电话具有较差的频谱效率(第 3 章将讨论其原因),并且由于蜂窝市场的快速增长,网络运营者对于为更多用户提供可用频谱有很高的兴致。此外,通信领域的研究已经开始不可逆转地转向了数字通信,而且还包括数字无线通信。20 世纪 70 年代后期和整个 20 世纪 80 年代,在遍及全世界的许多研究实验室中,人们都在进行着这方面的探索,研究内容涉及频谱的高效调

^① 当人们逐渐意识到原先的决定的荒唐之处时,这些商业活动在 20 世纪 90 年代初期得到重新启动。AT&T 随即花了超过 100 亿美元收购麦克考(McCaw)公司,并将它重新命名为 AT&T Wireless。

制方式、信道失真与时变性对数字信号的影响、多址方案及更多其他问题。因此,对于该领域的那些行家而言,实际系统随着研究的深入将很快变为现实已经是非常明了的事情了。

欧洲再一次走在了前头。欧洲电信标准协会(ETSI)的下设小组开始研究制定全欧洲强制性的数字蜂窝标准,并且该标准后来被世界上绝大多数国家所采用,这就是全球移动通信系统(GSM)。该系统的开发贯穿于整个20世纪80年代,于20世纪90年代初期开始实际的部署,并且立即为用户所接受。由于其附加特性、更好的语音质量及安全通信的潜力,基于GSM业务的市场份额在其引入市场的两年内就超越了模拟业务。在美国,向数字系统的转变稍微慢了一些,但截至20世纪90年代末,这个国家的通信业务也基本上完全数字化了。

数字电话使得已经迈上成功之路的蜂窝通信变成了一颗“重磅炸弹”。2004年,蜂窝数字电话在西欧的市场占有率已超过80%,而在某些斯堪迪纳维亚(Scandinavian)国家,其市场占有率则接近了100%(许多人拥有两部或三部手机)。在美国其占有率也超过了50%,在日本则达到约70%。以绝对数目而论,中国已成为最大的独立市场,在2004年就拥有了约3亿数字电话用户。

无线系统的发展也使人们清楚地意识到了标准化的必要性。只有相互兼容,并且每一部接收机能够“理解”每一部发射机,即它们遵从同样的标准,设备之间才能交流。但应该如何来确立这些标准呢?不同的国家提出了不同的处置方式。美国采用的是“不插手(hands-off)”的处置方式:准许各种不同标准投入使用,由市场来决定谁是最后的赢家(或几个赢家)。20世纪90年代,当用于数字蜂窝通信的频率被拍卖时,频谱牌照的购买者可以选择他们想采用的系统标准。因此,如今在美国有3个不同的标准同时被采用。日本采用了类似的方式,有两种不同的系统为争夺其第二代(2G)蜂窝系统的市场而相互竞争。在日本和美国,同一地理区域上可以运作基于不同标准的多个网络,这就允许用户在不同技术标准之间进行选择。

欧洲的情况就不同了。当数字通信引入时,每个国家通常只有一家运营商(一般就是当时的公共电话运营商)存在。如果这些运营商各自采用不同的标准,就会导致严重的市场分裂(即每个标准只拥有一个小市场),这对运营商之间的相互竞争没有好处。此外,由于欧洲的地理区域划分比美国和日本频繁得多,因此国家之间的漫游是经常性的,若各自采用不同的标准,将不可能实现漫游。因而,为整个欧洲设定单一的共同标准是理所当然的。这一决定总的来说被证明对无线通信的发展有利,因为它扩大了经济规模并且降低了新业务的成本,从而提高了新业务的可接受程度。

1.1.5 新无线系统及电信泡沫的爆裂

尽管蜂窝通信在一般人群中为无线通信勾画出了美好的发展图景,但20世纪90年代还是有各种不同层次的新业务陆续进入市场。无绳电话开始取代许多家庭里的“常规”电话。这些无绳电话的最初版本采用的都是模拟技术;然而,对于这一应用,数字技术同样被证明是更为优越的。在其他方面,模拟通话被窃听的可能性、邻近的非法使用者“劫持”了模拟无绳基站并进行非法通话却由他人付费的可能性,导致必须转向数字通信。然而无绳电话从来就未能取得蜂窝电话那样壮观的市场规模,它们只占有一个固定的市场。

20世纪90年代充满希望的另一个市场是固定无线接入和无线本地环路(WLL),也就是将通向用户住所的铜线替换为无线链路,但它并没有移动性所具有的特殊好处。这期间提出了许多技术方案,但最终无一例外地都失败了。究其原因,经济、政治方面的因素不比技术因

素少。无线本地环路的原始动机是使另外的电话业务提供商可以争取到用户,同时又绕开了属于当前业务提供商的用于接入的铜线。然而,20世纪90年代中期,全世界所有的电信管理者都规定现有的有线业务提供商必须向其他的业务提供商租借其线路,通常出租价格还非常优惠。这就消除了无线本地环路大部分的经济基础。同样地,固定无线接入被作为以具有竞争力的价格来提供宽带数据接入的方案加以推广。然而,与数字用户线(DSL)技术和有线电视之间的价格战已经大大减弱了这种接入方式的经济吸引力。

因此,最大的宝藏似乎在于创建“第三代(3G)”系统(在模拟系统和诸如 GSM 之类的 2G 系统之后)[Bi et al. 2002],使蜂窝系统得以进一步的发展。2G 系统基本上是纯话音传输系统,尽管也包括一些简单的数据业务,如短消息(SMS)。新的系统将在时速达到 500 km/h 的情况下,提供可与运气不佳的综合业务数字网(ISDN)相比的数据速率(144 kbps),甚至高达 2 Mbps。经过漫长的商讨以后,订立了两个标准:3GPP(第三代伙伴计划,受到欧洲、日本和一些美国公司支持)和 3GPP2(受到另一部分美国公司支持)。在世界绝大多数地区,新的标准还需要新的频谱划分,而且对这些频谱使用权的销售已成为一些国家的国库收入的一个重要来源。

3GPP 的发展及早前美国对 IS-95 CDMA(码分多址)的引入,触发了关于码分多址和用于无线通信的其他扩谱技术(见第 18 章)的大量研究;到 20 世纪 90 年代末,多载波技术(见第 19 章)也已在研究界取得了强势的地位。多用户检测,即利用多址干扰的结构信息,能够大大减轻其影响,特别是在 20 世纪 90 年代初期,多用户检测就已经是许多研究者关注的另一个领域了。最后,多天线系统(见第 20 章)领域从 1995 年起有了迅猛发展,有时该领域的研究论文数几乎占无线通信物理层设计研究领域发表论文总数的一半之多。

3G 蜂窝系统频谱的销售及一些以无线业务为主导的公司的首次公开募股(IPO),昭示着 20 世纪 90 年代的“电信泡沫”已经登峰造极了。2000 年和 2001 年,市场在狂跌之下终于崩溃了。随着其倡导者的破产,许多新的无线系统(如固定无线)的发展停滞了,同时另一些系统(包括 3G 蜂窝系统)的部署也大大放缓。最令人不安的是,许多公司减缓甚至完全停止了研究,整个经济的萎靡不振也导致学术研究的资金匮乏。

1.1.6 无线通信的复兴

从 2003 年起,几个方面的发展使得人们重拾起对无线通信的兴趣。第一个就是 2G 和 2.5G 蜂窝通信在新兴市场和新应用的刺激下所取得的持续性增长。只需举出一例:2008 年,甚至在其第一个 3G 网络正式投入运营之前,中国已拥有了超过 5 亿的手机用户。世界范围内,2008 年,在用手机大约有 35 亿部,其中绝大多数是基于 2G 和 2.5G 标准的。

此外,尤其在日本、欧洲和美国,3G 网络已实现了广泛的覆盖并广受欢迎(2008 年,在西欧,总的手机市场渗透率超过了 100%;在美国,这个数字则接近了 90%)。与有线方式具有可比性的数据传输速率(5 Mbps)得以实现。继而,这一进展带动了相应设备的层出不穷,这些设备不仅支持语音通话,也能够支持互联网浏览和流媒体形式的音频和视频接收。这类设备之一的 iPhone,首次推出时即获得了公众的巨大关注,而实际上还有很多手机都具有相似的性能。在美国,这些被称为“智能手机”的设备占有了 20% 的手机市场。所有这些进展导致了针对手机的数据传输业已成为一个规模庞大的市场。

甚至是在 3G 网络刚刚开始部署时,就已经开始了下一代(有时称为 4G 或 3.9G)系统的开

发。大多数基础设施制造商正集中精力于主流 3G 标准的长期演进 (LTE) 方面的研发。一种植根于固定无线接入系统的替代性标准也已开始部署。另外,用手机收看电视节目(无论是直播电视还是预录制的剧集)正在变得越来越普及。对 4G 网络和电视传输而言,多输入多输出的正交频分复用(MIMO-OFDM)(见第 19 章和第 20 章)成为选定的调制方式,已推动了这一领域的研究。

第二个重要进展是无线计算机网络,即无线局域网(WLAN)取得了意想不到的成功。符合 IEEE 802.11 标准(见第 29 章)的设备已使计算机几乎可以像手机那样灵活而机动地使用。这一标准化进程早在 20 世纪 90 年代中期就已经启动了,却形成了几个不同的版本,而制造商之间的深度竞争最终影响了如今的大规模生产和应用。目前,无线接入点比比皆是,不仅在家庭和办公室,在机场、咖啡店和其他类似场所也都可以进行无线接入。因此,依靠便携式电脑和互联网接入办公的许多人士在选择何时何地进行工作时拥有了更大的自由度。

第三方面的进展在于,无线传感器网络为从远端站点对工厂甚至住所的监测与控制提供了新的可能性,并且在军事和安全监控领域也可以得到应用。对传感器网络的兴趣激发了对 ad hoc 和 P2P(peer to peer)网络的研究热潮。这类网络无须专门的基础结构。如果信源和信宿之间的距离过于遥远,则网络中的其他节点将为信息最终到达信宿提供传递信息的帮助。由于它们的网络结构与传统的蜂窝网络有很大的不同,需要进行大量的新的研究。

总的来说,“无线通信的复兴”基于以下 3 种趋势:(i)产品类型的宽泛化,(ii)已有产品数据传输的高速率化,(iii)用户密度的高密集化。这些趋势决定了该领域的研究走向,也为许多晚近的科学开发提供了原动力。

1.2 业务类型

1.2.1 广播

第一种无线业务是广播电台。在此项业务中,信息传送给不同的、也可能是移动的用户,见图 1.1。将广播电台与诸如蜂窝电话这样的业务区分开的属性有如下 4 个。

1. 信息仅在一个方向上传送。只有广播电台(或电视台)将信息发送到广播(或电视)的接收者;听众(或观众)不向电台发送任何信息。
2. 向所有用户传送的信息都是一样的。
3. 信息是连续发送的。
4. 许多情况下,多部发射机发送相同的信息。在欧洲尤其如此,在那里由国家广播网络完成整个国家的覆盖,并在国家的各个角落播出同样的节目^①。

上述属性极大地简化了广播电台网络的设计。发射机的设计不必对接收机有更多的概念或进行更多的考虑。也不需要提供双工信道(不需要将信息从用户接收机传送到电台发射机)。该业务的可能用户数目也对发射机的结构不构成影响,即不必操心是有上百万用户还是只有一个用户,发射机都发送出同样的信息。

以上描述对于传统的模拟广播电视和电台而言基本上是正确的。卫星电视和广播就有所

^① 美国的情形略有不同,在那里通常是“地方台”只完成单个大城市地区的覆盖,而且常常只有一部发射机。

不同了,因为其发射常常只面向所有可能用户的一小部分(付费电视或者依观看次数计费的用户),因此需要对播放内容加密,以阻止未经授权的观看。然而,要注意的是此处的“保密性”问题不同于常见的蜂窝电话。对于付费电视,授权用户群中的所有成员都应能方便地获取相关内容(多播, multicast),而对于蜂窝电话,每个呼叫仅能由呼叫本身所针对的那个用户听到(单播, unicast),而不让其网络提供商的所有用户都听到。

尽管它们在经济上具有毋庸置疑的重要性,但广播网络不是本书的关注重点,篇幅的限制影响了本书对其进行更详尽的讨论。虽然如此,记住它们是无线信息传送的一种特殊情况仍然是有意义的,并且诸如同播(simulcast)数字电视和交互式电视,尤其是诸如供计算机和手机接收的流媒体电视之类的一些新进展,趋向于使广播和蜂窝电话之间的差别进一步模糊化。

1.2.2 寻呼

与广播相似,寻呼系统也是单向无线通信系统。这些系统具有以下特点(见图1.2)。

1. 用户只能接收信息,不能发送信息。所以,一个“寻呼”(信息)只能由寻呼中心发起,而不能由用户发起。
2. 信息的发送目标和接收方都只限于一个单独的用户。
3. 发送信息的数量非常小。刚开始时,接收信息由单比特信息构成,向用户指示“有人发送了一条信息给你”。然后用户必须打一个电话(常常是公用电话)到寻呼中心,在那里由一个寻呼服务人员重复之前保留的信息内容。后来,寻呼系统变得更加复杂,允许传送短消息。例如,给出需要用户回复的另一个电话号码或者一个紧急事件的即时状况,但能提供的信息量很有限。

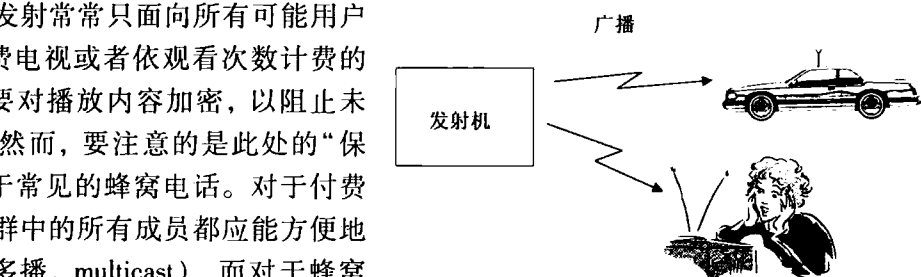


图 1.1 广播发送原理

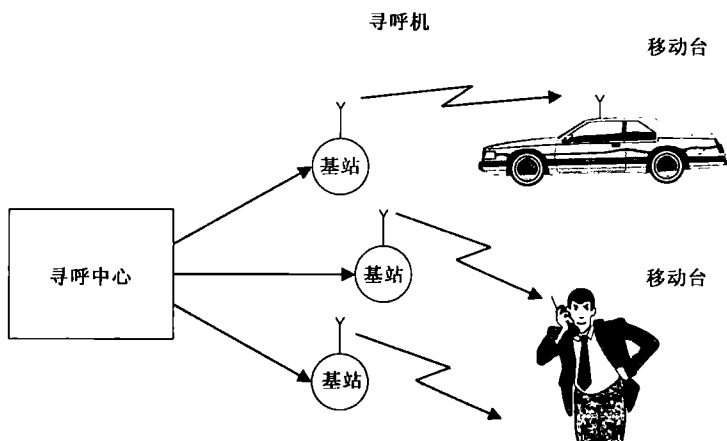


图 1.2 寻呼机原理

由于通信的单向性及信息量较少,此业务所需带宽较小。这就允许该项业务工作工作在较低

载频上(如 150 MHz),这一频段只有少量频谱可供使用。稍后将会了解到,这样的较低载频更容易做到,只用很少的发射机来完成一个大区域的良好覆盖。

寻呼机在 20 世纪 80 年代至 20 世纪 90 年代早期非常流行。对于某些专业群体(如医生),寻呼机是该群体的必备工具,使他们可以在较短的时间内响应紧急事件。然而,蜂窝电话技术的成功大大降低了寻呼的吸引力。蜂窝电话可以提供寻呼机的所有业务,此外还增加了许多其他特性。2000 年后寻呼系统的主要吸引力就在于其可达到的更好的区域覆盖。

1.2.3 蜂窝电话

蜂窝电话是经济上最重要的无线通信形式,具有以下特征。

1. 信息流是双向的。用户可以同时发送和接收信息。
2. 用户可以位于(全国性的或国际性的)网络的任何一个地方。他(她)或者通话双方都不需要知道用户的位置,网络必须考虑到用户的移动性。
3. 一次呼叫可以由网络发起,也可以由用户发起。换句话说,蜂窝用户可以被呼,也可以主呼。
4. 一个通话仅针对某单个用户,其他网络用户不应该听到这一通话内容。
5. 用户的高度移动性。在一次通话期间用户位置可能会显著变化。

图 1.3 显示了蜂窝系统的框图。移动用户同与其建立了良好无线连接的基站通信。各个基站都连接到移动交换中心,移动交换中心与公共有线电话系统相连。

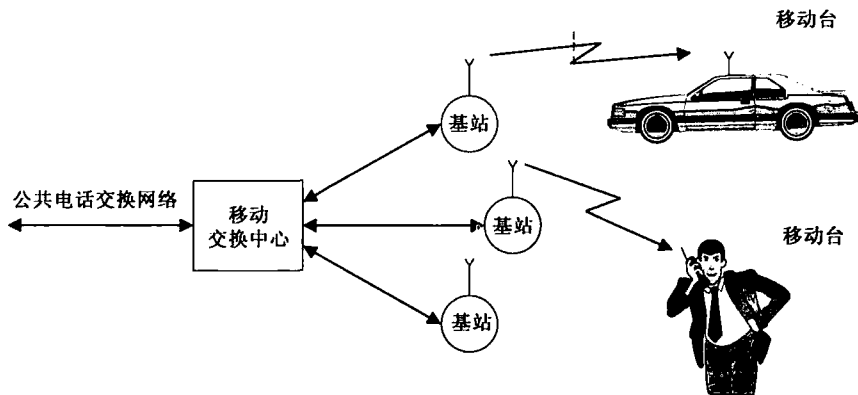


图 1.3 蜂窝系统原理

由于每个用户都想要发送或接收不同的信息,一个蜂窝网络中的有效用户数是受限的。可用带宽必须在不同用户之间共享,这可以通过不同的“多址”方案加以实现(见第 17 章至第 20 章)。这是与广播系统的一个重要差别。在广播系统中,用户(接收机)的数目是没有限制的,因为所有的接收机接收同样的信息。

为增加可能用户的数目,采用了蜂窝原理:一家网络提供商的服务区域被分成许多子区域,称为小区(cell)。在每个小区内,不同用户必须分享可用带宽,以下考虑每个用户占用不同载频的情况。即使是相邻小区也必须使用不同的频率,以使同道干扰足够低。但对于相互之间距离足够远的小区,可以使用相同的频率,这是因为各自发射机发出的信号随着传输距离的增加就会变得更弱。因此,在一个国家里,可能有成百上千的小区使用着相同的频率。

蜂窝电话技术的另一个重要方面是不限制用户的移动性。为了能够实现通信,用户可以位

于网络覆盖区域的任何位置(即,并不限制用户在特定小区内)。而且,他(她)也可以在通话期间从一个小区移动到另一个小区。蜂窝网络具有与公共电话交换网络及其他无线系统的接口。

正如无线通信简史回顾中所指出的,蜂窝电话在20世纪80年代开始流行起来,现已成为在全世界拥有超过10亿用户的最具优势的通信形式。正是由于这个原因,本书将经常从蜂窝电话技术里提炼示例,尽管所阐明的一般原理同样适用于其他无线系统。第24章至第28章将对最流行的蜂窝系统进行详尽的描述。

1.2.4 集群无线电

集群无线电系统是蜂窝电话的一种重要变形,这种无线系统和公共电话交换网络之间没有连接,因此它支持封闭用户群之间的通信。显而易见的应用包括警察局、消防局、出租车和其他类似业务。封闭用户群使一般蜂窝系统不可能(或者很难)实现的如下几项技术革新成为可能。

1. 群呼。可同时与多个用户通信,或者可以实现系统中多个用户之间的会议通话。
2. 呼叫优先级。一般蜂窝系统运作的基本程序是“先到先服务”。一旦建立起一个呼叫,就不能中断该呼叫^①。在蜂窝系统下这是合情合理的,因为网络运营者不能确定呼叫的重要性或紧迫程度。然而,对于集群无线电系统,如用于消防部门的系统,这一程序是不可行的。有关紧急事件的通知必须传达到受其影响的群体中,尽管这样做意味着中断正在进行中的优先级较低的呼叫。因而,集群无线电系统必须能决定呼叫的优先级,并允许为支持高优先级的呼叫而挂断低优先级的呼叫。
3. 中继网络。网络的覆盖范围可以通过将每个移动台都作为其他移动台的中继站来获得延伸。因此,在基站覆盖范围以外的移动台可以向基站覆盖范围内的另一移动台发送信息,由后者将消息传递给基站;为了最终将消息传递到基站,系统甚至可以采用多次中继的方式。这种方式增加了网络的有效覆盖区域和可靠性。然而,此方式却只能用于集群无线电系统,而不能用于蜂窝系统,因为一般的蜂窝用户都不愿接受为其他用户中继消息所必然造成的对其电池能量的消耗。

1.2.5 无绳电话

无绳电话技术所描述的无线链路是手持机和直接与公共有线电话系统相连的基站之间的链路。与蜂窝电话的主要差别是,与无绳电话相关联并且可以与之通信的仅仅是一个孤立的基站(见图1.4)。因此,不存在移动交换中心;反之,基站直接与公共电话交换网络相连。这样就出现了以下几种重要结果。

1. 基站不需要具备任何网络功能。当出现来自于公共电话交换网络的呼叫时,没必要查找移动台的位置,同样也没必要提供不同基站之间的切换。
2. 无中心系统。用户一般在其居所或办公地点有一个基站,但该基站不会影响任何其他基站。因为这个原因,没必要(也不可能)进行频率规划。
3. 无绳电话由用户控制的事实也意味着不同的计费模式。不存在可以为移动台和基站的

^① 除了因技术问题导致的通话中断,比如用户移动到了覆盖区域之外。

连接计费的网络运营商；反之，唯一出现的费用是由基站接入公共电话交换网络的费用。

在其他许多方面，无绳电话与蜂窝电话是相似的。无绳电话支持呼叫区内的移动性；信息流是双向的；呼叫既可以由公共电话交换网络发起，也可以由移动用户发起，并且还必须具有使呼叫不被非法用户截获或监听，以及禁止非法呼叫的措施。

无绳系统还演变为专用自动小交换机(PABX)(见图1.5)。在其最简形式下，一个专用自动小交换机配备有一个基站，可以同时服务于几部手持机，既可以将它们连接到公共电话交换网络，又可以将它们(同一公司或房间内的呼叫)相互连接。在其更高级的形式下，专用小交换机包含有连接至一个中心控制站的多个基站。这种系统基本上具备了与蜂窝系统一样的功能，仅能根据覆盖范围的大小来区分这种全功能专用小交换机和蜂窝网络。

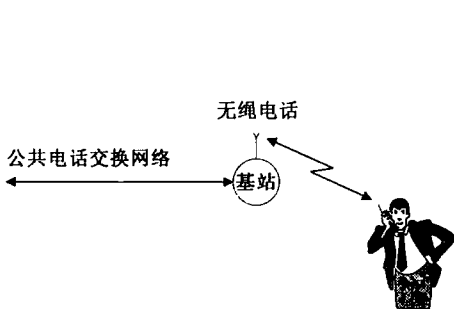


图 1.4 简单无绳电话原理

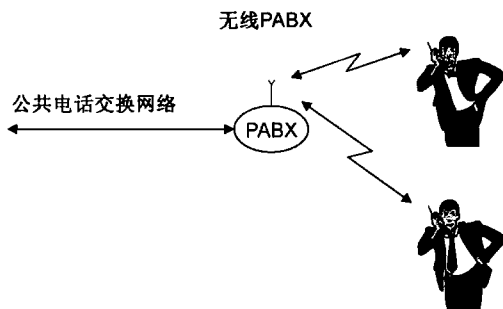


图 1.5 无线专用自动小交换机(PABX)原理

第一个无绳电话系统是仅可建立手持机和基站之间的简单无线链路的模拟系统。通常，它们甚至不能提供起码的安全性(即阻止非法呼叫)。目前的系统是数字的，并可提供更复杂的功能。在欧洲，增强型数字无绳电话(DECT)系统(见配套网站 www.wiley.com/go/molisch)是最重要的标准；日本有一个类似的系统，称为个人手持电话系统(PHS)，它提供了无绳电话及蜂窝系统的替代系统(全功能的专用自动小交换机系统覆盖了日本的绝大部分地区，并可以提供公共接入)的可能性。这两种系统都工作在 1800 MHz 频段，使用的是专门为无绳应用预留的频谱。在美国，数字无绳电话主要工作在与其它无线业务共享的 2.45 GHz ISM(工业、科学和医疗)频段。

1.2.6 无线局域网

无线局域网(WLAN)的功能属性与无绳电话非常相似，将单个移动用户设备连接到公共陆上有线系统(public landline system)。这种情况下，“移动用户设备”通常是便携式电脑，而公共陆上有线系统指的是 Internet。无绳电话系统最主要的好处是其支持用户的移动性，从而方便了用户。无线局域网甚至对将固定位置的电脑(台式机)连接到 Internet 也是有用的，因为可以节省铺设电缆到计算机所在位置的费用。

无线局域网和无绳电话之间最重要的差别是所需数据速率的不同。无绳电话传送(数字)语音，最多需要 64 kbps，而无线局域网至少要与其所连接的 Internet 速度一样快。对于个体用户(家庭)应用，这意味着速率要在 700 kbps(美国的 DSL 速率)至 3~5 Mbps(美国和欧洲的电视电缆接入提供商支持的速率)，以及 20 Mbps 或更高(日本的 DSL 速率)之间。对于具有更高速互联网连接的公司，速率要求相应地也更高。现已开发出许多标准以满足这些高数据速

率的需求,并且所有这些标准都带有 IEEE 802.11 标识。最原始的 IEEE 802.11 标准支持以 1 Mbps 的速率传输;最流行的 802.11b 标准(即人们熟知的 WiFi)支持高达 11 Mbps 的速率,802.11a 标准将速率提升到 55 Mbps。2008 年/2009 年推出的 802.11n 标准已经能够实现更高的传输速率。

在原理上,无线局域网设备可以连接到任何采用相同标准的基站(接入点)。然而,接入点的拥有者可以对接入进行限制(如通过适当的安全设置)。

1.2.7 个域网

当网络覆盖区域变得比无线局域网更小时称之为个域网(Personal Area Network, PAN)。此类网络多数是为了达到简单的“电缆替代”效果。例如,符合蓝牙(Bluetooth)标准的设备支持手机和头戴式免持耳机之间的无电缆的无线连接,这种情况下的两设备之间距离小于 1 m。在这类应用中,数据速率相当低(小于 1 Mbps)。近来,娱乐系统的组件之间(DVD 播放器到电视)、计算机与其外设(打印机,鼠标)之间,以及其他类似的应用受到了重视,并且许多个域网标准已由 IEEE 802.15 小组开发完成。这些应用采用了超过 100 Mbps 的数据速率。

用于更短距离的网络称为体域网(Body Area Network, BAN),这种网络使置于用户身体不同部位的设备之间的通信成为可能。这类体域网在患者健康和医疗设备(如心脏起搏器)的监测领域扮演着日渐重要的角色。

最后要指出,个域网和体域网既可以采用类似于蜂窝方式的网络结构,也可以形成 1.2.9 节将讨论的 ad hoc 网络。

1.2.8 固定无线接入

固定无线接入也可以看成无绳电话或无线局域网的演化形式,它从根本上替代了用户和陆上有线系统之间的专用线缆连接。与无绳系统的主要差别在于:(i)用户设备不具移动性;(ii)基站几乎总是服务于多个用户。此外,通过固定无线接入设备中继之后的传输距离(在 100 米到几十千米之间)要比通过无绳电话中继后的传输距离长得多。

固定无线接入的目的在于:不必从中心交换局到用户所在的办公室或居所铺设电缆,就能向用户提供电话和数据连接。考虑到铺设电缆施工所需的高昂劳动力成本,这种接入方式应该是一种较经济的方式。然而,值得关注的是,绝大多数建筑(尤其是在发达国家的市区)都已经存在现成的某种形式的线缆,普通电话线、电视电缆甚至光纤。不同的国家电信主管部门的管理条例均强调拥有这些线缆的运营商必须允许与之竞争的公司使用其线缆。结果,固定无线接入的主要市场就在于向农村地区提供覆盖,以及在不具备任何有线基础设施的发展中国家建立连接。总之,固定无线接入的商业运行状况已经很让人失望(见 1.1.5 节)。IEEE 802.16(WiMAX)标准试图通过在系统中引入某些有限的移动性来缓解这一问题,因而也就模糊了与蜂窝电话技术之间的界限。

1.2.9 ad hoc 网络和传感器网络

到目前为止,我们已讨论过“基于基础设施”的无线网络,在这种网络中,特定的设施(基站、电视发射机等)被有计划地刻意设置在固定的地点,以进行网络控制并实现与其他网络之间的接口。网络的尺寸可能各不相同(从仅可覆盖一间公寓的局域网到能够覆盖整个国家的

蜂窝网络),但区分“基础设施”和“用户设备”是它们共同的核心原则。然而,存在着另一种网络,其中只有一类设备并且所有的设备都可以移动,它们根据位置和需要自行组织起来形成网络。这类网络称为 ad hoc 网络(见图 1.6)。在一个 ad hoc 网络中仍然可以存在“控制器”,但选择哪个设备作为主终端和选择哪个设备作为从终端,都是在网络形成时相机而定的。ad hoc 网络也不分等级。实际的数据传输(即物理层通信)几乎与基于基础设施的网络完全相同,但在媒体接入和网络功能上都是非常不同的。

ad hoc 网络的好处在于低成本(因为不需要基础设施)和高度灵活性。缺点则包括效率较低、通信范围较小,以及可以囊括进网络的设备数目受限。ad hoc 网络在目前传感器网络的迅速发展中发挥着重要作用,后者用于支持楼宇控制(基于传感数据对空调、灯光等进行控制)、工厂自动化、安全监控等领域设备之间的通信。ad hoc 网络在应急通信(当基础设施被毁坏时,比如遭地震毁坏)和军事通信领域也发挥着相应的作用。

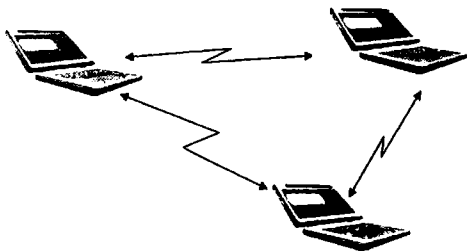


图 1.6 ad hoc 网络原理

1.2.10 卫星蜂窝通信

除了在卫星市场创造最大收益的电视以外,蜂窝通信是卫星的第二个重要应用。卫星蜂窝通信与陆基蜂窝通信虽然有着同样的工作原理,但却存在着重要的差别。

“基站”(即卫星)距移动台的距离相当远。对于静止轨道卫星,距离为 36 000 km;对于近地轨道(LEO)卫星,距离为几百千米。因此,需要更大的发射功率,卫星上必须使用高增益天线(许多情况下移动台也用这种天线),并且若用户在建筑物内部,则通信几乎无法进行。

另一个与陆基蜂窝系统的重要差别在于蜂窝尺寸。由于卫星与地球之间的遥远距离,即使对近地轨道卫星而言,蜂窝直径小于 100 km 也是不可能的;对静止轨道卫星,蜂窝区域会更大。这样大的蜂窝尺寸对卫星系统而言,既是最大的优点也是最大的缺点。积极的一面在于,即使针对地广人稀的区域,也容易做到良好的覆盖,单个小区就可以覆盖撒哈拉地区的绝大部分。另一方面,这个区域内的频谱效率非常低,也就意味着(假定分配给这项业务的频谱有限)只有很少的人可以同时通信。

建立一个“基站”(即卫星)的费用比为陆基系统建造基站的费用高得多。不仅通信卫星的发射非常昂贵,而且还必须作为卫星和公共电话交换网络之间链接点的地面站建造适用的基础设施。

以上所有问题所产生的后果就是,卫星通信系统的商业状况迥异于其他系统,该系统定位于以昂贵的价格向少量用户提供必不可少的通信手段。灾区 and 战区的救护工人与记者、船基通信,以及远离海岸的石油钻井平台上的工人,都是这种系统的典型用户。国际海事卫星(INMARSAT)系统是这种通信的最主要的提供者。20 世纪 90 年代后期,铱(IRIDIUM)计划试图通过 60 多颗近地轨道卫星向人们提供更低价格的卫星通信服务,但最后以破产告终。

1.3 业务需求

理解无线设计的关键在于,要明确不同应用在数据速率、覆盖范围、移动性、能量消耗等方面有不同的需求。设计用户以 500 km/h 速度移动时还能在 100 km 范围内支持 Gbps 数据速

率的系统是完全没有必要的。之所以强调这一事实是因为在工程师中有一种不良趋向,即意图设计“可以做除洗碗之外的任何事”的系统;而诉诸于科学的观点,这样的系统势必是价格昂贵且频谱效率低下的。下面将罗列系统设计中将遇到的种种需求的变动范围,还将列举哪些需求会出现在哪些不同的应用之中。

1.3.1 数据速率

无线业务的数据速率取决于具体的应用,覆盖了从几 bps 到几 Gbps 的广大范围。

- 传感器网络通常要求的数据速率从几 bps 到大约 1 kbps。一般来说,传感器测量某种关键性参数,比如温度和速度等,然后以范围从毫秒级到小时级的一定间隔发送当前数值(仅对应于几比特)。从大量传感器搜集信息再将其传送出去以便进一步处理的传感器网络,其中心节点往往需要较高的数据速率。这种情况下,可能需要高达 10 Mbps 的数据速率。这些“中心节点”显示出与无线局域网或固定无线接入更多的相似性。
- 语音通信通常需要 5 ~ 64 kbps 的速率,这取决于所需的音质和压缩的程度。对于需要较高频谱效率的蜂窝系统,标准的信源数据速率约为 10 kbps。对于无绳电话,由于无须复杂的语音压缩,因此可采用更高的数据速率(32 kbps)。
- 基本的数据业务需要 10 ~ 100 kbps 的数据速率。这些业务中有一类是采用手机的显示屏来提供类互联网信息的。由于显示屏较小,所需数据速率通常也低于传统互联网应用的速率。另一类数据业务提供了到便携式电脑的无线移动连接。这种情况下,大多数用户需要至少与拨号上网相当的速率(约 50 kbps),尽管基本业务有时也采用 10 kbps 的速率(利用事先为语音准备的同类通信信道)。在美国、欧洲和日本,基本数据业务基本上已被高速数据业务所替代,但在世界的其他地方它仍发挥着重要作用。
- 计算机外设与类似设备之间的通信。为了取代连接计算机外设(如鼠标和键盘)和主机的线缆(对于手机也具有类似的情况),使用具有约 1 Mbps 数据速率的无线链路。这些链路在功能特性上与以前流行的红外连接相似,但通常可以提供更高的可靠性。
- 高速数据业务。无线局域网和 3G 蜂窝系统用于提供快速互联网接入,速率范围为 0.5 ~ 100 Mbps(目前仍处于开发之中)。
- 个域网主要用于明确这种无线网络的覆盖范围(至多 10 m),但常常也含有高数据速率(超过 100 Mbps)的意思。大多用于连接消费娱乐系统的各个组件(从计算机或 DVD 播放器向电视输送视频)或者高速的计算机连接(无线 USB)。

1.3.2 覆盖范围和用户数目

不同网络之间的另一明显差异在于覆盖范围和所服务的用户数目。这里所说的“覆盖范围”是指一对发射机和接收机之间的距离。在大型网络中,一个系统可以提供的覆盖区域是通过大量基站共同作用实现的,而这与此处所说的覆盖范围几乎无关。

- 体域网(BAN)完成附着于身体的不同设备之间的通信覆盖,如从腰际皮套里的手机到挂于耳际的耳机。因此,覆盖范围是 1 m 级的。体域网一般包含于个域网中。
- 个域网包括那些可以达到约 10 m 距离的网络,覆盖了一个用户的“个人空间”。连接计算机组件的网络和家庭娱乐系统网络都是个域网的实例。由于较小的覆盖范围,个域网内

的设备数目较少,并且都与单个的“主设备”相关联。同时,相重叠的个域网(即共享相同的空间或房屋)数目也不多,通常少于5个。这就使小区规划和多址更加简单。

- 无线局域网和无绳电话覆盖更大的范围,无线局域网的覆盖范围可以达到100 m。用户数通常限制在10个左右。当用户数增多的情况(如会议或集会)发生时,每个用户的数据速率会有所下降。无绳电话可以达到300 m的覆盖范围,与一个基站连接的用户数与无线局域网是同一数量级的。注意,无线专用自动小交换机却可以覆盖更大的范围和容纳更多的用户,如前所述,基本上可以把它看成小型专用的蜂窝系统。
- 蜂窝系统的覆盖范围要大于无线局域网之类的系统。微小区的典型覆盖半径为500 m,而宏小区的覆盖半径可以是10 km或30 km。小区内的有效用户数目常在5~50之间,这取决于可用带宽和多址方案。如果系统向某个用户提供高速数据业务,有效用户数目一般会有所减少。
- 固定无线接入业务的覆盖范围与蜂窝系统接近,即在100 m到几十千米之间。同时,用户数目也与蜂窝系统有着相近的数量级。
- 卫星系统提供相当大的蜂窝尺寸,常常覆盖若干个国家甚至若干个洲的全部区域。小区尺寸完全取决于卫星的轨道,静止轨道卫星提供比近地轨道卫星更大的小区尺寸(半径为1000 km)。

图1.7是数据速率和覆盖范围的关系图。显然,如果所需覆盖范围较小,则比较容易达到更高的数据速率。固定无线接入是一个例外,它在相当长的传输距离上需要高的数据速率。

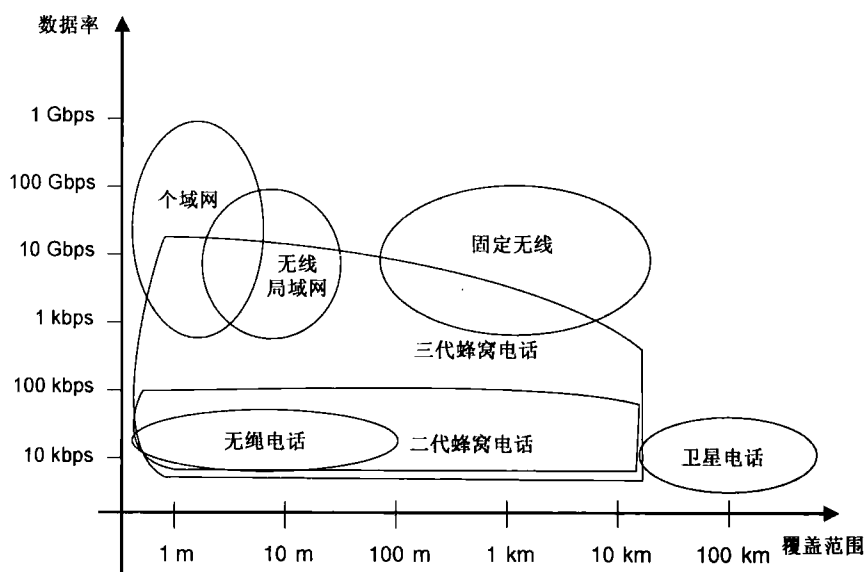


图1.7 不同应用情况下数据速率和覆盖范围的关系

1.3.3 移动性

无线系统还在对用户移动程度的支持上存在差别。无线通信对用户的主要吸引力之一在于允许用户在通信过程中到处移动。然而,对移动性的需求存在如下不同等级。

- 固定设备的安装一次到位,随后无论与基站通信还是彼此之间通信,其位置都是固定不

变的。对这些设备而言,采用无线传输技术的主要动机在于避免铺设线缆。尽管这些设备不移动,它们所使用的传播信道却可能是时变的:既可能是由于有人走过,又可能是因为环境的变化(机器、家具等重新摆放)。固定无线接入是典型的例子。还要注意所有有线通信(如公共电话交换网络)都可归入此类。

- 游牧型设备。游牧型设备在特定位置停留一定的时间(若干分钟到若干小时),然后会移动到另一不同的位置。这意味着,在一次“降落”(设备摆放妥当)期间,可认为设备是固定设备,然而从一次降落到下一次降落,环境可能发生彻底的变化。便携式电脑是典型的例子:人们不会在走来走去时操作他们的便携式电脑,而是将它们放置在桌面上使用。他们可能会将电脑携带至另一位置,然后在那里继续使用。
- 低移动性。许多通信设备会在步行速度下使用。步行用户操作的无绳电话和蜂窝电话是典型的例子。低移动性产生的效应是信道变化得相当慢,并且在多基站系统中,从一个小区到另一个小区的切换很少发生。
- 高移动性通常描述的是速度范围约 30 ~ 150 km/h 的情形。在行驶中的汽车里,人们所操作的蜂窝电话是一个典型的例子。
- 极高移动性以高速列车和飞机为代表,速度为 300 ~ 1000 km/h。这样的速度对物理层设计(多普勒频移,见第 5 章)和小区间切换这两方面形成了前所未有的挑战。

图 1.8 显示了移动性和数据率的关系。

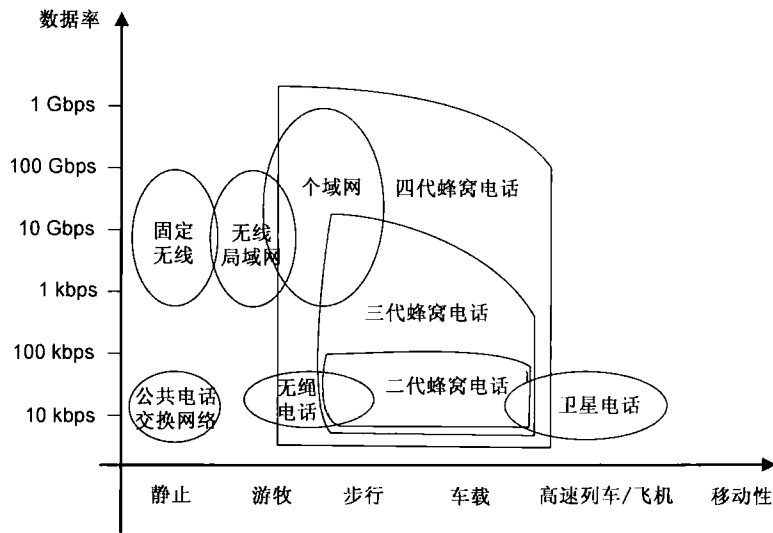


图 1.8 不同应用情况下，数据率和移动性之间的关系

1.3.4 能量消耗

能量消耗是无线设备的一个关键问题。大多数无线设备使用(一次性或可充电)电池,因为无线设备应不受任何形式的线缆的束缚,无论是用于通信的线缆,还是用于提供电能的连线。

- 可充电电池。游牧型和移动设备,如便携式电脑、蜂窝电话和无绳电话,常借助于可充电电池来工作。待机时间和使用时间是用户满意度的决定性因素之一。能量消耗一方面取决于数据传送的距离(记住,必须维持一个最小的信噪比),另一方面取决于所传

送的数据数量(信噪比正比于每比特能量)。过去100年间,电池的能量密度增加缓慢,以至于在使用和待机时间方面的主要进展来自于尽可能地降低设备的电能消耗。对于蜂窝电话,最低要求是通话时间应超过2 h,而待机时间应超过48 h。对于便携式电脑,电池能量的消耗并非主要由无线发射机所决定,而相反主要是由硬盘使用和处理器速度之类的其他因素所决定。

- 一次性电池。传感器网络的节点通常使用价格低廉又可提供较高能量密度的一次性电池。而且,通常不考虑更换电池,一般是在电池用尽以后就把包括电池和无线收发信机的传感器废弃不用。显然,与使用可充电电池的设备相比,在这种情形下能量高效的操作就显得更为重要了。
- 电力电源。可以将基站和固定设备连接到电力电源,所以能量消耗对于它们而言并非主要的关注点。因此,如果可能,将尽可能多的功能特性的实现(及由此带来的能量消耗)由移动台转移到基站是人们所希望的。

用户对于电池的需求也是重要的市场卖点,尤其是在蜂窝手机市场。

- 电池占了一部移动台重量的绝大部分(70%~80%)。手机的重量和尺寸都是重要的卖点。20世纪80年代中期,蜂窝电话一般称为“车载电话”,因为移动台只能放在汽车行李箱内运输,并通过汽车电池供电。到20世纪80年代末,电池的重量和尺寸下降到约2 kg,以至于用户可以用背包携带它。到2000年,电池重量已减少到约200 g。这一进展部分源于更高效的电池技术,但更大程度上源于手机功耗的降低。
- 手机(原材料)的成本也在相当大的程度上取决于电池。
- 用户既要求几天的待机时间,又要求在再次充电之前至少可以通话2 h。

这些“商业”因素决定了电池的最大尺寸(及电能容量),因此也决定了待机和通话期间的允许能耗。

1.3.5 频谱的使用

频谱的分配可以是基于专用的,也可以是基于共享的。这在很大程度上决定了多址方案和系统所必须提供的抗干扰能力。

- 业务和运营商专用频谱。在这种情况下,一部分确定的电磁波谱以专用的方式分配给某业务提供商。主要的例证就是蜂窝电话,网络运营者以专用的方式购买或者租用(价格一般非常高)频谱。由于是专用的分配方式,运营者对分得的频谱有完全控制权,也就可以去规划将频谱的不同部分用于不同的区域,以使干扰最小化。
- 允许存在多个运营商的频谱。
 - 业务专用频谱。在这种情况下,频谱只能供特定业务使用,而不是分配给特定的运营商。相反,用户可在无须牌照的情况下建造符合规格的设施。这种方式不需要(或不允许)干扰规划,而是按照能够避免相同区域上其他用户所带来干扰的思路进行系统设计。由于唯一的干扰源来自于相同类型的设施,相对而言,不同设备之间的相互协调就比较简单。对发射功率加以限制(对所有用户均加以限制),是这种方式得以有效实现的关键所在。否则,如果每个用户都仅为压过其他的干扰而盲目增加功率,势必导致一场用户之间的“军备竞赛”。

- 分配给不同业务和不同运营者的免费频谱。2.45 GHz 的 ISM 频段是最广为人知的例子。这一频段允许使用微波炉、WiFi 局域网和蓝牙无线连接,以及其他许多应用。对于这种情况,也要求每个用户必须遵守严格的发射功率限制,以便不对其他系统和用户造成太大的干扰。然而,用户间的相互协调(以最小化干扰)变得几乎不可能实现,因为不同系统之间无法交换协同工作的信息,甚至常常无法确定干扰的确切属性(带宽、占空比)。

2000 年以后,两项新的频谱使用举措已经公布,但至今仍未能真正广泛应用。

- 超宽带系统将信息扩展到非常宽的带宽,同时具有非常低的功率谱密度。因此,传输频段可以包括那些已分配给其他业务使用的频段,而不会造成明显的干扰。
- 自适应频谱占用。先确定某特定频段当前的使用情况,然后使用频谱中尚未占用的部分。

1.3.6 传输方向

并非所有的无线业务都需要在两个方向上传递信息。

- 单工系统仅在一个方向上发送信息,例如广播系统和寻呼系统。
- 半双工系统可以在两个方向上发送信息。然而,在任何时刻只允许一个方向的发送。步话机(Walkie-talkies)在讲话时需要用户按下一个按键,是这类系统的典型例子。注意,在使用中,一方用户在结束其发送时必须向另一方用户示意,例如表达“通话完毕”来结束本次通话,这样另一方的用户就知道自己现在可以发送了。
- 全双工系统允许同时在两个方向上传输,例如蜂窝电话和无绳电话。
- 非对称双工系统。就数据传输而言,我们常常发现一个方向(常常是下行链路)的数据速率要比另一个方向高得多。即便如此,它仍然保持着全双工的能力。

1.3.7 服务质量

不同无线业务对服务质量的要求也有着很大的差别。服务质量的第一个重要指标是针对语音业务的语音质量和针对数据业务的文件传送速度。常采用平均印象分值(MOS)来度量语音质量,它表示了对接收语音质量的大量(主观)个人判断(按 1 到 5 分的分值来衡量)的平均值(见第 15 章)。数据传输的速率简单地以 bps 来度量,显然速率越高越好。

业务的可获得性也是一个相当重要的指标。对于蜂窝电话和其他语音业务,服务质量常按照对“阻塞呼叫^①比例加上中途掉话的呼叫比例的 10 倍”取补^②的方式来计算。这个公式考虑到了一个正在进行的通话被中断要比根本无法打通一个电话更使人困扰的事实。对于欧洲和日本的蜂窝系统,这项服务质量的测量值常常可以超过 95%,而这个值在美国相当低^③。

对于紧急业务和军事应用而言,服务质量按对“阻塞呼叫比例加上中途掉话的呼叫比例”取补来测算更好一些。在紧急情况下,打不通电话的情况与正在进行的通话被中断的情况一样使人困扰。并且系统必须以更加牢靠的程度来规划,因为这时需要服务质量高于 99%。对

① 这里,“阻塞呼叫”包括了所有不成功的呼叫尝试,既包括那些由于信号强度不够所造成的不成功呼叫,也包括因网络容量不足所造成的那些。

② 取补即用 1 去减。——译者注

③ 造成这种差异的原因部分源于历史和经济因素,部分源于地理因素。

于诸如工厂自动化系统这样的应用,需要“超可靠(ultrareliable)系统”,这类系统要求服务质量超过99.99%。

还有一个相关的指标是通信的允许时延(等待时间)。对于话音通信,从一个人讲出到另一个人听到信息的时延应该不超过约100 ms。对于视频和音乐流,时延可以更大一些,因为对大多数用户而言数据流的缓冲(至多几十秒)是可接受的。在话音和视频流通信中,先发送的数据先被接收者收到这一点是很重要的。对于数据文件,可接受的时延常常会更大一些,并且数据到达接收端的顺序并不重要(例如,在从服务器下载电子邮件时,是第1封还是第7封邮件先到达客户端对用户而言并不重要)。然而,对某些数据应用而言,即使很小的时延也是致命的,例如对于控制类应用、保安和安全监控等。

1.4 经济和社会因素

1.4.1 构建无线通信系统的经济条件

无线系统的设计不仅要着眼于针对特定的应用取得优化的性能,还应在合理成本的前提下达到上述目的。由于经济因素影响系统的设计,所以科学家和工程师必须至少对市场和销售环节所施加的约束有基本的认识。无线设备设计的一些指导原则如下。

- 将尽可能多的功能由采用(更昂贵的)模拟器件实现转变为采用数字电路实现。数字电路的造价跟模拟电路的造价相比,前者随时间而降低的速度要快得多。
- 对大批量市场应用,要设法将尽可能多的元件集成到单个芯片上。大多数系统力求仅采用两片芯片,一片用于模拟射频电路,一片用于数字(基带)处理。进一步将其集成到单个芯片(片上系统)上去是人们所希望的。为利基(niche)市场所生产的产品就不同了,这些产品一般都试图采用通用处理器、专用集成电路(ASIC)或者有现货供应的器件,因为在这种市场应用场合,售出器件的数量有限,更高集成度芯片的设计显得得不偿失。
- 由于人工劳动力非常昂贵,应避免需要人力介入的任何形式的电路(如射频元件的调谐)。而且,这一点对于大批量市场产品的生产更为重要。
- 为提高开发过程及生产的效率,应将同样的芯片用到尽可能多的系统中。

谈及无线系统和业务的设计,必须辨别如下两种不同的类型。

- 对于那些移动性对其系统本身具有重要性的系统,例如蜂窝电话技术。这些业务会向用户收取额外的费用,即收费会比可获得同样通信效果的有线系统更昂贵。蜂窝电话就是一个恰当的例子。在过去,其每分钟价格一直高于陆上有线电话,并且看来还将持续高于后者,尤其是在与基于IP协议的话音(VoIP)电话业务相比时。尽管存在这一事实,如果价格差别不太大,这些业务仍然可以与传统的有线业务竞争,并最终超过有线业务。20世纪90年以来的发展状况无疑已经使人们看到了这一趋势,因为许多消费者(甚至许多公司)不再使用有线业务,而仅仅依靠蜂窝电话技术进行通信。
- 无线接入仅被当成廉价的线缆替代方案,不具有附加属性的那些业务,例如固定无线接入。这些系统必定相当注重经费的多少,因为无线基础设施的构建成本必须低于铺设新的有线连接线路或购买现有线路的使用权所需的费用。

1.4.2 无线通信市场

蜂窝电话是一个已取得巨大增长的充满活力的市场。然而,在不同的国家,其市场占有率是不同的。影响市场占有率的一些因素如下。

- 所提供业务的价格。业务价格依次受到竞争程度、运营商为获得更大市场占有率而接受一些损失的意愿,以及运营商的额外成本投入(特别是频谱执照的购买费用)的影响。然而,业务的价格并不总是市场占有率的决定性因素,斯堪迪纳维亚地区虽然有着相对高的价格,但仍然具有世界上最高的市场占有率。
- 移动台的价格。如果消费者同意订立长期合同,移动台通常由运营商赞助,并且要么是免费的,要么只以象征性的价格销售。“预付费”业务是例外情况,这种情况下用户购买指定的若干分钟的业务使用权(此时,手机以全价卖给消费者)。在市场分布的另一端,高端设备通常需要由消费者进行数目不小的共同支付。
- 所提供业务的吸引力。许多市场中,由不同网络运营商提供的业务的价格几乎是相同的。运营商们试图通过不同的特性(如更好的覆盖、文字和图片消息业务等)来区分彼此。由于允许用户来确定符合自己需要的业务,这些改进特性的提供一般也有助于扩大市场规模。
- 普遍的经济状况。显然,良好的普遍经济状况为一般客户在诸如移动通信业务之类的“非基本需求”上消费更多创造了条件。在那些收入的很大比例用来满足诸如食物和居住之类基本需求的国家,蜂窝电话市场显然就很有有限了。
- 现有电信基础设施。在那些现有陆上有线电信基础设施缺乏的国家和地区,蜂窝电话和其他无线业务可能是通信的唯一方式。这将导致高的市场占有率。遗憾的是,这些地区往往又是上面提到的经济状况糟糕的地区(大部分收入都被用来满足基本需求)。这样的事实状况尤其对固定无线接入的发展形成了阻碍^①。
- 大众的倾向性。有一些可以扩大蜂窝市场的社会因素:(i)人们对新技术(新发明)抱有积极的态度,例如在日本和斯堪迪纳维亚地区;(ii)人们认为通信是生活中必不可少的组成部分,例如在中国;(iii)人们的高度移动性,由于人们在一天绝大部分时间里都不在其办公室或家里,就像在美国那样。

无线通信已经成为了一个如此巨大的市场,以至于这一行业中绝大多数公司甚至都不为大多数消费者所知。消费者容易了解网络运营商和手机制造商。然而,零部件提供商和其他辅助行业比比皆是。

- 蜂窝电话基础设施制造商。大多数手机制造商同时也向网络运营商提供基础设施(基站、交换机等)。
- 零部件制造商。大多数手机制造商从外部提供商那里购买芯片、电池、天线等零部件。随着许多制造商和系统集成商剥离了他们的半导体部门,这一外购趋势得到了进一步的加速。甚至有些手机公司根本就不制造任何东西,而仅仅从事设计和市场运作。
- 软件提供商。软件及应用正在成为市场中日益重要的部分。例如,手机铃声业已成就

^① 简单明了地说:“这一产品的市场是那些负担不起该费用的人群。”

了数十亿欧元的市场。类似地,由于手机要具备越来越多的个人数字助理(PDA)的功能,手机操作系统和应用软件变得愈加重要了。

- 系统集成商。无线局域网和传感器网络要么需要集成到大网络中,要么需要与其他硬件结合使用(例如,传感器网络必须集成到工厂的自动化系统中)。这就为原始设备制造商(OEM)和系统集成商提供了新的行业领域。

1.4.3 行为影响力

工程技术并非凭空出现的,人们的需求改变着工程师们的研发内容,而他们的劳动成果也影响着人们的行为。蜂窝电话使我们可以随时进行通信联系,大多数人希望如此。然而应该意识到,这种应用改变了我们的生活方式。从前,我们不可能去呼叫某个人,而是去呼叫某个位置。那意味着职业或个人生活之间基本上是完全隔离的。正是由于蜂窝电话,我们可以在任何时间呼叫到任何一个人,既可能在夜晚被工作中的同事呼叫,也可能在会议进行中被自己的熟人呼叫;换句话说,职业和私人生活会掺杂在一起。较积极的一面是这样还会带来新的更便利的工作形式,以及更大的灵活性。

另一重要的行为影响力在于蜂窝电话礼节的发展(或者缺失)。大多数人倾向于认同在观赏歌剧表演期间听到手机振铃是令人不愉快的事情,同时仍然有相当数量的人在同样场景下不愿意将他们的电话转入“静音”模式。此外,人们看上去并不情愿对已经振铃的手机全然不顾。人们往往愿意打断正在进行的任何事情来接听手机。主叫识别、自动应答功能等是工程师们所提供的用于缓解这些问题的方案。

更值得关注的严肃事情是,无线设备尤其是蜂窝电话有时可能会关乎生死。登山事故发生以后能够在野外打出求助电话无疑是一种救生方法。用于雪崩幸存者的定位设备具有相似的益处。不利的一面是,被电话交谈分散注意力的司机对公路交通构成了严重的危险。佐治亚理工大学最近的研究表明,在驾驶中进行手机通话(即使使用了免提设备),所造成的危险性与酒驾无异。驾驶中进行短消息的收发情况更糟。本书作者希望读者在记住书中所呈现的许多技术知识的同时,从本书中汲取的最最重要的启示就是:千万不要在驾驶中发送短消息或通话!!!再强调一次,这个问题并非纯技术问题,尽管已提出大量方案(包括“关机”按键)来试图解决这一问题。这是一个用户应改善行为的问题,也是工程师可以做些什么来促进这一改善的问题。

第2章 无线通信的技术挑战

前一章已从应用和用户需求的角度描述了对各种无线通信系统的不同要求，本章将对无线通信系统所面临的实际挑战给出高层次的描述。最值得注意的几种挑战如下所示。

- 多径传播。即发射信号可借助不同的路径到达接收机，例如从不同的房屋或山体反射后到达接收机。
- 频谱限制。
- 能量限制。
- 用户移动性。

这将是本书的其余部分的基础，后面将会详细讨论这些挑战及其应对措施。

作为第一步，研究有线和无线通信的差别是有必要的。表 2.1 首先回顾了有线和无线系统的一些重要属性。

表 2.1 有线和无线通信

有 线 通 信	无 线 通 信
通信的发生借助于相对而言较为稳定的媒体，如铜线或光纤。媒体的属性可以明确界定并且是时不变的	由于用户的移动性和多径传播，传输媒介随时间剧烈变化
可通过在已有线缆上使用一个不同的频率来增加传输容量，亦或铺设新的线缆	当可用频谱的数量受限时，增加容量必须借助于更复杂的收发信机概念和更小的小区尺寸(在蜂窝系统中)完成
无转发站点时，通信可达范围主要受限于媒体(及相应噪声)引起的衰减；对于光纤通信，传输脉冲的失真也会限制数据传输的速率	覆盖范围既受到传输媒介(衰减、衰落和信号失真)的限制，又受到频谱效率(小区尺寸)要求的限制
来自其他用户的干扰和串话要么不会发生，要么干扰属性是不变的	来自其他用户的干扰和串话是蜂窝通信的原理本身所固有的。由于用户的移动性，它们还是时变的
传输过程中的时延也是常数，决定于线缆长度和可能要采用的转发放大器的群延迟	传输时延部分取决于基站和移动台之间的距离，因而是时变的
增加信噪比(SNR)可以使误比特率(BER)大幅度下降(近似指数规律)。这意味着相对小的发送功率提升可以大大降低误码率	就简单的系统而言，平均误比特率在增加平均信噪比的情况下只会缓慢下降(线性规律)。增加发送功率一般不会导致误比特率的明显降低。不过，更复杂的信号处理会有助于误比特率的进一步降低
由于传输媒体的性能稳定，有线传输的质量一般较高	由于令人困扰的传输媒介，不采取特殊应对措施的情况下，传输质量一般较低
在未经网络运营者许可的情况下，对有线传输进行干扰和截获几乎是不可能做到的 ^a	除非采取特殊措施，否则对无线链路的干扰是直截了当的。截获空中的信号是可能的。因此，加密就成为了阻止未经授权的信息使用的必要措施
链路的建立是基于位置的，换句话说，从一个节点到另一个节点来建立链路，而不管与节点连接的是谁	无线连接的建立是基于移动设备的，而设备一般又与特定的人相关联。连接与固定位置无关
电源要么来自于通信网络本身(如传统的陆上传输线电话)，要么来自于传统的电源专线(如传真)。在这两种情况下，能量的消耗都不是设备的设计者主要关注的问题	移动台使用可充电或一次性电池。能量效率因此成为主要关注的问题

^a 但是，由于有线互联网这一特殊通信协议的设计，对这种通信方式的截获是容易做到的。

2.1 多径传播

无线通信的传输媒介是发射机和接收机之间的无线信道。信号从发射机到达接收机可以借助许多不同的传播路径。某些情况下,在发射机和接收机之间会存在视距(LOS)路径。此外,电波传播环境中存在着各种不同的相互作用体(如,房屋、户外环境下的山体、窗户和墙壁等),信号由发射机到达接收机的过程还可以借助于在这些相互作用体表面所发生的反射或受其阻挡而发生的绕射来实现。这些可能的传播路径数目众多。如图 2.1 所示,信号经每条路径传输后的幅度、延迟(信号的传播时延)、各条路径信号在发射机处的出发方向,以及在接收机处的到达方向都不相同;最重要的是,各个多径分量(MPC)相互之间有着不同的相移。接下来将在系统设计层面上就多径传播的一些结论展开研讨。

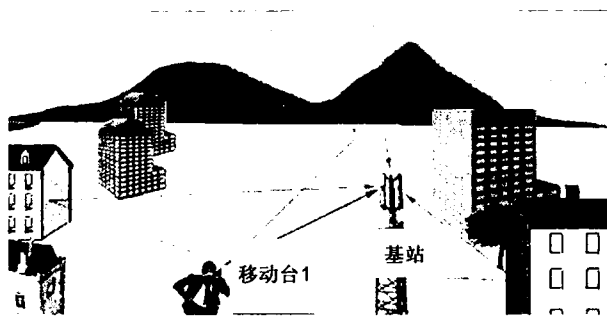


图 2.1 多径传播

2.1.1 衰落

简单的接收机无法辨别不同的多径分量,而仅仅是将其相加,以至于它们彼此之间相互干涉。这种干涉可能是相长的,也可能是相消的,取决于各个多径分量的相位状态(见图 2.2)。相位状态反过来主要取决于相应多径分量的传播路径长度,因而取决于移动台及相互作用体的位置。因此,如果发射机、接收机或相互作用体处于运动中,干涉信号及相应的合成信号幅度都会随着时间而变化。这种效应,即由于不同多径分量的相互干涉而引起的合成信号幅度的变化,称为小尺度衰落(small-scale fading)。2 GHz载频下,小于 10 cm 的移动已经足以引起从相长干涉到相消干涉,再从相消干涉到相长干涉的整整一轮变化了。换句话说,即使很小的移动也可能使信号幅度发生很大的变化。所有车载电台的用户都了解这类效应,汽车小于 1 m 的移动(例如,遇到被迫停下、启动时的行车状况)就能非常明显地影响接收信号质量。就蜂窝电话而言,为了改善接收信号质量,常常是走开一步就足够了。

作为一种附加效应,每个单独多径分量的幅度随时间(或随位置)而变化。阻挡物会造成对一个或几个多径分量的遮蔽。例如,假设图 2.3 中的移动台最初(在位置 A 处)与基站之间存在视距传输路径。当移动台移动到高层建筑的后面(在位置 B 处)时,基站与移动台之间沿直接路径(视距)传播的分量幅度会大大降低。这是由如下事实所决定的:移动台目前位于高层建筑物形成的无线电阴影中,任何穿过或者绕过该建筑物的电波都将被大大衰减,这一效应称为阴影效应(shadowing)。当然,阴影效应不仅可能对视距分量起作用,也可能对其他任何多径分量起作用。而且还要注意,阻挡物所投射的并不是“尖锐”的阴影,从“光明”(即视距不被遮蔽)区域到

“黑暗”(即视距被遮蔽)区域的转变是逐渐的^①。从光明区域进入黑暗区域,移动台必定要在较长的距离(从几米到几百米)上移动。因此,阴影效应导致大尺度衰落(large-scale fading)。

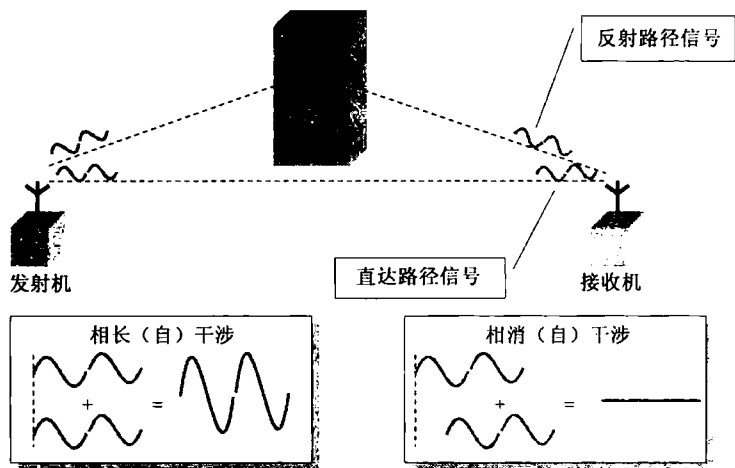


图 2.2 小尺度衰落原理

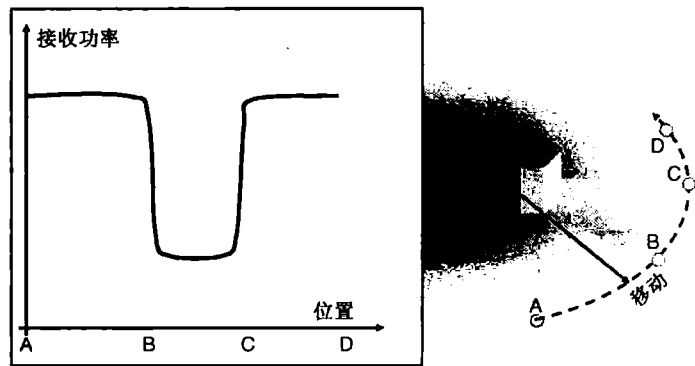


图 2.3 阴影效应原理

大尺度衰落和小尺度衰落重叠在一起,以至于接收信号看上去如图 2.4 所示。显然,在信号幅度较低时(或位置),传输质量也会较差。这将导致糟糕的话音质量(对话音通信而言)、高的误比特率及低的数据速率(对数据传输而言),并且如果传输质量在较长的持续时间内只能维持过低的水平,将导致传输中断。

对于非衰落的通信链路,若不采取特殊措施,误比特率将随着信噪比的增加而呈近似指数规律的下降,这在常规的数字通信里是众所周知的。然而,在衰落信道里,信噪比不保持恒定;相反地,链路处于衰落深陷(fading dip,即处在低信噪比位置的情况)的概率决定了误比特率的表现。由于这个原因,随着平均信噪比的增加,平均误比特率仅呈现出线性规律的下降。因此,通过简单地增加发射功率,常常并不能实现对误比特率的改善。反之,必须通过采用更复杂的发射机和接收机方案来达到改善信噪比的目的。本书第三部分和第四部分的绝大多数内容(第 13 章、第 14 章、第 16 章、第 18 章至第 22 章)专注于讨论这样一些技术。

^① 这是由于(i) 绕射效应(将在第 4 章更详细地解释);(ii) 诸如房屋之类的二级辐射源在空间上会被扩展(可类比于长的荧光灯管从不会投射出狭窄的阴影)。

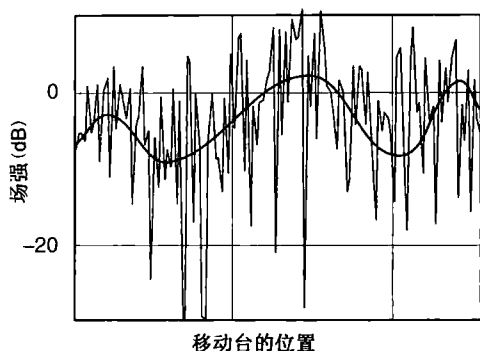


图 2.4 衰落的典型示例。细线表示归一化的瞬时场强；粗线表示每 1 m 距离上的平均场强

由于衰落的存在,准确地预测特定位置的接收信号幅度几乎是不可能的。就系统开发和布置的许多方面而言,人们认为对平均幅度及围绕平均值波动的统计规律进行预测就已经足够了。对信号幅度的完全确定性预测,如通过求解麦克斯韦方程组^①在给定环境下的近似式,常常呈现出 3 ~ 10 dB 的误差(对于总的合成信号幅度),并且就单个多径分量的特性而言,这种预测可靠性更低。有关衰落的更多细节可以在第 5 章至第 7 章找到。

2.1.2 符号间干扰

不同多径分量的传播时间是不同的。上面已经指出,这会导致多径分量之间的不同相位,从而造成窄带系统中的干扰。在具有较大带宽因而有着较好时间域分辨率的系统中^②,多径传播的主要后果是信号扩展。换句话说,信道脉冲响应不是单个 δ 脉冲,而更确切地说是一系列脉冲(对应于不同的多径分量),其中每个脉冲除了具有不同的幅度和相位之外,还具有不同的到达时间(见图 2.5)。这种信号扩展在接收端导致了符号间干扰(ISI)。具有较长传播时间并携带着第 k 比特信息的多径分量与具有较短传播时间并携带着第 $k+1$ 比特信息的多径分量同时到达接收端,相互形成干扰(见图 2.6)。若不采取特殊措施^③,则这种符号间干扰导致的差错无法用简单地增加发射功率来消除,因而常把这种错误称为不可减轻错误(irreducible errors)。

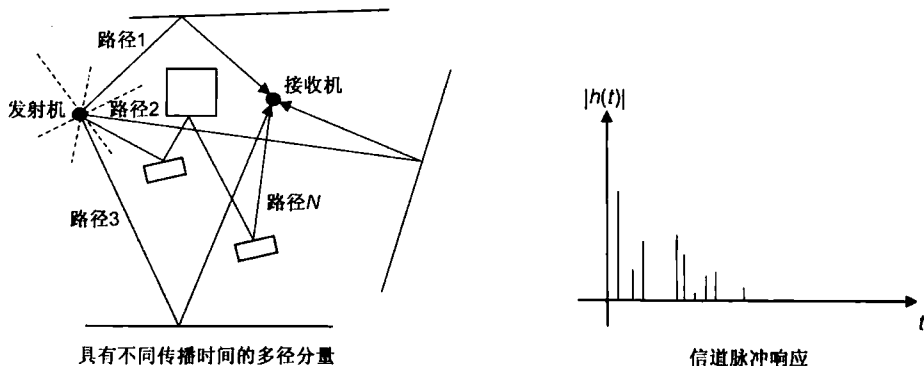


图 2.5 多径传播及其所引起的脉冲响应

- ① 最流行的确定性预测工具是“射线跟踪(ray tracing)”和“射线发射(ray launching)”,将在第 7 章中讨论。
 ② 严格地说,这里提到的是时延域的分辨率。时间域和时延域的差别将在第 5 章和第 6 章给出解释。
 ③ 特殊措施包括均衡器(见第 16 章)、Rake 接收机(见第 18 章)和 OFDM(见第 19 章)。

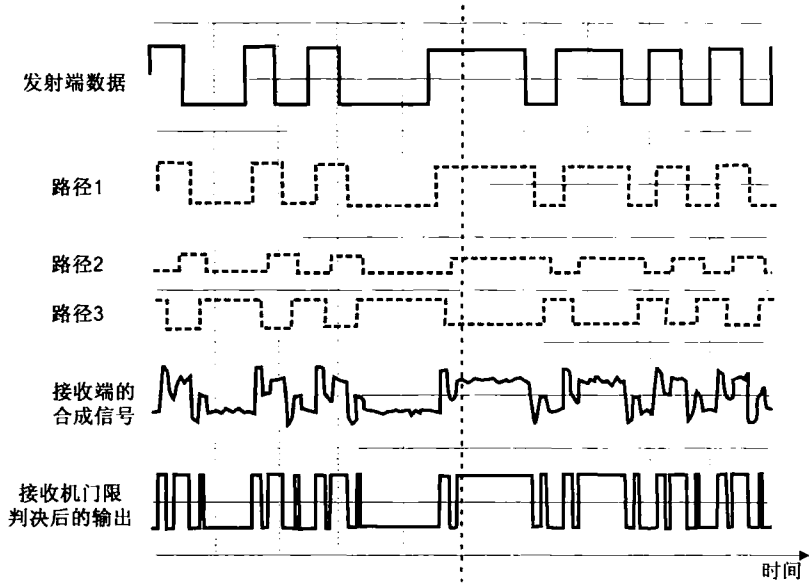


图 2.6 符号间干扰

符号间干扰 (ISI) 本质上是由符号持续时间和信道脉冲响应的持续时间之比所决定的。这意味着符号间干扰不仅对高数据传输速率很重要，而且对于会导致峰值发送速率增加的那些多址方式(如时分多址，见第 17 章)也相当重要。最后，值得注意的是，当脉冲响应持续时间相对短(并非特别短)于比特持续时间时，符号间干扰甚至也起着不小的作用(见第 12 章和第 16 章)。

2.2 频谱限制

无线通信业务的可用频谱是受限制的，并且通过国际协议来管制。因此，频谱必须以高效的方式使用。频谱的使用有两个途径：受管制的频谱使用(单个网络运营商具有频谱使用的控制权)和不受管制的频谱(每个用户可以不受额外控制地发射，只要他(或她)遵守关于发射功率和带宽的一定的限制)。下面首先回顾分配给不同通信业务的频率范围。然后，讨论受管制和不受管制接入的频率重用基本原理。

2.2.1 频率的分配

联合国(UN)的分支机构之一，即国际电信联盟(ITU)负责管理为不同无线业务分配频率的相关事务。在其三年一次的会议(世界无线电通信大会，WRC)上，确立不同地区和国家使用频谱的世界性指导原则。进一步的规定由各个国家的频率管理机构发布，这些机构包括美国联邦通信委员会(FCC)、日本无线工业及商业协会(ARIB)和欧洲邮政电信管理会议(CEPT)等。虽然确切的频率分配存在差异，但整个世界范围内针对相同的业务还是倾向于采用同一频率范围，如下所示。

- 100 MHz 以下。在这些频率上，存在公民波段(CB)电台、寻呼机和模拟无绳电话。
- 100 ~ 800 MHz。这些频率主要用于广播(无线电台和电视)业务。
- 400 ~ 500 MHz。许多蜂窝和集群无线系统使用此频段。通常是那些需要良好覆盖但又呈现出低用户密度的系统。

- 800 ~ 1000 MHz。几个蜂窝系统使用此频段(模拟系统和第二代蜂窝系统)。也有一些应急通信系统(集群无线电)使用此频段。
- 1.8 ~ 2.1 GHz。这是蜂窝通信的主要频段。目前的(第二代)蜂窝系统工作于此频段,绝大多数第三代系统也工作于此频段。许多无绳系统也工作于此频段。
- 2.4 ~ 2.5 GHz。工业、科学和医疗(ISM)频段。无绳电话、无线局域网和无线个域网工作于此频段;它们与其他的许多设备(包括微波炉)分享这一频段。
- 3.3 ~ 3.8 GHz。为固定无线接入系统预留。
- 4.8 ~ 5.8 GHz。此范围内,可找到大多数的无线局域网。5.7 ~ 5.8 GHz 之间的频率范围也可用于固定无线接入,以补充 3 GHz 频段。汽车之间的通信也工作在这一频段。
- 11 ~ 15 GHz。此范围内可以找到最普及的卫星电视业务,这些业务将 14.0 ~ 14.5 GHz 用于上行链路,将 11.7 ~ 12.2 GHz 用于下行链路。

特定业务所使用的确切频率的更多细节可以从国家的频率管理机构和 ITU 处获得。

针对不同的应用,不同的频率范围得到了最优化的使用。低载频通常更容易传播(见第4章),以至于单个基站可以覆盖一个大的区域。另一方面,低频段的可划分绝对带宽更小,较低频率的频率重用也不像更高频率那样有效^①。因此,低频段最适合于那些需要良好的覆盖,却具有较小的交互信息总速率的业务。典型的应用实例是寻呼系统和电视,寻呼之所以合适是因为发送给每个用户的信息量都较小,而电视系统只存在要传送到所有用户的单向信息流。就蜂窝系统而言,低载频对于覆盖那些低用户密度的广袤地区(美国中东部乡村地区和俄罗斯乡村地区、北斯堪迪纳维亚、阿尔卑斯山区等)再合适不过了。而对于较高用户密度的蜂窝系统和无线局域网,更高的载频通常是人们所期望的^②。

分配给不同业务的频谱数量并不总是服从于技术需求,有时更服从于历史的演进。例如,许多年来,分配给电视台使用的宝贵的低端频谱数量比按技术需求所确定的数量多得多。如果使用适宜的频率规划和不同的传送技术(包括同播),1 GHz 以下的相当一部分频谱可以被释放出来用于其他用途。2008 年至 2009 年前后,在美国和欧洲已采取了这一措施,但在其他国家这一措施的落实可能还需要更长的时间。电视台反对这一改进,因为这将要求他们对其发射机进行必要的更改。由于这些电视台有着相当大的大众舆论影响力并具备游说上层的能力,所以频率管理机构在推进合理的规则改进方面就有些犹豫不决了。

另外,值得注意的是,为不同业务分配频谱的财政条款在各个国家和各种业务之间存在着巨大的差异,有时甚至取决于频谱分配的时机。显然,分配给公共安全业务(警察、消防部门、军事)的频谱是不计成本的。甚至电视台分配到的频谱通常也是免费的。20 世纪 80 年代,蜂窝电话业务常常只需支付相当少的费用就能获得频谱,那是为了鼓励这一在当时还很新的业务的发展。20 世纪 90 年代中后期,一些国家将拍卖频谱作为增加国家收入的一种方式(可回顾 1995 年美国对个人通信系统,即 PCS 频段的拍卖,以及 2000 年前后英国和德国对通用移动通信系统,即 UMTS 频段的拍卖)。其他一些国家则选择了基于“选美比赛”的竞争方式来分配频谱——频谱的申请者为了获得牌照,必须保证在业务质量和覆盖等方面满足一定的条件。诸如无线局域网之类不受管制的业务则可以免费获得频谱。

① 正如将在下一节看到的,频率重用要求:分配给小区的信号在小区以外被大大地衰减。然而,低载频导致良好的传播,以至于信号在其被分配的小区之外很远的地方仍能保持较强的电平。

② 然而,载频不应该过高,极高的载频甚至难以覆盖很小的区域。

2.2.2 受管制频谱的频率重用

因为频谱是有限的,所以对不同位置的不同无线连接不得不使用相同的频谱。为简化讨论,以下考虑这样一种蜂窝系统:不同的连接(不同用户)通过所使用的频道(以一个特定载频为中心的频带)来区分。如果一个区域得到单个基站的服务,则可用频谱可以划分成 N 个频道,从而可以同时服务于 N 个用户。如果要服务的用户数超过 N 个,就需要多个基站,并且频道必须在不同位置重复使用。

为达到这一目的,我们将一个区域(可能是一个地区、一个国家或整个洲)划分成许多个小区(cell),并将可用频道划分成几个组,然后将频道组分配到小区上。重要的是频道组可以在多个小区上使用。唯一的要求就是使用相同频道组的小区彼此之间不会形成明显的干扰^①。显然,同一载频可以在诸如罗马(Rome)和斯德哥尔摩(Stockholm)这样两个不同的地方同时用于不同的连接。两座城市之间遥远的距离保证了斯德哥尔摩的移动台发出的信号无法到达罗马的基站,因而也就根本不会造成任何干扰。但是,为了获得高的效率,实际上必须更频繁地重复使用频率,典型情况是每个城市里重复使用好几次。结果,小区间干扰(又称为同道干扰,co-channel interference)就成为了制约传输质量的主要因素。关于同道干扰的更多细节可以在本书第四部分找到。

频谱效率描述了频率重用的有效程度,即每单位带宽和单位面积可以达到的业务密度。因而对于话音业务,频谱效率的单位为 Erlang/(Hz·m²),对数据业务则为 bit/(s·Hz·m²)。由于一个网络提供商的覆盖区域和可用带宽是固定不变的,所以增加可服务的用户数目的唯一方式就是增加频谱效率,同时也可增加收益。因而,增加频谱效率的各种方法就成为了无线通信研究的核心问题。

因为网络运营商购买了频谱的使用牌照,所以就能按照自己的规划来使用相应的频谱,即网络规划可以确保不同小区相互之间不形成明显的干扰。网络运营商可以根据需要使用足够大的发射功率,还能规定不同用户的移动台的发射功率限制^②。网络运营商还可以保证网络中的干扰仅由网络自身和其中的用户产生。

2.2.3 非管制频谱的频率重用

相对于管制频谱,有几种业务使用一般公众均可使用的频段。例如,某些无线局域网工作于分配给 ISM 业务的 2.45 GHz 频段。任何人都可以在这些频段发射,只要他们(i)限制其发射功率为规定数值;(ii)遵守关于信号形状和带宽的一定规则;(iii)按照频率管理者约定好的(有相当宽泛的定义)用途来使用频段。

结果,无线局域网的接收机会面对大量的干扰。干扰或者来自于其他无线局域网发射机,或者来自于微波炉、无绳电话及其他工作于 ISM 频段的设备。因此,一个无线局域网链路必须具备对付干扰的能力。可以通过选用 ISM 频段中干扰较少的频段并采用扩展频谱技术(见第 18 章)或者其他适宜的技术来达到这一目的。

还有将频谱分配给特定业务(如 DECT)而不是分配给特定运营商的情况。这种情况下,接收机仍将不得不面对强烈的干扰,只是这种干扰的结构是已知的。这就允许采用特别的干

① 有效干扰的门限值(即允许的干扰信号比)由调制和接收方案及传播条件共同决定。

② 此规则有一些例外情况。例如,出于对健康的关注来规定对发射功率的限制,以及由运营商采用的系统标准(如 GSM)所强加的限制。

扰减轻技术,如动态频率分配,读者可参考配套网站(www.wiley.com/go/molisch)关于 DECT 的资料。

动态频率分配可以看成认知无线电(cognitive radio)(见第21章)的一种特殊情形。在认知无线电中,发射机侦测所关心的位置处的当前哪一部分频谱未被使用,并据此动态地调整发射频率。

2.3 能量限制

真实的无线通信不仅需要 will 信息发送到空中(而非借助于线缆),而且要求采用一次性的或者可充电的电池向移动台供电,否则移动台就要受到电源“线”的束缚。而电池反过来又对设备的功耗施加了限制。低能耗的要求导致了以下几方面的技术制约。

- 发射机中的功率放大器必须具有高的效率。在移动台中,功率放大器占用了相当大比例的功率消耗,所以基本上必须采用效率大于 50% 的放大器。这些放大器,特别是 C 类或 F 类放大器,是高度非线性的^①。结果,无线通信就趋向于采用对非线性失真不敏感的调制方式。例如,恒包络信号常被优先选用(见第 11 章)。
- 必须以节能的方式来完成信号处理。这意味着对于移动台,数字逻辑的实施宜采用互补金属氧化物半导体(CMOS)这样的低功耗半导体技术,而射极耦合逻辑(ECL)这样的虽高速但功耗却更高的实现方式就不适合了。这一限制对于干扰抵消、对抗符号间干扰等方面所能采用的算法带来了重大的影响。
- 接收机(尤其在基站处)必须具有很高的灵敏度。例如,全球移动通信系统(GSM)这样规定:即使接收信号功率为 -100 dBm,也应该能得到满意的传输质量。这样的接收机比电视接收机在接收信号幅度上敏感好几个数量级。假设 GSM 标准早已规定好 -80 dBm 的接收机灵敏度,则为达到相同的覆盖范围,发送功率必须比目前高 100 倍。这同时又意味着,为了支持相等的通话时间,电池体积也将是现在的 100 倍,即质量将是 20 kg 而不是现在的 200 g。但是,对接收机灵敏度的高要求将对接收机结构(为全面利用接收信号,就要采用低噪放大器和复杂的信号处理)和网络规划带来重大影响。
- 只应在需要时才使用最大发射功率发射。换句话说,发射功率应该适应信道状态,而后者又取决于发射端和接收端之间的距离(应依此进行功率控制)。如果移动台靠近基站,则信道衰减将会很小,应该保持较低的发射功率。此外,对于语音传输,移动台应该只在用户讲话时才发射,而实际讲话时间仅占全部通信时间的 50% (非连续传输, DTX)。
- 必须定义节能的“待机”或“睡眠”模式,对于蜂窝电话要这样做,对传感器网络更应如此。

以上提到的要求中有几种是相互矛盾的。例如,构造高灵敏度接收机(因而需要复杂的信号处理)的要求与低功耗信号处理的要求恰恰相反。这就要求进行工程上的折中处理。

^① 线性放大器(例如 A 类、B 类或 AB 类)所具有的效率均低于 30%。

2.4 用户移动性

移动性是绝大多数无线系统的固有特性,对系统设计有着重要的影响。与移动性紧密关联的衰落已在 2.1.1 节讨论过。在蜂窝系统中,特别针对移动用户的第二种重要影响是:无论任何时候,系统都必须知道用户位于哪个小区中^①。

- 如果对某一移动台(用户)有呼入的呼叫,网络必须知道用户目前位于哪个小区。第一个要求是移动台应以固定间隔发射一个信号,用于通知附近的基站,它目前位于邻近区域。然后有两个数据库使用这一信息:归属位置寄存器(HLR)和访问位置寄存器(VLR)。HLR 是跟踪用户当前所在位置的中央数据库;VLR 是与特定基站有关的数据库,用于记录目前位于特定基站覆盖区域内的所有用户。假设用户 A 在旧金山(San Francisco)注册,但目前位于洛杉矶(Los Angeles)。该用户会通知离得最近的(在洛杉矶的)基站目前正位于其覆盖区域内;基站将这一信息输入进 VLR。同时,信息被传递到中央 HLR(比如,位于纽约)。现在如果有人呼叫用户 A,一条询问命令将被发送到 HLR,以找出用户的当前位置。在收到应答以后,呼叫将被路由至洛杉矶。对于洛杉矶的基站而言,用户 A 仅仅是一个“常规”用户,其数据都保存在 VLR 中。
- 如果移动台移动穿越小区边界,一个不同的基站将成为向其提供服务的新基站;换句话说,移动台从一个基站切换至另一个基站。这样的切换必须在不打断通话的前提下完成;事实上,用户根本就不应该察觉到这一变化。这将需要复杂的信令。不同的切换形式将在第 18 章(对于 CDMA 系统)和第 24 章(对于 GSM 系统)加以描述。

^① 也有无须考虑这种影响的情况。例如,对于简单的无绳系统,某个用户或者位于基站(一个且仅有一个)覆盖范围之内,或者位于其覆盖范围之外。

第3章 噪声受限和干扰受限系统

3.1 引言

本章解释链路预算的原理,以及单用户或多用户无线系统的规划问题。3.2节将讨论噪声受限系统的链路预算,并且计算不存在干扰时的最小可用发射功率(或可以达到的最大覆盖范围)。这些计算不仅能使我们对无线系统的基本能力有初步了解,而且还具有实用价值。例如,在附近不存在其他基站的情况下,无线局域网和无绳电话通常工作在噪声受限模式。如果用户密度较低(例如网络建设初期),甚至蜂窝系统有时也工作在此模式下。

3.3节将讨论干扰受限系统。正如前两章所述,不受管制的频谱使用导致用户无法控制干扰。当频谱受到管制时,网络运营商可以决定各个基站的位置,从而能够影响信号干扰比(SIR)。对于这两种情况,在进行链路预算时,将干扰的存在考虑在内是很重要的;3.3节将描述这些链路预算。第17章将讨论这些计算是如何与蜂窝原理及不同小区上的频率重用联系在一起的。

3.2 噪声受限系统

无线系统应该向用户提供一定的最低传输质量保障(见1.3节)。要获得这一传输质量,就要求在接收机(RX)处达到最小的信噪比。现在考虑只有单个基站发送而且只有一个移动台接收的情形,此时系统性能仅由(有用)信号和噪声的强度决定。当移动台逐渐远离基站时,其接收信号功率会逐渐减小,并且在达到某个特定距离时,信噪比就会达不到维持可靠通信所需的门限值。因此,系统的覆盖范围是噪声受限的(noise-limited),同样也可以称为信号功率受限(signal-power-limited)。这要看如何解释了,因为糟糕的链路质量正是由过大的噪声或者过低的信号功率造成的。

暂且假定接收功率按基站与移动台之间距离的平方(d^2)规律衰减。更确切地讲,接收功率 P_{RX} 可表示为

$$P_{RX} = P_{TX} G_{RX} G_{TX} \left(\frac{\lambda}{4\pi d} \right)^2 \quad (3.1)$$

其中, G_{RX} 和 G_{TX} 分别是接收天线和发送天线的增益^①, λ 是波长, P_{TX} 是发送功率(这一公式的推导和更多细节见第4章)。

噪声对信号的扰乱可能由以下几种成分组成。

1. 热噪声。热噪声的功率谱密度取决于天线“所见的”环境温度 T_e 。地表温度大约是300 K,而(寒冷)天空的温度近似地取 $T_e \approx 4$ K(当然,太阳照射到的方向上温度会更

^① 粗略地讲,“接收天线增益”是在一个确定方向上,与采用各向同性天线相比,使用某特定天线可以多接收到的功率的度量。发射天线增益的定义是类似的。更多细节见第9章或 Sutzman and Thiele[1997]。

高)。通常，作为一级近似，我们假定环境温度是 300 K，且各个方向都是相同的。则噪声功率谱密度为

$$N_0 = k_B T_c \tag{3.2}$$

其中， k_B 是玻尔兹曼 (Boltzmann) 常数， $k_B = 1.38 \times 10^{-23} \text{ J/K}$ ，噪声功率为

$$P_n = N_0 B \tag{3.3}$$

其中， B 是接收机带宽 (单位为 Hz)。通常采用对数单位 [功率 P 用 dBm 单位来表示，即 $10\log_{10}(P/1 \text{ mW})$]。将式(3.2)写成

$$N_0 = -174 \text{ dBm/Hz} \tag{3.4}$$

这意味着每 1 Hz 带宽包含 -174 dBm 的噪声功率。在带宽 B 中包含的噪声功率为

$$-174 + 10\log_{10}(B) \text{ dBm} \tag{3.5}$$

带宽 B 的对数，即 $10\log_{10}(B)$ ，其单位为 dBHz。

2. 人为噪声。我们可以区分如下两种类型的人为噪声。

(a) 杂散发射。许多电子设备和其他频段的无线发射机 (TX) 会在很宽的带宽上 (将无线通信系统工作的频率范围包括在内) 存在杂散发射。就城市室外环境而言，汽车点火和其他脉冲源是影响相当大的噪声源。与热噪声相反，脉冲源产生的噪声随频率的增加而减弱 (见图 3.1)。在 150 MHz，它比热噪声强 20 dB；在 900 MHz，一般要强 10 dB。在通用移动通信系统 (UMTS) 工作频段，Neubauer et al. [2001] 测得城市环境下的人为噪声有 5 dB 的噪声增强作用，乡村环境的噪声增强作用约为 1 dB。值得注意的是，大多数国家的频率管理者都对所有电子设备的“杂散”或“带外”发射进行了限制。此外，对于已获得牌照频段上的通信，这类杂散发射是唯一的人为噪声来源。在需要许可牌照的频段 (通常使用权已被牌照持有者购买了) 是不允许其他人为设置的发射机工作的，而人为噪声天生存在于这些频段里。相对于热噪声而言，人为噪声未必是服从高斯分布的。然而，为简便起见，大多数系统规划工具和理论设计中都假定其分布具有高斯性。

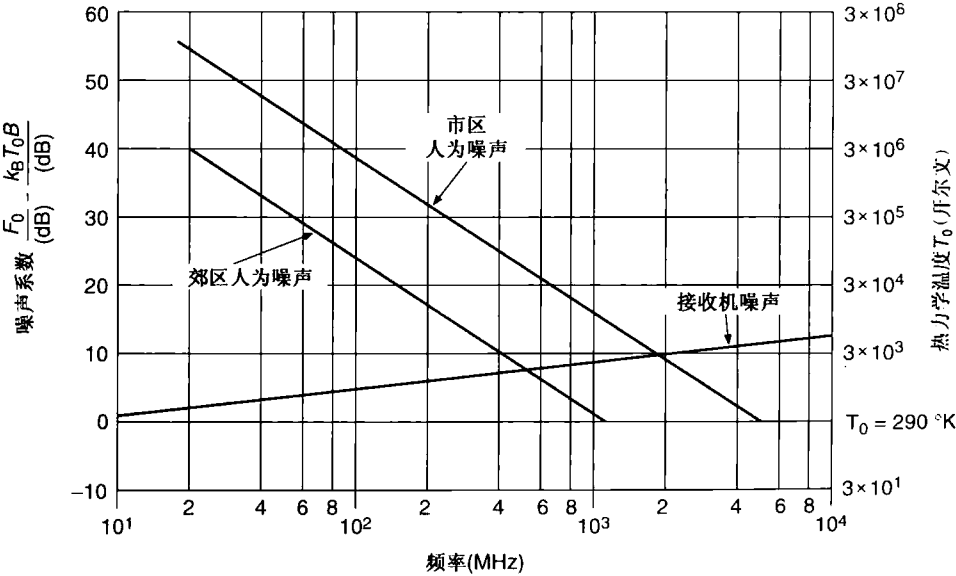


图 3.1 作为频率函数的噪声。引自 Jakes[1974] © IEEE

(b)其他故意发射源。几种无线通信系统工作于不需要牌照的频段上。每个人只要服从关于发射功率等的一些特定约束条件,就可以使用这些频段(在其上发射电磁波)。这些频段中最重要的就是2.45 GHz的工业、科学和医疗(ISM)频段。这些频段上的干扰数量相当可观。

3. 接收机噪声。接收机中的放大器和混频器是会引入噪声的,因而就增加了总的噪声功率。这种效应用噪声系数 F 描述, F 定义为接收机输入信噪比除以接收机输出(一般指下变换到基带后)信噪比。由于放大器具有增益,接收机中后面各环节所引入的噪声不如第一个环节引入的噪声影响那么大。一系列级联部件的总噪声系数 F_{eq} 的数学表达式为

$$F_{eq} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots \quad (3.6)$$

其中, F_i 和 G_i 是单个处理环节的噪声系数和增益,以绝对量值(而非dB)计。注意,在这一公式中,像衰减器这样的增益 $m < 1$ 的无源器件,可以解释为具有 $F = 1/m$ 的噪声系数和 $G = 1$ 的单位增益,也可以解释为具有 $F = 1$ 的单位噪声系数和 $G = m$ 的增益。

对于一个数字系统,常采用误比特率(BER)来描述传输质量。对于每种数字通信系统,信噪比与误比特率之间有一定的关系,这取决于调制方案、编码及一定范围内的其他因素(将在本书第三部分讨论)。这样,最低传输质量就可以与最小信噪比 SNR_{min} 通过图3.2所示的映射关系联系起来。从而,在噪声受限环境下,所有模拟和数字链路的规划方法都是相同的,目的就在于确定最小信号功率 P_s :

$$P_s = SNR_{min} + P_n \quad (3.7)$$

这里所有的量值都以dB计。然而,应当注意,实际数值因系统不同而异。

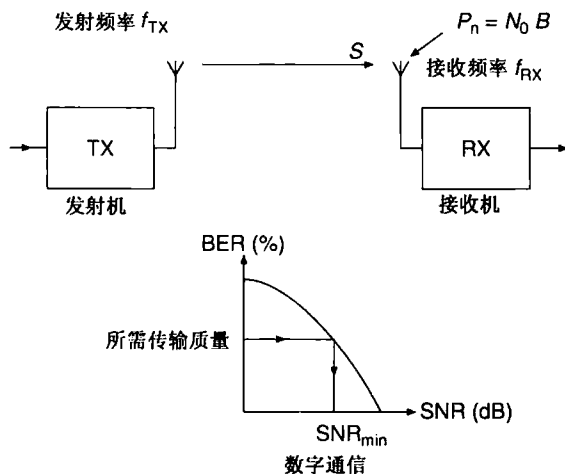


图3.2 噪声受限系统。引自 Oehrvik[1994]© Ericsson AB

3.2.1 链路预算

链路预算是计算发射机所需发射功率的最清楚、最直观的方式。它将联系发射功率和接收信噪比的所有关系式列成表格。由于大多数影响信噪比的因素以乘积形式引入,将所有关系式按对数形式表达更方便,即用dB(分贝)来表达。然而,还必须注意到,链路预算只是就

总的信噪比给出了一个近似(通常是对某种最坏情况的估计),这是因为它并没有考虑不同效应之间的某些相互作用。

在给出一些例子之前,先要强调以下几点。

- 第4章和第7章将对路径损耗,即发射机和接收机之间由于传播效应带来的衰减进行深入讨论。为了达到本章的目的,这里采用一个简单的模型,即所谓“断点”模型:对于距离 $d < d_{\text{break}}$ 的情况,根据式(3.1),功率正比于 d^{-2} 。超过断点,功率正比于 d^{-n} ,其中 n 一般为 $3.5 \sim 4.5$ 。因此接收功率表示为

$$P_{\text{RX}}(d) = P_{\text{RX}}(d_{\text{break}}) \left(\frac{d}{d_{\text{break}}} \right)^{-n}, \quad d > d_{\text{break}}$$

(3.8)

- 无线系统,尤其是移动系统,所面对的是时间和空间均存在变化的传输信道(衰落)(见2.1节)。换句话说,即使距离近似不变,接收功率也可能随着发射机和/或接收机的微小移动而发生显著变化。式(3.8)计算得到的功率仅仅是一个平均值。发射功率和这一平均接收功率的比值也称为路径损耗(路径增益的倒数)。

如果用平均接收功率作为链路预算的基础,则只有大约 50% 的时间和位置上的传输质量可以超过这一门限值^①。这是完全不能接受的服务质量。因此,我们必须追加一个衰落余量(fading margin),以确保在所有接收情况中,至少有比如 90% 的情况,超过最小接收功率(见图 3.3)。衰落余量的数值取决于衰落的幅度统计量,更多细节将在第 5 章讨论。

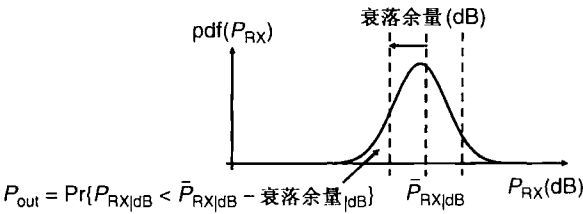


图 3.3 确保一定中断概率的衰落余量

- 天线端口处的电压和电流是互易的,在这个意义上,上行链路(移动台到基站)和下行链路(基站到移动台)也是互易的(只要上、下行链路采用相同的载频)。然而,基站和移动台的噪声系数一般会差别很大。因为移动台必须要大批量生产,所以使用低成本器件来制造更为可取,这些器件一般具有更高的噪声系数。此外,对电池使用时间的考虑决定了基站可以比移动台发射更大的功率。最后,基站和移动台在天线分集、与干扰的接近程度等方面也存在差异。因此,上行和下行链路的链路预算是不同的。

例 3.1 链路预算

考虑 GSM 系统的下行链路(还可参照第 24 章)。载波频率为 950 MHz,接收机灵敏度(依据 GSM 规范)为 -102 dBm。发射机放大器的输出功率为 30 W。发射天线的天线增益为 10 dB,连接器和合并器等的总衰减为 5 dB。衰落余量为 12 dB,断点 d_{break} 在距离 100 m 处。那么可以覆盖的距离是多少呢?

① 如果式(3.8)表示中值功率,则意味着接收功率超过这一门限值的情况正好占所有可能接收情况的 50%。

解：

发送侧：

发送功率	P_{TX}	30 W	45 dBm
天线增益	G_{TX}	10	10 dB
损耗(合并器、连接器等)	L_I		-5 dB
有效各向同性辐射功率(EIRP)			50 dBm
接收侧：			
接收机灵敏度：	P_{min}		-102 dBm
衰落余量：			12 dB
最小接收功率(均值)			-90 dBm
允许的路径损耗(EIRP与最小接收功率的差值)			140 dB
断点 $d_{break} = 100\text{ m}$ 处的路径损耗	$[\lambda(4\pi d)]^2$		72 dB
断点之外的路径损耗	$\propto d^{-n}$		68 dB

取决于路径损耗指数：

$n = 1.5 \cdots 2.5$ (视距)①

$n = 3.5 \cdots 4.5$ (非视距)

可以得到覆盖距离：

$d_{cov} = 100 \times 10^{68/(10n)}\text{ m}$ (3.9)

例如，假定 $n = 3.5$ ，则覆盖距离为 8.8 km。

这个例子非常简单，因为接收机灵敏度由系统规范所规定。要是得不到灵敏度的数据，接收端的计算会变得更复杂一些，如下例所示。

例 3.2 链路预算

考虑载频为 900 MHz 的移动无线系统，带宽为 25 kHz，仅受到热噪声的影响(环境温度 $T_c = 300\text{ K}$)。发射机和接收机的天线增益分别为 8 dB 和 -2 dB②。发送端在线缆、合并器等处的损耗为 2 dB。接收机噪声系数为 7 dB，信号的 3 dB 带宽为 25 kHz。所需的工作信噪比为 18 dB，期望的覆盖范围是 2 km。断点在 10 m 距离处；断点之外，路径损耗指数为 3.8，且衰落余量为 10 dB。则最小发射功率是多少？

解：

对这一问题的陈述方式让我们觉得以从接收端向发送端反推的方式进行求解更便捷。

噪声功率谱密度	$k_B T_c$	-174 dBm/Hz
带宽		44 dBHz
⋮		
接收端热噪声功率		-130 dBm

① 注意，即使传播环境中不存在建筑物或山体的阻挡，超过一定距离之后，视距路径也将不再存在。地球曲率造成在某距离之外视距路径被阻断，该距离大小取决于基站和移动台的高度。

② 在大多数链路预算中，假定移动台的天线增益为 0 dB。然而，近来的一些测试表明，用户头部和身体对电波的吸收及反射会降低天线增益，所引起的损耗可达 10 dB。更多细节将在第 9 章讨论。

接收机附加噪声		7 dB
所需信噪比		18 dB
⋮		
所需接收功率		- 105 dBm
从 10 m 到 2 km 距离上的路径损耗	$(200^{3.8})$	87 dB
发射机到 10 m 断点处的路径损耗	$[\lambda(4\pi d)]^2$	52 dB
移动台天线增益 G_{RX}	(2 dB 损耗)	- (-2) dB
衰落余量		10 dB
所需 EIRP		46 dBm
发射天线增益 G_{TX}	(8 dB 增益)	- 8 dB
发射端在线缆、合并器等处的损耗	L_l	2 dB
所需发射功率(放大器输出)		40 dBm

因此，所需发射功率等于 40 dBm，也就是 10 W。图 3.4 同样表示了这一链路预算过程。

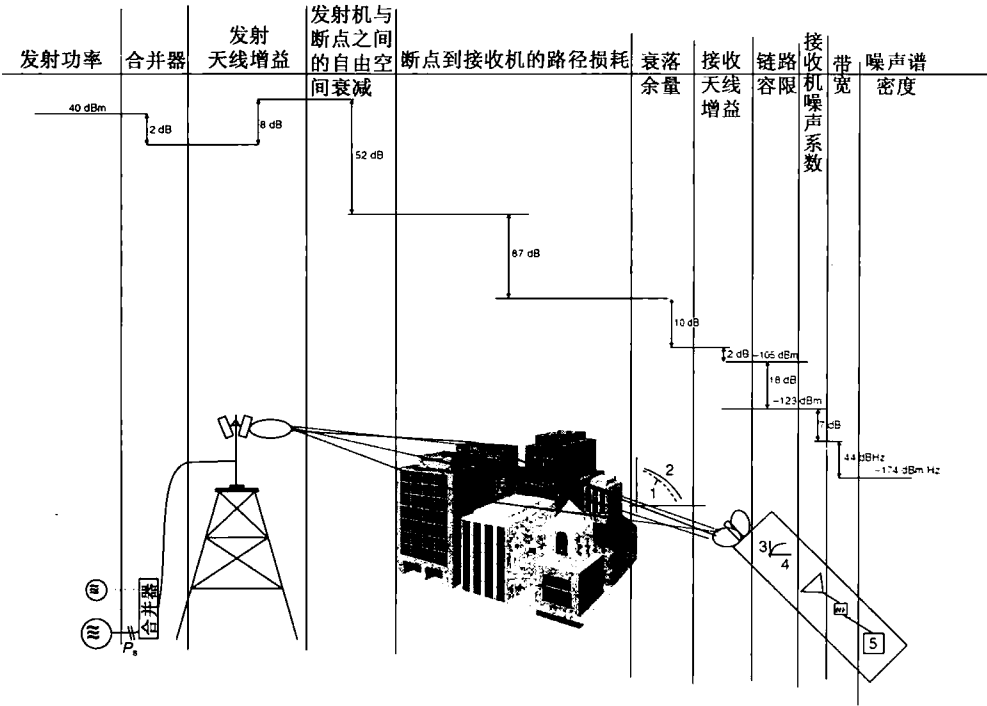


图 3.4 例 3.2 的链路预算。1 代表 10%，即十分位；2 代表中值；3 代表平均印象分值；4 代表信噪比；5 代表检测器

3.3 干扰受限系统

现在考虑的情况是干扰强大到完全决定了系统的性能，以至于噪声可以被忽略。假定基站覆盖的区域(小区)可以近似描述为一个以基站位置为圆心、半径为 R 的圆面。另外，距离“目标”基站为 D 处存在一个工作在相同频率且发射功率也相同的干扰发射机。那么，假定移

动台位于小区边界(最坏的情形), 为保证 90% 的时间达到满意的传输质量, 距离 D 应该取多大? 此处的计算可以仿照前一节的链路预算计算。作为一级近似, 我们把干扰看成高斯型的。这就允许我们将干扰看成等效噪声, 而且最小信干比 $SIR_{\min}$ 也就具有了和噪声受限情况下的最小信噪比 $SNR_{\min}$ 相同的数值。

干扰和噪声的一个不同之处在于干扰会经历衰落, 而噪声功率一般是常数(在一个短时间间隔上取平均)。因此, 为确定衰落余量, 必须考虑如下事实: (i) 所需信号在一半的时间内小于它的中值; (ii) 干扰信号在一半的时间里大于它的中值。从数学角度讲, 即信干比的累积分布函数就是在所有可能情况里, 信号与干扰这两个随机变量的比值大于某给定数值的概率, 以 $x\%$ 来表示(x 是可以得到满意传输质量的位置占所有可能位置的百分数值), 参见第 5 章。作为一级近似, 我们可以为所需信号追加衰落余量(即发射时必须增加的额外功率, 以确保所需信号的电平超过某给定值的时间达到 $x\%$, 而不是 50%), 并添加对干扰衰落余量的考虑, 即减小功率来保证只有 $(100 - x)\%$ (而不是 50%) 的时间干扰超过某个确定数值(见图 3.5)。这样做会过高地估计真实的衰落余量, 因此在系统规划中采用这个数据(指距离 D)是很稳妥的。

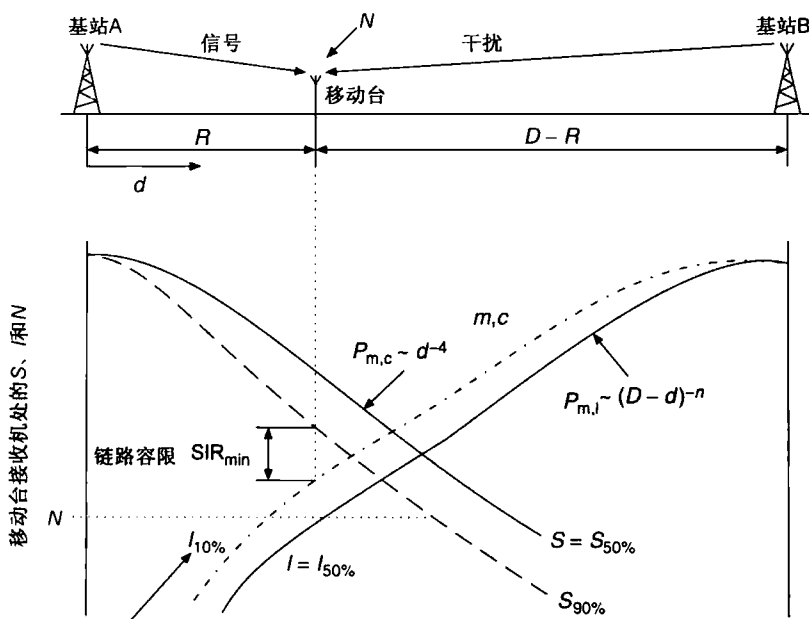


图 3.5 小区半径和重用距离的关系。实线代表中值; 虚线代表所需信号的 90% 十分位; 点划线代表干扰信号的 10% 十分位。引自 Oehrvik [1994] © Ericsson AB

第二部分 无线传播信道

- 第4章 传播机制
- 第5章 无线信道的统计描述
- 第6章 宽带和方向性信道的特性
- 第7章 信道模型
- 第8章 信道探测
- 第9章 天线

无线传播信道是连接发射机和接收机的媒介。其特性决定了信息论容量,即无线通信的最终性能限制,也决定了特定无线系统的表现。因此,知道和理解无线信道并将这些知识应用于系统设计是非常必要的。本书第二部分的目的是为了帮助读者理解无线信道。

无线信道与有线信道的区别在于多径传播(见第2章),即从发射机到接收机存在多条路径,信号在其传播过程中,可能经过反射、绕射或散射。理解信道的一种方法就是考虑所有这些不同的传播现象,以及它们如何影响每一个多径分量(Multi Path Component, MPC)。传播现象是第4章的核心。7.5节将解释如何将这些知识应用于确定信道模型和预测(射线跟踪)。

另一种思路是从现象入手。考虑重要的信道参数,如接收功率并分析其统计特性。换句话说,我们不关心信道在特定位置看起来如何,或者它如何被特定的多径分量影响,而是关注描述信道参数达到某个值的概率。最感兴趣的参数为接收功率或场强,它决定着窄带系统在噪声受限和干扰受限(见第3章)情况下的性能。读者将发现,平均接收功率随着距离而减弱。这种减弱的物理原因在4.1节和4.2节描述,模型则在7.1节详述。然而,围绕着这个均值存在着变化,这些变化可以用随机的方法建模。这些随机变化在第5章进行详细描述。

不同多径分量产生的干扰不仅产生衰落(即接收功率随时间和/或空间变化),而且产生时延色散。时延色散指的是,如果传输一个持续时间为 T 的信号,则接收信号的持续时间将是一个更长的 T' 。很自然,这将引起符号间干扰(Inter Symbol Interference, ISI, 又称码间干扰)。这种作用对早期的模拟系统关系不大,但在数字系统如全球移动通信系统中却非常重要。第三代和第四代蜂窝系统及无线局域网更受无线信道时延色散或许还有角度特性的影响。正如第6章所述,这要求引入新的参数和描述方法来量化这些特性。7.3节和7.4节给出这些参数在室外和室内环境下的典型值。

为了理解传播现象并描述信道的随机统计特性,需要进行测量。第8章描述测量设备及如何处理设备的输出。测量接收功率相当直截了当,但获取信道的时延色散并确定角度特性相当棘手。最后,第9章描述无线信道的天线。天线代表收发信机和传播信道的接口,决定着信号如何送到传播信道及如何从信道中得到信号。

第4章 传播机制

这一章讨论几个影响电磁波传播的机制，重点放在它们与无线通信相关的几个方面。最简单的情况是自由空间传播。换句话说，即一个发送天线和一个接收天线存在于自由空间中。在更为实际的情况下，还存在绝缘和导电的障碍物，即相互作用体 (Interacting Object, IO)。如果这些相互作用体有光滑的表面，电磁波就会被反射，而另一部分能量则穿透相互作用体传播。如果相互作用体表面粗糙，电磁波就会发生散射，最终电磁波将在相互作用体边缘发生绕射。这些作用将在下文逐一讨论^①。

4.1 自由空间衰减

首先讨论可能的最简单情况：自由空间中单发单收天线的情形，导出接收功率随距离变化的函数关系。

能量守恒表明，对围绕发送天线的任何一个闭合表面上的能量密度积分，都应该等于发送功率。假设某一闭合表面是以发射机天线为圆心的半径为 d 的球面，并且假设发送天线的辐射各向同性，那么该表面的能量密度为 $P_{TX}/(4\pi d^2)$ 。接收机天线有一个“有效面积” A_{RX} ：可以认为撞击到该区域的所有能量都被接收天线收集到 [Stutzman and Thiele 1997]²，于是接收能量为

$$P_{RX}(d) = P_{TX} \frac{1}{4\pi d^2} A_{RX}$$

如果发送天线不是各向同性的，那么能量密度必须乘以接收天线方向上的天线增益 G_{TX} ³，即

$$P_{RX}(d) = P_{TX} G_{TX} \frac{1}{4\pi d^2} A_{RX} \quad (4.1)$$

发送天线功率与所考虑方向增益的乘积也称为等效各向同性辐射功率 (EIRP)。

对于给定功率密度，有效天线面积正比于从天线连接处收到的功率。例如，对于一个抛物面天线，有效天线面积大约是表面的几何面积。然而，也有可能天线的几何面积非常小 (如偶极子天线)，却有相当大的有效面积。

可以证明天线有效面积与天线增益有一个简单的关系式 [Stutzman and Thiele 1997]：

$$G_{RX} = \frac{4\pi}{\lambda^2} A_{RX} \quad (4.2)$$

上式中最值得注意的事实是，对于一个固定的天线面积，天线增益随频率增加而增长。这是很直观的，因为天线的方向性取决于由波长确定的尺寸。

将式(4.2)代入式(4.1)，得到接收功率 P_{RX} 为以自由空间距离 d 为变量的函数，也称为 Friis 定律：

^① 本书中，相互作用体经常称为散射体，即使相互作用过程不是散射的。

^② 除非特别说明，本书总是将天线增益定义成与各向同性辐射体相比的增益，天线特性的深入讨论参见第9章。

$$P_{RX}(d) = P_{TX} G_{TX} G_{RX} \left(\frac{\lambda}{4\pi d} \right)^2 \quad (4.3)$$

因子 $(\lambda/4\pi d)^2$ 也称为自由空间损耗因子。

Friis 定律似乎指出自由空间中的“衰减”随频率增加而变大。这与直观不符，因为能量并没有丢失，而是在 $4\pi d^2$ 的球形表面上进行重新分配。这个机制必须独立于波长。这个表面上的矛盾就是由假设发送天线增益与波长相独立这一事实造成的。另一方面，如果假设接收天线的有效天线面积区域独立于频率，那么接收功率也和频率相独立，见式(4.1)。对于无线系统来说，假定恒定增益通常很有用，因为不同的系统(例如，工作于 900 MHz 和 1800 MHz 的系统)使用相同的天线类型(例如 $\lambda/2$ 偶极子或单极子)，而不是采用相同的天线尺寸。

Friis 定律适用于天线远场，例如发送天线和接收天线至少要间隔一个瑞利距离。瑞利距离(也称为费琅禾费距离)定义如下：

$$d_R = \frac{2L_a^2}{\lambda} \quad (4.4)$$

其中 L_a 是天线的最大尺寸，并且远场要求 $d \gg \lambda$ 且 $d \ll L_a$ 。

例 4.1 计算 20 dB 增益的方形天线的瑞利距离。

解：

增益在线性刻度下是 100。这样，有效面积约为

$$A_{RX} = \frac{\lambda^2}{4\pi} G_{RX} = 8\lambda^2 \quad (4.5)$$

对于一个方形天线， $A_{RX} = L_a^2$ ，瑞利距离为

$$d_R = \frac{2 \times 8\lambda^2}{\lambda} = 16\lambda \quad (4.6)$$

为了确定链路预算，最好将 Friis 定律写成对数刻度，那么式(4.3)变为

$$P_{RX|dBm} = P_{TX|dBm} + G_{TX|dB} + G_{RX|dB} + 20 \log \left(\frac{\lambda}{4\pi d} \right) \quad (4.7)$$

其中“|_{dB}”表示“以 dB 为单位”。为了更好地指出距离相关，最好先计算 1 m 距离处的接收功率

$$P_{RX}(1\text{ m}) = P_{TX|dBm} + G_{TX|dB} + G_{RX|dB} + 20 \log \left(\frac{\lambda|_m}{4\pi \times 1} \right) \quad (4.8)$$

式(4.8)右边最后一项在 900 MHz 时约为 -32 dB，在 1800 MHz 时约为 -38 dB。距离为 d (单位为 m)时的实际接收功率为

$$P_{RX}(d) = P_{RX}(1\text{ m}) - 20 \log(d|_m) \quad (4.9)$$

4.2 反射和透射

4.2.1 斯涅尔(Snell)定律

电磁波在到达接收机之前通常被一个或多个相互作用体所反射。相互作用体的反射系数和反射发生的方向，决定了到达接收机处的功率。这一节讨论镜面反射。这种反射发生在当光波射向光滑、巨大(相对于波长)的物体时。一个相关的机制是波的透射，例如入射波射

入和穿过某个相互作用体。透射对于波在建筑内的传播非常重要。如果基站在建筑物以外,或在另一个房间里,那么波就要穿透一座墙(绝缘层),从而能够到达接收机。

现在推导均匀平面波入射到电介质半空间的反射和透射系数。绝缘物质用介电常数 $\epsilon = \epsilon_0 \epsilon_r$ 和电导率 σ_r 来描述,其中 ϵ_0 是真空介电常数,为 8.854×10^{-12} Farad/m(法拉第/米), ϵ_r 是物质的相对介电常数。此外,还假设材料各向均质,相对磁导率 $\mu_r = 1^{(1)}$ 。介电常数和电导率可合并成一个参数,即复介电常数:

$$\delta = \epsilon_0 \delta_r = \epsilon - j \frac{\sigma_r}{2\pi f_c} \tag{4.10}$$

其中, f_c 是载波频率, j 是虚数单位。虽然这个定义仅对单一频率严格正确,但实际上可用于所有窄带系统,只要带宽远小于载波频率,同时也应该远小于 σ_r 和 ϵ 发生显著变化时的带宽^②。

平面波以入射角 Θ_e 射向半空间, Θ_e 定义为波矢量 $\mathbf{k}$ 与垂直于电介质边界的单位矢量之间的夹角。我们必须辨明横向磁性(TM)情形和横向电性(TE)情形。在 TM 的情况下,磁场分量方向平行于两个电介质的交界面,而在 TE 的情况下则是电场分量平行于该交界面(见图 4.1)。

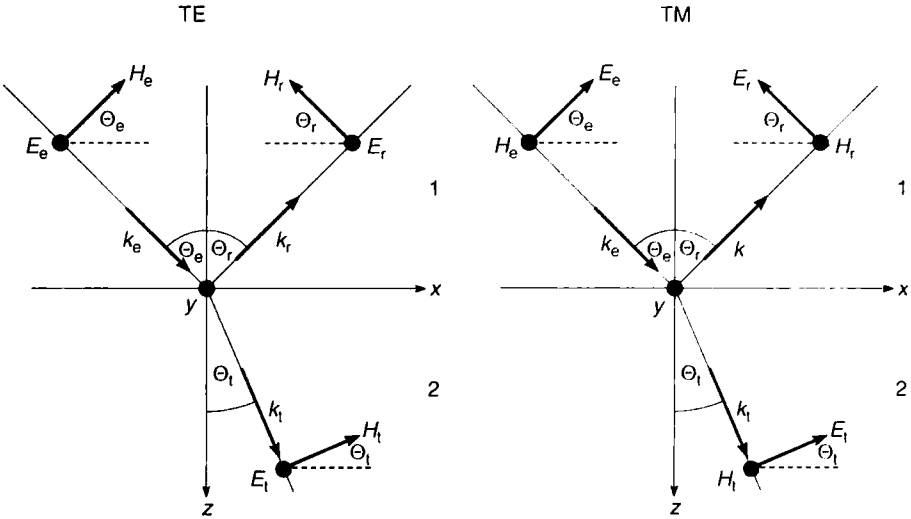


图 4.1 反射和透射

从假定的入射情况、反射和透射的平面波,以及强加了交界面处的连贯性条件,可以计算出反射和透射系数,参见 Ramo et al. [1967]。基于这些考虑,得到了斯涅尔定律,即入射角与反射角相同:

$$\Theta_r = \Theta_e \tag{4.11}$$

且透射波的角度由下式给出:

$$\frac{\sin \Theta_t}{\sin \Theta_e} = \frac{\sqrt{\delta_1}}{\sqrt{\delta_2}} \tag{4.12}$$

其中下标 1 和下标 2 表示相应的媒介。

① 这对影响移动无线电波传播的大多数物质近似正确。
② 注意,这是指在射频频意下的“窄带”。下面将遇到窄带的另一种不同定义,它和无线信道的时延色散相关

对于 TE 和 TM 情形来说, 反射和透射系数是不同的。对于 TM 极化:

$$\rho_{\text{TM}} = \frac{\sqrt{\delta_2} \cos \Theta_e - \sqrt{\delta_1} \cos(\Theta_i)}{\sqrt{\delta_2} \cos \Theta_e + \sqrt{\delta_1} \cos(\Theta_i)} \quad (4.13)$$

$$T_{\text{TM}} = \frac{2\sqrt{\delta_1} \cos(\Theta_e)}{\sqrt{\delta_2} \cos \Theta_e + \sqrt{\delta_1} \cos(\Theta_i)} \quad (4.14)$$

对于 TE 极化:

$$\rho_{\text{TE}} = \frac{\sqrt{\delta_1} \cos(\Theta_e) - \sqrt{\delta_2} \cos(\Theta_i)}{\sqrt{\delta_1} \cos(\Theta_e) + \sqrt{\delta_2} \cos(\Theta_i)} \quad (4.15)$$

$$T_{\text{TE}} = \frac{2\sqrt{\delta_1} \cos(\Theta_e)}{\sqrt{\delta_1} \cos(\Theta_e) + \sqrt{\delta_2} \cos(\Theta_i)} \quad (4.16)$$

其中, $\rho_{\text{TM}} = E_r/E_e$, $T_{\text{TM}} = E_t/E_e$ (ρ_{TE} 和 T_{TE} 是类似的)。注意, 反射系数既有幅度又有相位。图 4.2 给出了一个复介电常数 $\delta = (4 - 0.25j)\epsilon_0$ 的介电物质的幅度和相位。值得注意的是, TE 和 TM 波的反射系数在掠入射 ($\Theta_e \rightarrow 90^\circ$) 时变为 -1 (幅度为 1, 相移为 180°)。这与发生在理想导体表面的反射具有相同的反射系数。稍后读者会看到这将对无线系统中的地面反射波的影响有重要的意义。

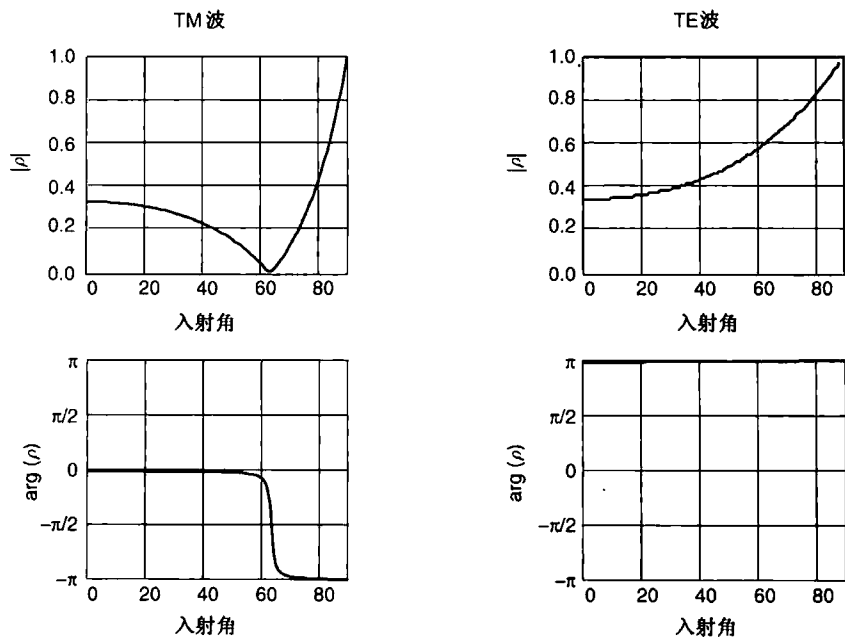


图 4.2 复介电常数为 $\delta = (4 - 0.25j)\epsilon_0$ 的介电物质的反射系数

在高损耗的物质中, 透射波不再是各向同质的平面波, 所以斯涅尔定律将不再适用, 而是在电介质界面产生一个导波。然而, 这些因素多是在理论上而非实际上有用。

4.2.2 分层电介质结构的反射和透射

前一节讨论过电介质半空间中的反射和透射。这很有意义, 例如, 对地面反射及山体等地形特征引起的反射研究就很有价值。一个相关的问题是穿过介电层的透射问题。这种情况通常发生在建筑物内的一个用户正与户外的基站通信, 或者发生在一个微微小区 (picocell) 中当

移动台和基站处于不同房间时。在这种情况下,我们关心一个波穿越墙壁所产生的衰减和相移。幸运的是,介电层的基本问题在电气工程的其他领域已经非常清楚,如光学薄膜技术[Heavens 1965],其结果可以很容易地应用到无线通信中。

最简单也是实际中最重要的情况发生在介电层两侧被空气包围时。反射和透射系数可以用部分波之和确定,导致总的透射系数为

$$T = \frac{T_1 T_2 e^{-j\alpha}}{1 + \rho_1 \rho_2 e^{-2j\alpha}} \quad (4.17)$$

反射系数为

$$\rho = \frac{\rho_1 + \rho_2 e^{-j2\alpha}}{1 + \rho_1 \rho_2 e^{-2j\alpha}} \quad (4.18)$$

其中, T_1 是波从空气到介电半空间(与考察的层有相同的介电特性)的透射系数, T_2 是波从介电层到空气的透射系数,它们都可以由 4.2.1 节的结果计算。对于与介电层 Θ_1 成某角度的波来说,量 α 是介电层的电气长度:

$$\alpha = \frac{2\pi}{\lambda} \sqrt{\epsilon_r} d_{\text{layer}} \cos(\Theta_1) \quad (4.19)$$

其中, d_{layer} 是介电层的几何长度。注意,损耗材料中有一个波导效应(见前一节的讨论),因此这一节的结果并不严格适用于有损电介质。

在多层结构中,问题变得相当复杂[Heavens 1965]。然而在实际中,即使多层结构也可以用“有效”介电常数或反射/透射常数来描述。可以直接测量复合结构的参数。相反,如果单独测量每一层的介电特性,再计算最终的有效介电常数,就很容易出错,因为不同层的测量误差将累加起来。

例 4.2 计算 4 GHz 载波的垂直入射波穿透 50 cm 砖墙时的有效 ρ 和 T 。

解:

既然 $\Theta_e = 0$, 式(4.11)和式(4.12)指出 $\Theta_r = \Theta_t = 0$ 。在 $f = 4$ GHz ($\lambda = 7.5$ cm) 时,砖的相对介电常数 $\epsilon_r = 4.44$ [Rappaport 1996], 此处忽略了电导率。在空气中有 $\epsilon_{\text{air}} = \epsilon_1 = \epsilon_0$, 式(4.13)至式(4.16)给出了空气和砖表面的反射和透射系数:

$$\left. \begin{aligned} \rho_{1,\text{TM}} &= \frac{\sqrt{\epsilon_2} - \sqrt{\epsilon_1}}{\sqrt{\epsilon_2} + \sqrt{\epsilon_1}} = 0.36 \\ \rho_{1,\text{TE}} &= \frac{\sqrt{\epsilon_1} - \sqrt{\epsilon_2}}{\sqrt{\epsilon_2} + \sqrt{\epsilon_1}} = -0.36 \\ T_{1,\text{TM}} &= \frac{2\sqrt{\epsilon_1}}{\sqrt{\epsilon_2} + \sqrt{\epsilon_1}} = 0.64 \\ T_{1,\text{TE}} &= \frac{2\sqrt{\epsilon_1}}{\sqrt{\epsilon_2} + \sqrt{\epsilon_1}} = 0.64 \end{aligned} \right\} \quad (4.20)$$

砖表面和空气之间的反射和透射系数为

$$\left. \begin{aligned} \rho_{2,\text{TM}} &= \rho_{1,\text{TE}} \\ \rho_{2,\text{TE}} &= \rho_{1,\text{TM}} \\ T_{2,\text{TM}} &= T_{1,\text{TE}} \cdot \sqrt{\epsilon_2/\epsilon_1} = 1.36 \\ T_{2,\text{TE}} &= T_{1,\text{TM}} \cdot \sqrt{\epsilon_2/\epsilon_1} = 1.36 \end{aligned} \right\} \quad (4.21)$$

注意,透射系数可能大于1(例如 $T_{2,TE}$)。这并不违反能量守恒:透射系数定义为入射与反射场的幅度比值。能量守恒仅仅指出反射和透射场的磁通(能量)密度与入射场的密度相等。

墙的电气长度 α 由式(4.19)确定:

$$\alpha = \frac{2\pi}{\lambda} \sqrt{\epsilon_r} d = \frac{2\pi}{0.075} \sqrt{4.44} \cdot 0.5 = 88.26 \quad (4.22)$$

最终,总的反射和透射系数可以由式(4.17)和式(4.18)确定,为

$$\left. \begin{aligned} T_{TM} &= \frac{T_{1,TM} T_{2,TM} e^{-j\alpha}}{1 + \rho_{1,TM} \rho_{2,TM} e^{-2j\alpha}} = \frac{0.64 \cdot 1.356 e^{-j88.26}}{1 - 0.36^2 e^{-j176.53}} = 0.90 - 0.36j \\ \rho_{TM} &= \frac{\rho_{1,TM} + \rho_{2,TM} e^{-2j\alpha}}{1 + \rho_{1,TM} \rho_{2,TM} e^{-2j\alpha}} = \frac{0.36(1 - e^{-j176.53})}{1 - 0.36^2 e^{-j176.53}} = 0.086 + 0.22j \\ T_{TE} &= \frac{T_{1,TE} T_{2,TE} e^{-j\alpha}}{1 + \rho_{1,TE} \rho_{2,TE} e^{-2j\alpha}} = \frac{T_{2,TM} T_{1,TM} e^{-j\alpha}}{1 + \rho_{2,TM} \rho_{1,TM} e^{-2j\alpha}} = T_{TM} \\ \rho_{TE} &= \frac{\rho_{1,TE} + \rho_{2,TE} e^{-2j\alpha}}{1 + \rho_{1,TE} \rho_{2,TE} e^{-2j\alpha}} = \frac{\rho_{2,TM} + \rho_{1,TM} e^{-2j\alpha}}{1 + \rho_{2,TM} \rho_{1,TM} e^{-2j\alpha}} = \frac{-\rho_{1,TM} - \rho_{2,TM} e^{-2j\alpha}}{1 + \rho_{2,TM} \rho_{1,TM} e^{-2j\alpha}} = -\rho_{TM} \end{aligned} \right\} \quad (4.23)$$

很容易验证两种情况下都有 $|\rho|^2 + |T|^2 = 1$ 。

4.2.3 d^{-4} 功率定律

无线通信的一个“民间定律”是指,接收信号功率与收发天线距离的四次方成反比。这个定律通常可以通过计算只有一个直接的(视距)波加一个地面反射波情况下的接收功率来证明其有效性。对于这种特殊情况,可以推出以下等式,细节参见附录4.A(www.wiley.com/go/molisch):

$$P_{RX}(d) \approx P_{TX} G_{TX} G_{RX} \left(\frac{h_{TX} h_{RX}}{d^2} \right)^2 \quad (4.24)$$

其中, h_{TX} 和 h_{RX} 分别是发送和接收天线的高度。该公式在距离大于如下值时有效:

$$d_{break} \geq \frac{4h_{TX}h_{RX}}{\lambda} \quad (4.25)$$

其中,“ $\geq$ ”表示“大约大于”。这个等式替代了标准的Friis定律,表明接收功率变得独立于频率。而且,从式(4.24)中可以看出,接收功率随着基站和移动台高度的平方而增加。对于距离 $d < d_{break}$,斯涅尔定律保持近似有效。

用对数刻度重写功率定律对链路预算是很有用的。假设功率以 d^{-2} 衰减,到断点 d_{break} 之后以 d^{-n} 衰减,那么接收功率变为(见第3章):

$$P_{RX}(d) = P_{RX}(1\text{ m}) - 20 \log(d_{break}/1\text{ m}) - n 10 \log(d/d_{break}) \quad (4.26)$$

图4.3表明了存在一个直射波和一个地面反射波时的接收功率,这都是从严格的方程(见附录4.A)和式(4.24)得出的。可以发现,衰减系数 $n=2$ 和 $n=4$ 之间的变化实际上并不是明显的断点,而是很平滑。因此对 d^{-4} 法则进行严格试验不太可能。根据式(4.25),断点应该在 $d=90\text{ m}$ 处。从图上看,这似乎是近似合理的。

上面推导的几个公式(及附录4.A中的公式)是一致的,但必须强调的是,它们并不是对无线信道的通用描述。毕竟,传播不总是发生在平坦的地面上,而是存在多条传播路径的可能性,视距路径经常被相互作用体遮挡等等。因此,推导的理论在几个方面都与实际信道中的测量结果不符。

- 1. $n=4$ 并不是一个在较大距离时普遍适用的衰减指数。更确切的讲, $1.5 < n < 5.5$ 内的值都测到过, 并且实际的值在很大程度上取决于周围的环境。 $n=4$ 最多是对各种环境的一个平均值。
- 2. $n=2$ 和 $n=4$ 之间的变化几乎不发生在式(4.25)的预测值 d_{break} 处。
- 3. 测量也表明还会有第二个断点, 超过断点后适用指数 $n > 6$ 。上面的模型中根本没有预测到这种效果。对于某些情形, 这种效果可以用无线电地平线(即地球的曲率)来解释, 这没有包括在上面的模型中[Parsons 1992]。

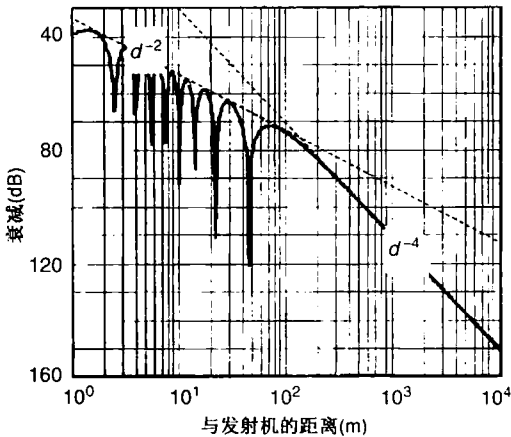


图 4.3 理想反射地面的传播。基站的高度为 5 m；移动台的高度为 1.5 m

4.3 绕射

到现在为止, 推导的所有公式都是针对无限延伸的相互作用体的。然而, 真正的相互作用体(比如大楼, 汽车等)都是空间有限的。而有限大小的物体并不会产生尖锐的阴影(这一点可以通过几何光学证明), 而是发生绕射, 这是由于电磁波辐射的波特特性决定的。只有在波长非常小(高频率)的限制下, 几何光学变得准确。

接下来, 首先考虑两个经典的绕射问题: (i) 一个均匀平面波被刀刃或屏绕射; (ii) 一个均匀平面波被一个楔形物绕射。首先推导其绕射系数, 这将反映出一个障碍物之后的阴影区域里可以接收到多少功率; 其次考虑多屏级联所产生的影响。

4.3.1 单屏或楔形绕射

绕射系数

最简单的绕射问题是一束均匀平面波被一个半无限的屏所绕射, 如图 4.4 所示。根据惠更斯原理, 可以这样理解绕射: 波阵面的每一点都可以看成球面波的源点。对一个均匀平面波来说, 多个球面波的叠加产生了另一个均匀平面波, 如图 4.4 中平面 A'到 B'之间的变化所示。然而, 如果屏阻挡住一部分点源(及其所关联的球面波), 产生的波阵面就不再是平面波了, 如图 4.4 中平面 B'到 C'之间的变化。相长和相消干涉发生在不同的方向^①。

① 出于更精确的考虑, 值得注意的是惠更斯原理并不严密。一个根据麦克斯韦方程的推导, 同时包括了必要假设的讨论, 可以参阅 Marcuse[1991]等书。

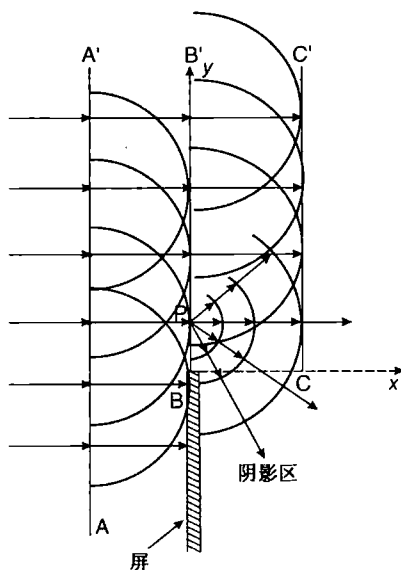


图 4.4 惠更斯原理

屏右侧($x \geq 0$)的任何一点的电场可以用仅包含一个标准积分的形式表示,即菲涅尔(Fresnel)积分。入射场用 $\exp(-jk_0x)$ 表示,总的场变为[Vaughan and Andersen 2003]:

$$E_{\text{total}} = \exp(-jk_0x) \left(\frac{1}{2} - \frac{\exp(j\pi/4)}{\sqrt{2}} F(v_F) \right) = \exp(-jk_0x) \tilde{F}(v_F) \quad (4.27)$$

其中 $v_F = -2y \sqrt{\lambda x}$, 并且菲涅尔积分 $F(v_F)$ 定义为

$$F(v_F) = \int_0^{v_F} \exp\left(-j\pi \frac{t^2}{2}\right) dt \quad (4.28)$$

图 4.5 画出了这个函数。有意思的是,对于 v_F 的某些值, $\tilde{F}(v_F)$ 的取值将会变得大于 1。这暗示某些特定位置的接收功率实际上可能由于屏的出现而增加了。此外,惠更斯原理也提供如下解释:在特定位置通常会产生相消干涉的球面波被屏蔽掉了。但是,要注意到全部能量(对全部波阵面进行积分)不会因为屏的存在而增加。

现在考虑图 4.6 所示的更一般的几何图形。发送天线的高度为 h_{TX} , 接收天线的高度为 h_{RX} , 屏从 $-\infty$ 扩展到 h_s 。那么绕射角 θ_d 就是

$$\theta_d = \arctan\left(\frac{h_s - h_{\text{TX}}}{d_{\text{TX}}}\right) + \arctan\left(\frac{h_s - h_{\text{RX}}}{d_{\text{RX}}}\right) \quad (4.29)$$

菲涅尔参数 v_F 可以根据 θ_d 求得

$$v_F = \theta_d \sqrt{\frac{2d_{\text{TX}}d_{\text{RX}}}{\lambda(d_{\text{TX}} + d_{\text{RX}})}} \quad (4.30)$$

其中,要求 d_{TX} 和 d_{RX} 远大于 h 和 λ 。可以再次用式(4.27)计算场强,其中菲涅尔参数用式(4.30)来计算。

注意,上面给出的结果是近似合理的,它忽略了入射场的极化。对 TE 和 TM 两种情况的精确公式可以在 Bowman et al. [1987] 中找到。

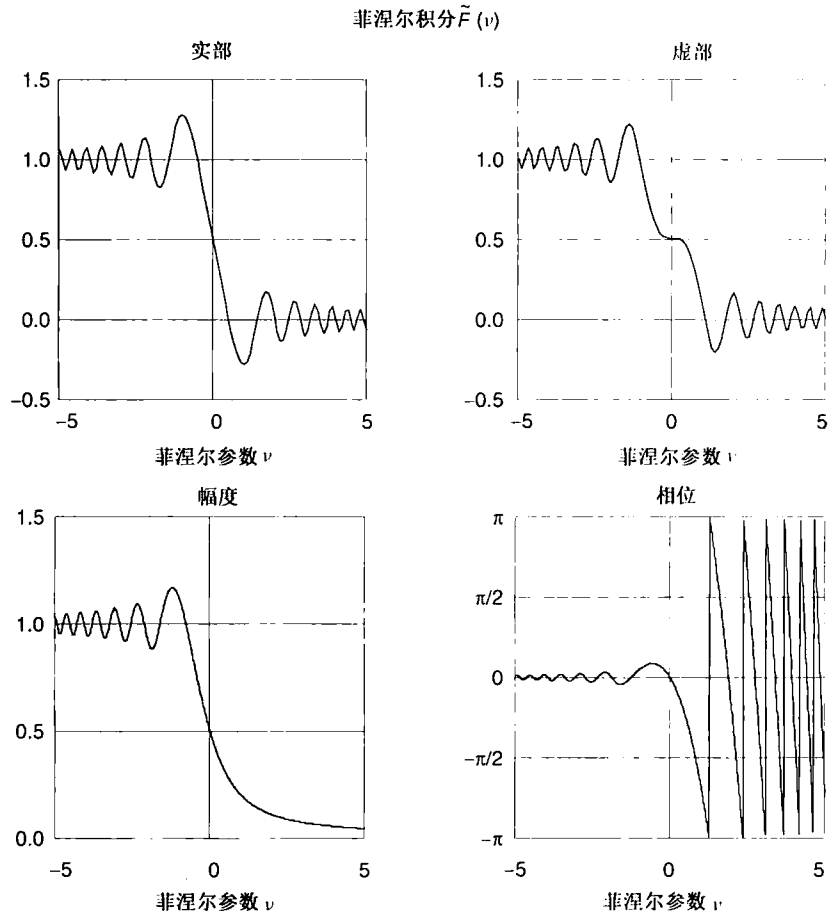


图 4.5 菲涅尔积分

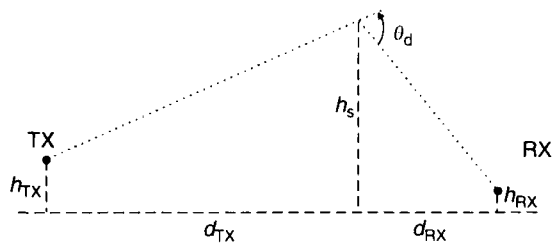


图 4.6 计算菲涅尔参数的几何图形

例 4.3 计算绕射系数。考虑由一个屏阻挡产生的绕射， $d_{TX}=200\text{ m}$ ， $d_{RX}=50\text{ m}$ ， $h_{TX}=20\text{ m}$ ， $h_{RX}=1.5\text{ m}$ ， $h_s=40\text{ m}$ ，中心频率为 900 MHz 。计算其绕射系数。

解：

中心频率为 900 MHz 意味着波长 $\lambda=1/3\text{ m}$ 。用式(4.29)计算绕射角 θ_d 可得

$$\begin{aligned}\theta_d &= \arctan\left(\frac{h_s - h_{TX}}{d_{TX}}\right) + \arctan\left(\frac{h_s - h_{RX}}{d_{RX}}\right) \\ &= \arctan\left(\frac{40 - 20}{200}\right) + \arctan\left(\frac{40 - 1.5}{50}\right) = 0.756\text{ rad}\end{aligned}\tag{4.31}$$

再用式(4.30)计算菲涅尔参数如下:

$$\nu_F = \theta_d \sqrt{\frac{2d_{TX}d_{RX}}{\lambda(d_{TX} + d_{RX})}} = 0.756 \sqrt{\frac{2 \cdot 200 \cdot 50}{1/3 \cdot (200 + 50)}} = 11.71 \quad (4.32)$$

计算式(4.28), 用 MATLAB 或根据参考文献 Abramowitz and Stegun[1965] 可得

$$F(11.71) = \int_0^{\nu_F} \exp\left(-j\pi \frac{t^2}{2}\right) dt \approx 0.527 - j0.505 \quad (4.33)$$

最后, 式(4.27)给出了总的接收场如下:

$$\begin{aligned} E_{\text{total}} &= \exp(-jk_0x) \left(\frac{1}{2} - \frac{\exp(j\pi/4)}{\sqrt{2}} F(11.71) \right) \\ &= \exp(-jk_0x) \left(\frac{1}{2} - \frac{\exp(j\pi/4)}{\sqrt{2}} (0.527 - j0.505) \right) \\ &= (-0.016 - j0.011) \exp(-jk_0x) \end{aligned} \quad (4.34)$$

菲涅尔环带

按照菲涅尔环带的概念, 一个障碍物的影响可以定性地、直观地估算出来。图 4.7 示出了其基本原理。以基站和移动台为焦点画一个椭圆, 根据椭圆的定义, 被椭圆上任意一点反射的光线有相同的路径长度(相同的传播时间)。椭圆的偏心率决定与视距(即两个焦点之间的直接连接)相比增加的路径长度。附加距离为 $\lambda/2$ 的整数倍的椭圆称为“菲涅尔椭圆”。附加路径也会导致一个附加相移, 所以椭圆也可以用它们引起的相移来描述。更明确地, 第 i 个菲涅尔椭圆引起的相移为 $i \cdot \pi$ 。

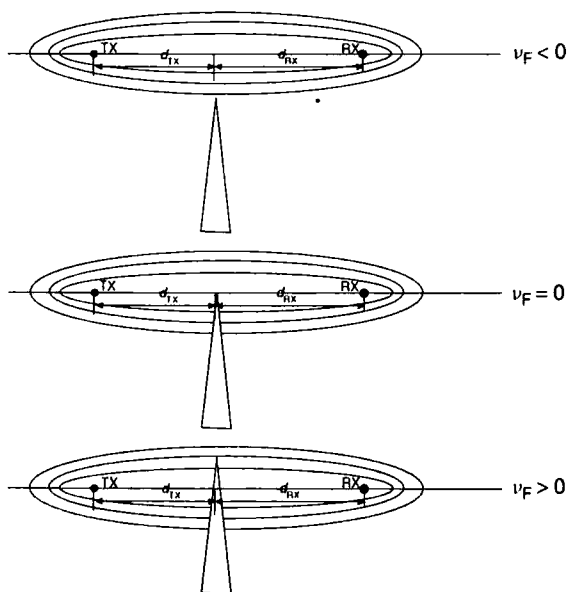


图 4.7 菲涅尔椭圆原理

菲涅尔环带也可以用来解释 d^{-4} 法则。传输遵循自由空间法则, 直到第一个菲涅尔椭圆接触到地面的距离。在这个距离上, 也就是断点距离上, 直射波与反射波的相位差为 π 。

楔形物绕射

半无限的吸收屏是解释绕射的一个有用工具,因为它是最简单的结构。然而,市区环境中的许多障碍物用楔形结构表示会更好,如图 4.8 所示。楔形物绕射问题已经讨论了大约 100 年,并且现在仍是一个活跃的研究领域。根据边界条件,可以获得适用于任意观察点的精确解,或者仅适用于远场(即远离楔形物)的近似解。后者通常简单得多,这里也只考虑这种情况。

由绕射产生的场的部分可以写成入射场与以下 3 项的乘积:一个相位因子 $\exp(-jk_0 d_{RX})$ 、一个几何因子 $A(d_{TX}, d_{RX})$ 和一个绕射系数 $D(\phi_{TX}, \phi_{RX})$, 其中几何因子仅仅取决于发射机和接收机距离楔形物的距离,绕射系数取决于绕射角[Vaughan and Andersen 2003]:

$$E_{diff} = E_{inc,0} D(\phi_{TX}, \phi_{RX}) A(d_{TX}, d_{RX}) \exp(-jk_0 d_{RX})$$

(4.35)

绕射场必须加到由几何光学^①计算出的场上。

几何参量的定义如图 4.8 所示。几何因子由下式给出:

$$A(d_{TX}, d_{RX}) = \sqrt{\frac{d_{TX}}{d_{RX}(d_{TX} + d_{RX})}}$$

(4.36)

绕射系数 D 取决于边界条件,即反射系数 ρ_{TX} 和 ρ_{RX} 。明确的公式在附录 4. B 中给出(www.wiley.com/go/molisch)。

4.3.2 多屏绕射

单屏绕射已被广泛研究,因为它可以用闭式数学来计算,并且构成了解决其他更复杂问题的基础。实际上,我们通常会遇到发射机和接收机之间有多个相互作用体的情况,比如越过市区环境的房顶传播时,正如在图 4.9 中所看到的,这种情况可以用多屏绕射来很好地近似。遗憾的是,多屏绕射是一个非常具有挑战性的数学问题,并且除了几种特殊情况,没有求精确解的一般方法。尽管如此,文献中还是提出了大量的近似方法,在这一节的余下部分会给出概述。

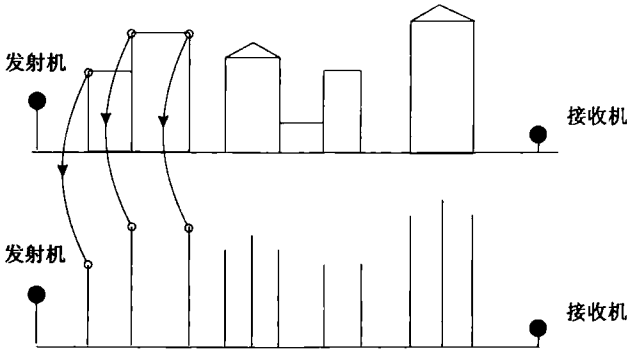


图 4.9 采用一系列屏近似多个建筑物

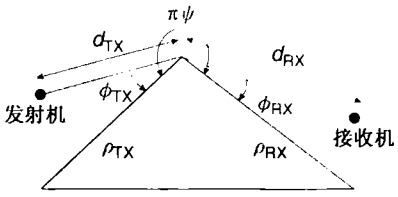


图 4.8 楔形物绕射的几何结构

① 如果楔形物上的入射场可以写成 $E_{inc,0} = E_0(-jk_0 d_{TX})/d_{TX}$, 上式就变得完全关于 d_{TX} 和 d_{RX} 对称。

Bullington 方法

Bullington 方法是使用一个“等价”的单屏来替代多屏。这个等价屏是用如下方法推导的：从发射机出发作各个实际障碍物的切线，并且选择最陡峭的那一条（即上升角最大的那一条），那么所有的障碍物或者与这条直线相接触，或者在这条直线以下。同样，从接收机出发作各个障碍物的切线，选择最陡峭的那一条。等价屏就取决于最陡峭的发射机切线和最陡峭的接收机切线的交界面（见图 4.10），在该单屏处由绕射造成的场就可以用 4.3.1 节的方法来计算。

Bullington 方法最大的吸引力就是简单。然而，这种简单性同样也带来了相当大的不准确性。物理存在的大多数屏不会影响等效屏的位置，甚至是最高的障碍物也不会产生影响。考察图 4.10，如果最高的障碍物位于屏 01 和屏 02 之间，它就能位于切线以下，即使它比屏 01 和屏 02 都高，也不会影响等效屏。在实际中，这些高的障碍物确实会对传输损耗产生影响并且产生一个附加的衰减。这样，Bullington 方法给出的往往是接收功率的乐观预测。

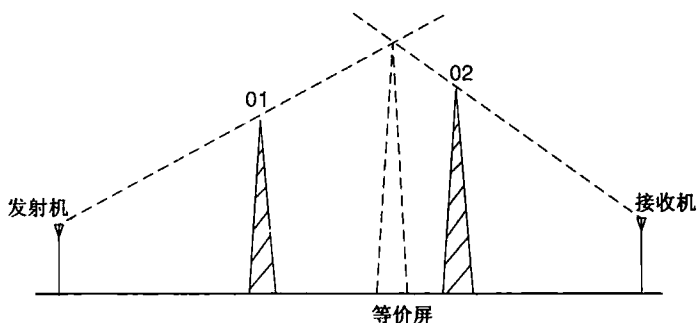


图 4.10 按照 Bullington 方法得到的等价屏。引自 Parsons[1992]© John Wiley & Sons, Ltd

Epstein-Petersen 方法

Bullington 方法的低精度源于仅仅两个障碍物就决定了等效屏的这一事实，并且从而决定了绕射系数。这个问题可以由 Epstein-Petersen 方法稍微缓解 [Epstein and Petersen 1953]。这种方法单独计算每个屏的绕射损耗。要计算某一特定屏的衰减，可分别在其相邻的左右两个屏的顶端放置虚拟的发射机 (TX) 和接收机 (RX)（见图 4.11）。这个屏的绕射系数和衰减可以用 4.3.1 节的原理很容易地计算出来。然后把不同屏引起的衰减相加（以对数刻度）。因而这种方法包含了所有屏的影响。

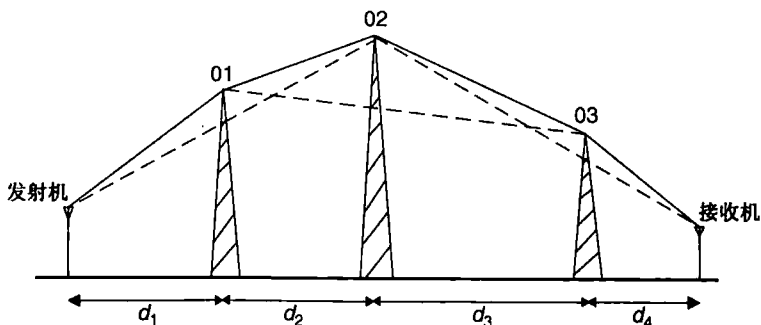


图 4.11 Epstein-Petersen 方法。引自 Parsons[1992]© John Wiley & Sons, Ltd

虽然这是一个更精确的模型，但这种方法仍然只是近似。它使用的绕射衰减式 (4.27) 基

于发射机在屏远场的假设。然而,如果两个屏距离很近,与该假设相违背,就会发生显著的误差。

“远场假设”引起的不准确可以用斜面绕射方法来进行改善。在这种方法中,场展开成泰勒级数。除了加强屏处电场连续性的零阶项(远场)以外,还要考察一阶项,用来加强场的一次导数的连续性。这导致改进的系数 A 和 D 由递归式来确定[Andersen 1997]。

Deygout 方法

Deygout 方法的体系与 Epstein-Petersen 方法近似,因为它也是要把每个屏引起的衰减相加[Deygout 1960]。然而 Deygout 方法中的绕射角是用不同的算法来定义的。

- 第一步,确定当只有第 i 个屏存在时发射机和接收机之间的衰减。
- 引起最大衰减的屏定义为“主屏”,其索引定义为 i_{ms} 。
- 计算发射机与主屏尖端由第 j 个屏引起的衰减(j 从 1 到 i_{ms})。引起最大衰减的屏定义为“次主屏”。同样地,计算主屏与接收机由第 j 个屏引起的衰减($j > i_{ms} + 1$)。
- 作为可选步骤,重复该过程以产生“辅屏”,等等。
- 把所有考虑的屏产生的损耗相加(以 dB 为单位)。

如果实际上有一个屏起主导作用,大部分损耗是由它引起的,则 Deygout 方法会工作得很好。否则,它就会产生相当大的误差。

例 4.4 有三个屏,相互间隔为 20 m,高度分别为 30 m、40 m 和 25 m。第一个屏距离发射机 30 m 远,最后一个屏距离接收机 100 m 远。发射机高度为 1.5 m,接收机高度为 30 m。用 Deygout 方法计算 900 MHz 绕射时所引起的衰减。

解:

由某一屏引起的衰减 L 如下:

$$L = -20 \log \tilde{F}(\nu_F) \quad (4.37)$$

其中, $\tilde{F}(\nu_F)$ 根据式(4.27)定义为

$$\tilde{F}(\nu_F) = \frac{1}{2} - \frac{\exp(j\pi/4)}{\sqrt{2}} F(\nu_F) \quad (4.38)$$

首先,确定屏 1 引起的衰减,绕射角为 θ_d , 菲涅尔参数为 ν_F , 衰减 L_1 如下:

$$\left. \begin{aligned} \theta_d &= \arctan\left(\frac{30-1.5}{30}\right) + \arctan\left(\frac{30-30}{140}\right) = 0.760 \text{ rad} \\ \nu_F &= 0.760 \sqrt{\frac{2 \cdot 30 \cdot 140}{1/3 \cdot (30 + 140)}} = 9.25 \\ L_1 &= -20 \log \left(\left| \frac{1}{2} - \frac{\exp(j\pi/4)}{\sqrt{2}} F(9.25) \right| \right) \\ &\approx -20 \log \left(\left| \frac{1}{2} - \frac{\exp(j\pi/4)}{\sqrt{2}} \cdot (0.522 - j0.527) \right| \right) = 32.28 \text{ dB} \end{aligned} \right\} \quad (4.39)$$

其中菲涅尔积分 $F(\nu_F)$ 是数值估计。同样,屏 2 和屏 3 所引起的衰减分别为 $L_2 = 33.59$ dB 和 $L_3 = 25.64$ dB。因此,主要衰减由屏 2 造成的,屏 2 就成为“主屏”。

接下来,计算发射机与屏 2 尖端之间的衰减,是由屏 1 引起的。绕射角 θ_d , 菲涅尔参数 ν_F 及衰减 L_4 变为

$$\left. \begin{aligned} \theta_d &= \arctan\left(\frac{30-1.5}{30}\right) + \arctan\left(\frac{30-40}{20}\right) = 0.296 \text{ rad} \\ \nu_F &= 0.296 \sqrt{\frac{2 \cdot 30 \cdot 20}{1/3 \cdot (30+20)}} = 2.51 \\ L_4 &= -20 \log\left(\frac{1}{2} - \frac{\exp(-j\pi/4)}{\sqrt{2}} F(2.51)\right) \\ &\approx -20 \log\left(\frac{1}{2} - \frac{\exp(-j\pi/4)}{\sqrt{2}} \cdot (0.446 - j0.614)\right) = 21.01 \text{ dB} \end{aligned} \right\} \quad (4.40)$$

同样, 屏2尖端到接收机之间由屏3引起的衰减为 $L_5 = 0.17 \text{ dB}$ 。绕射所引起的总衰减就是所有衰减之和, 即

$$\begin{aligned} L_{\text{total}} &= L_2 + L_4 + L_5 \\ &= 33.59 + 21.01 + 0.17 = 54.77 \text{ dB} \end{aligned}$$

经验模型

国际电信联盟提出了一个非常简单的半经验模型来计算绕射损耗(除自由空间衰减之外的损耗):

$$L_{\text{total}} = \sum_{i=1}^N L_i + 20 \log C_N \quad (4.41)$$

其中, L_i 是每个单独的屏的绕射损耗(以 dB 为单位), C_N 的定义如下:

$$C_N = \sqrt{\frac{P_a}{P_b}} \quad (4.42)$$

其中,

$$\left. \begin{aligned} P_a &= d_{p1} \prod_{i=1}^N d_{ni} \left(d_{p1} + \sum_{j=1}^N d_{nj} \right) \\ P_b &= d_{p1} d_{nN} \prod_{i=1}^N (d_{pi} + d_{ni}) \end{aligned} \right\} \quad (4.43)$$

其中, d_{pi} 是(地理上)到前面屏顶端的距离, d_{ni} 是到其后面的一个屏的距离。Li et al. [1997] 说明这个公式会产生大的误差并提出了一个改进的定义:

$$C_N = \frac{P_a}{P_b} \quad (4.44)$$

不同方法之间的比较

仅在一种特殊的情况下能够很容易地求出精确解, 即所有屏有相同的高度, 并且与发射机和接收机天线的高度相同。在这种情况下, 绕射损耗(在线性刻度下!)与屏数成比例, 为 $1/(N_{\text{screen}} + 1)$ 。让我们检查一下上面的近似方法能否给出这个结果(见图 4.12)。

- Bullington 方法独立于屏数, 因而显然地给出了错误的函数依赖关系。
- Epstein-Petersen 方法在对数刻度上对衰减求和, 因而导致了线性刻度上总的衰减呈指数增长。
- 同样, Deygout 方法和 ITU 经验模型方法得到的总损耗与屏数呈指数增长。
- 斜面绕射方法(15 屏以下)和改进的 ITU 经验模型方法导致总的衰减呈线性增长, 因而正确预测了这一趋势。

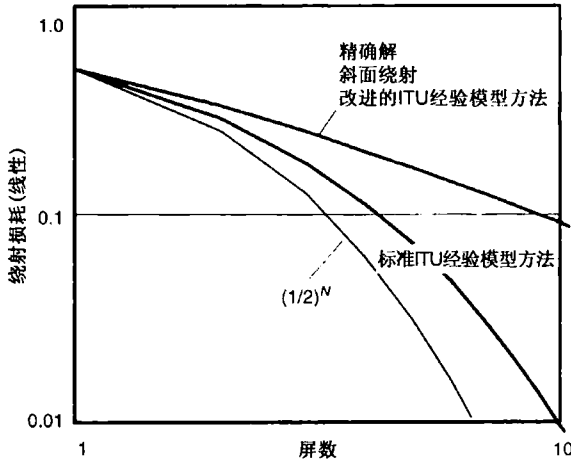


图 4.12 多屏绕射的不同计算方法的比较

然而，要注意上面的比较是基于一个特定的有限制的情形。对于少量不等高屏，Deygout 和 Epstein-Petersen 方法都可以成功地运用。

4.4 粗糙表面的散射

粗糙表面的散射(见图 4.13)对于无线通信来讲是一个非常重要的过程。散射理论通常假设粗糙程度是随机的。但是在无线通信中，通常也会把确定的或周期性的结构(如书架或窗沿)定义为粗糙的。对射线跟踪预测来说(见 7.5 节)，“粗糙”描述了所有(物理存在的)未包含在地图和建筑设计图中的物体。使用这种方法的理由有些启发式的：(i) 在射线跟踪预测中，产生的误差小于其他误差源；(ii) 没有更好的选择余地。

既然是这样，这一节余下的部分将会讨论真实粗糙表面的数学处理方法。过去 30 年中，人们在这一领域进行了大量的研究，主要由于它对雷达技术具有重要意义。由此发展出了两个主要理论：基尔霍夫(Kirchhoff)理论和微扰理论。

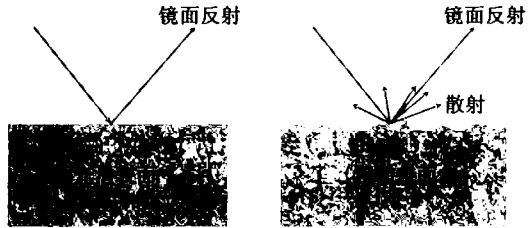


图 4.13 粗糙表面的散射

4.4.1 基尔霍夫理论

基尔霍夫理论的概念非常简单，并且只需很少量的信息，即平面振幅(高度)的概率密度函数。这个理论假设高度变化很小，以至于平面上的不同散射点并不会相互影响。换句话说，平面上的一点并不会给平面上的另一点“投下阴影”。这个假设在无线通信中实际上并不能很好地满足。

假设上面的条件实际上已经满足，表面粗糙导致镜面反射光线的功率减小，因为光线同时被散射到了其他方向(见图 4.13 的右边)。这种功率减小可以用有效反射系数 ρ_{rough} 来描述。在高斯高度分布状态的情况下，该反射系数变为

$$\rho_{\text{rough}} = \rho_{\text{smooth}} \exp \left[-2(k_0 \sigma_h \sin \psi)^2 \right] \tag{4.45}$$

其中, σ_h 是高度分布的标准偏差, k_0 是波数 $2\pi/\lambda$, ψ 是入射角(定义为波矢量和平面之间的夹角)。其中, $k_0\sigma_h\sin\psi$ 这一项也就是通常所说的瑞利粗糙。注意, 对于掠入射($\psi \approx 0$)来说, 粗糙的影响消失了, 反射又成为镜面反射。

4.4.2 微扰理论

微扰理论(perturbation theory)推广了基尔霍夫理论, 不仅使用了表面高度的概率密度函数, 还有它的空间相关函数。换句话说, 它考虑了这样一个问题: 当沿着表面移动某一距离时, 高度变化有多快(见图 4.14)?

数学上, 空间相关函数定义为

$$\sigma_h^2 W(\Delta_r) = E_r\{h(\mathbf{r})h(\mathbf{r} + \Delta_r)\} \quad (4.46)$$

其中, $\mathbf{r}$ 和 Δ_r 是(二维的)位置矢量, E_r 是关于 $\mathbf{r}$ 的期望值。需要利用这个信息来找出是否表面上的一点会在表面上的另一点“投下阴影”。如果

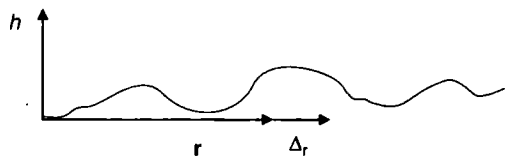


图 4.14 粗糙表面的微扰理论的几何表示

允许非常快速的振幅变化, 投影的情况就更加平常。上面的定义加强了空间统计量的平稳性, 即相关性独立于绝对位置 $\mathbf{r}$ 。相关长度 L_c 定义为使 $W(L_c) = 0.5 \cdot W(0)$ 的距离^①。

表面粗糙度对于镜面反射波振幅的影响可以用一个“有效”(复数的)介电常数 δ_{eff} 来描述, 该常数进而引出“有效”反射系数, 正如斯涅尔定律所计算的那样^②。对于垂直极化, δ_{eff} 表示如下 [Vaughan and Andersen 2003] :

$$\frac{1}{\sqrt{\delta_{r,\text{eff}}}} = \begin{cases} \frac{1}{\sqrt{\epsilon_r}} + j \frac{k_0 \sigma_h^2 \sin(2\psi)}{2L_c} \int_0^\infty \frac{1}{x} \frac{d\hat{W}(x)}{dx} dx, & k_0 L_c \ll 1 \\ \frac{1}{\sqrt{\epsilon_r}} + (k_0 \sigma_h)^2 (\sin \psi)^3, & k_0 L_c \gg 1, \psi \gg \frac{1}{\sqrt{k_0 L_c}} \\ \frac{1}{\sqrt{\epsilon_r}} - \frac{\sigma_h^2}{2L_c} \frac{\sqrt{k_0 L_c}}{\sqrt{2\pi}} \exp(j\pi/4) \int_0^\infty \frac{1}{x\sqrt{x}} \frac{d\hat{W}(x)}{dx} dx, & k_0 L_c \gg 1, \psi \ll \frac{1}{\sqrt{k_0 L_c}} \end{cases} \quad (4.47)$$

其中, $\hat{W}(x) = W(x/L_c)$ 。对于水平极化, 则为

$$\frac{1}{\sqrt{\delta_{r,\text{eff}}}} = \begin{cases} \frac{1}{\sqrt{\epsilon_r}} + j \frac{k_0 \sigma_h^2}{2L_c} \int_0^\infty \frac{1}{x} \frac{d\hat{W}(x)}{dx} dx, & k_0 L_c \ll 1 \\ \frac{1}{\sqrt{\epsilon_r}} + (k_0 \sigma_h)^2 \sin \psi, & k_0 L_c \gg 1, \psi \gg \frac{1}{\sqrt{k_0 L_c}} \\ \frac{1}{\sqrt{\epsilon_r}} - \frac{(k_0 \sigma_h)^2}{\sqrt{k_0 L_c}} \frac{2}{\sqrt{2\pi}} \exp(-j\pi/4) \int_0^\infty \frac{1}{\sqrt{x}} \frac{d\hat{W}(x)}{dx} dx, & \sqrt{k_0 L_c} \gg 1, \psi \ll \frac{1}{\sqrt{k_0 L_c}} \end{cases} \quad (4.48)$$

把这些结果与基尔霍夫理论相比, 可发现在 $k_0 L_c \gg 1$, $\psi \gg 1/\sqrt{k_0 L_c}$ 的情况下两者吻合得很好。这刚好符合了前面讨论的关于基尔霍夫理论的局限性: 假设相关长度大于波长, 表面上不可能有一个突然的“刺”引起绕射。并且, 只要满足 $\psi \gg 1/\sqrt{k_0 L_c}$, 就能保证一个小于 γ 的入射角的波不会在该表面的其他点上产生阴影。

① 数学期望值和相关函数的详细讨论将在第6章给出。

② 这里假设材料是不可导的, $\sigma_e = 0$ 。虚部的贡献由表面粗糙性产生。

4.5 波导

另一个重要过程是在(电介质)波导中的传播。这个过程模拟在街道峡谷、走廊、隧道中的传播。电介质波导的基本公式已经很完善地建立起来,参见 Collin [1991] 和 Marcuse [1991]。然而,那些发生在无线通信中的波导偏离了理论形式的理想化假设,原因如下。

- 材料是有损耗的。
- 街道峡谷(及大多数走廊)没有连续的墙,而是会被十字路口分割成或多或少的有秩序的间隔。而且,街道缺少波导的“上面的”墙面。
- 表面是粗糙的(窗沿等)。
- 波导管是非空的,而是充满了金属的(汽车)和非传导性的(步行者)相互作用体。

传播预测既可以通过计算波导模式来进行,也可以通过几何光学近似来进行。如果波导的横截面及内部的相互作用体比波长还长得多,后者的方法会给出好的结果[Klemenschits and Bonek 1994]。

传统的波导理论预测传播损耗随距离呈指数增长。一些在走廊的测量观察到类似的表现。然而,大多数测量都符合 d^{-n} 规则,其中 n 在 1.5 到 5 之间变化。注意,一个指数小于 2 的损耗不与能量守恒或其他物理定律相矛盾。自由空间传播的 d^{-2} 法则源于“随着距离的增加,能量将会扩展到更大的表面上”这一事实。如果可以引导能量,即使 d^0 在理论上也是有可能的。

4.6 附录

请访问 www.wiley.com/go/molisch。

深入阅读

如下几本关于无线传播过程的书籍给出了本章讨论的所有现象综述,包括 Bertoni [2000], Blaunstein [1999], Parsons [1992], Vaughan and Andersen [2003], Haslett [2008] 和 Barclay [2002]; 这些过程的 MATLAB 仿真程序可参考 Fontan and Espineira [2008]。基本的传播过程,特别是反射和透射,在许多电磁场的经典教材中都描述过,例如 Ramo et al. [1967], 这是本书作者本人最喜欢的,但还有很多其他教材。更复杂的结果可参考 Bowman et al. [1987] 和 Felsen and Marcuvitz [1973]。透过多层介电薄膜的传输的描述可以在 Heavens [1965] 中读到。绕射的几何理论最早由 Keller [1962] 提出,在 Kouyoumjian and Pathak [1974] 中扩展为统一的绕射理论,在 McNamara et al. [1990] 中给出总结。多重刀刃绕射的不同理论在 Parsons [1992] 中进行了很好的总结。然而,关于该主题的研究一直在持续进行,进一步的解决方案不断发表,如 Bergljung [1994]。粗糙表面的散射,多数受启发于雷达问题,在 Bass and Fuks [1979], de Santo and Brown [1986] 和 Ogilvie [1991] 中给出了描述,一个极好的总结可以见 Vaughan and Andersen [2003]。电介质波导的理论在 Collin [1991], Marcuse [1991] 和 Ramo et al. [1967] 中进行了描述。

第5章 无线信道的统计描述

5.1 概述

在许多环境中,要描述决定不同多径分量的所有的反射、绕射和散射是极其复杂的。通常,可取的方法是描述信道某一参数取某个值的概率。最重要的参数是信道增益,因为它决定接收功率和场强,它将是这一章要研究的中心。

为了达到对信道的更好理解,首先来看一个典型的描述接收功率和距离的函数关系的图形(见图5.1)。显然,接收功率可以有很大的变化,100 dB 或者更多。其次可注意到变化发生在不同的空间尺度上。

- 在距离很短的范围内,功率围绕一个(局部的)平均值上下波动,这一点从图5.1的小插图中可以看出。这些波动发生在大约一个波长的范围内,因此称为小尺度衰落。这些波动产生的原因是不同多径分量之间的干涉,如第2章所提到的。场强的波动可以很好地进行统计描述,即通过功率的(局部)平均值和围绕这个平均值的波动统计量来进行描述。

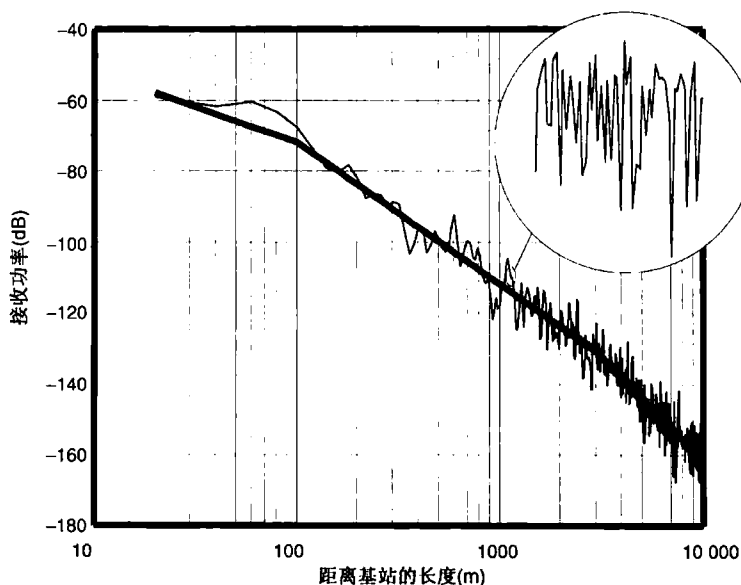


图5.1 接收功率与发射机、接收机间距的关系

- 平均功率,在大约10个波长上的平均,它本身也显示了波动性。这些波动发生在更大的空间距离上,典型地为几百个波长。当围绕着发射机移动一个圆环时,这种变化可以看得很清楚(见图5.2)。引起这些变化的原因是大型物体的阴影效应(见第2章),这与引起小尺度衰落的干涉在本质上是不同的。但是,这种大尺度衰落也同样可以用平均值和围绕这个平均值的波动统计量来描述。

- 大尺度平均值本身单调地依赖于发射机和接收机之间的距离。这种效应与自由空间路径损耗及其变化形式有关。这种效应通常采用确定的方式描述,并且已经在第4章中讨论过(更多细节和模型参见第7章)。

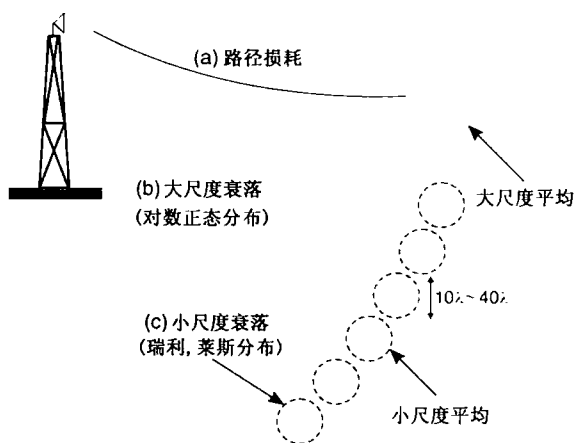


图 5.2 接收功率变化的类型

第5章仅关注信道增益的统计变化,时延色散及其他效应将在以后讨论。对这一章来说,考虑一个非调制的(正弦)载波信号的信道增益就足够了,虽然这种考虑同样也适用于窄带系统(如6.1节所定义的)。5.2节和5.3节解释两径模型,即解释小尺度衰落效应时最简单的模型。5.4节和5.5节把这些考虑因素推广到更实际的信道模型中,并且给出了各种情形下接收场强的统计量。随后两节描述衰落深陷(Fading dip,接收功率非常低的情形)的统计量,它们发生的频率及其平均延续时间。最后,由于阴影效应所导致的统计变化在5.8节中讨论。

5.2 时不变两径模型

作为“多径传播和衰落”这一复杂主题的入门,这里考察可能的最简单情况,即沿两条路径的时不变传播。发送一个正弦波形,来确定在接收机位置处的(复数)传输函数。

首先,考察一个单一的波形。设发送信号为正弦波:

$$E_{TX}(t) \propto \cos(2\pi f_c t) \quad (5.1)$$

让接收信号近似为一个同类的平面波。如果发射机和接收机之间的路径长度为 d ,则接收信号可以表示为

$$E(t) = E_0 \cdot \cos(2\pi f_c t - k_0 d) \quad (5.2)$$

其中,符号 k_0 是波数 $2\pi/\lambda$ 。使用复基带符号^①,可以写成

$$E = E_0 \exp(-jk_0 d) \quad (5.3)$$

注意,在用复数表示时,场的实部 $\text{Re}\{E\}$ 就等于场强在时间 $t=0$ 时的瞬时值。

现在考虑发送信号经过两条不同传输路径到达接收机的情形,由两个不同的相互作用体(10,见图5.3)引起。这些路径有不同的传播时间:

$$\tau_1 = d_1/c_0, \quad \tau_2 = d_2/c_0 \quad (5.4)$$

① 带通信号,即物理上存在的信号,与复基带(低通)表示的关系为 $s_{BP}(t) = \text{Re}\{s_{LP}(t) \exp[j2\pi f_c t]\}$ (见11.1节)。

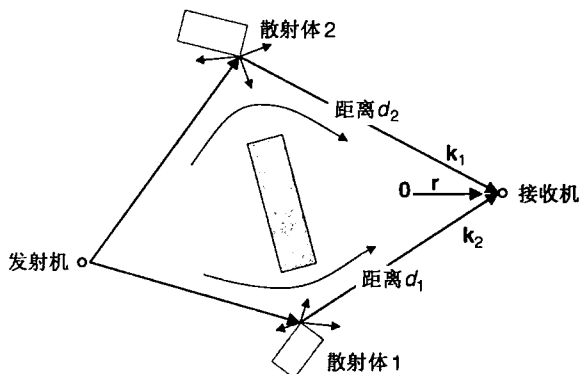


图 5.3 时不变两径模型的几何结构

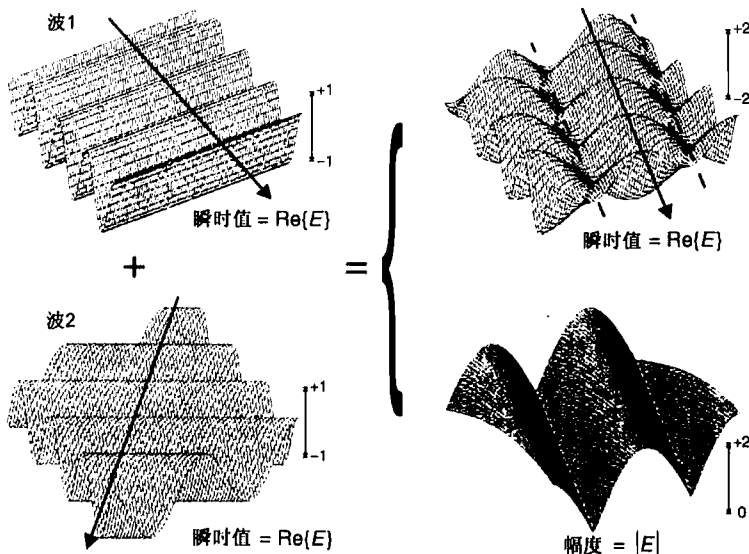
接收机在相互作用体的远场，因此到达波是均匀的平面波。进一步假设两路波都是垂直极化的，在参考位置 $\mathbf{r}=0$ (坐标系原点) 处的振幅为 E_1 和 E_2 。得到两个平面波叠加的表达式如下：

$$E(\mathbf{r}) = E_1 \exp(-j\mathbf{k}_1\mathbf{r}) + E_2 \exp(-j\mathbf{k}_2\mathbf{r}) \quad (5.5)$$

其中， k_i 是矢量值的波数 (即绝对值为 k_0 ，指向第 i 路波的方向)。

这里假设到达接收机位置 $\mathbf{r}$ 处的是两个绝对振幅不随接收位置而变化的平面波 (仅仅在小于 10λ 直径的范围内变化接收机的位置)。

图 5.4 的上半部分描绘了 $E(\mathbf{r})$ 的实部，下半部分显示了振幅，振幅与接收功率的平方根成正比。下面看看相长和相消干涉的位置，即位置取决于衰落。在由于不同时延而引起的相差为 180° 的位置处，存在衰落深陷。如果接收机处于相消干涉的位置处，可以看到一个总振幅为组成波振幅之差的信号。如果组成波振幅相等，相消干涉就能很彻底。在相长干涉的位置处，总的振幅是组成波振幅之和。这个现象可以在图 5.4 的右下角看到。

图 5.4 两个平面波的干涉，其中 $E_1 = E_2 = 1$, $\arg(\mathbf{k}_1, \mathbf{k}_2) = 30^\circ$

5.3 时变两径模型

通常,不同传播路径之间的时延(路径长度)差异随时间变化。这种变化可以由发射机、接收机、相互作用体的移动及其组合而引起。为了简化讨论,今后将假设只有接收机移动。于是接收机可以“看到”一个时变干涉模式,假设接收机沿着场强图上的“山峰和沟壑”移动。空间变化衰落成了时间变化衰落。因为衰落深陷的间隔大约为半个波长(当载波频率为 900 MHz 时对应为 16 cm),所以这种衰落称为小尺度衰落,也称为短时衰落或者快衰落^①。衰落率(每秒内的衰落深陷的数目)取决于接收机的移动速度。

接收机的移动也引起了接收频率的偏移,称为多普勒频移。为了解释这一现象,首先回顾单一正弦波到达接收机的情形,并考虑实的带通表示法。如果接收机以速度 v 远离发射机,那么两者之间的距离 d 随着该速度增长,则有

$$\begin{aligned} E(t) &= E_0 \cdot \cos(2\pi f_c t - k_0[d_0 + vt]) \\ &= E_0 \cdot \cos\left(2\pi t \left[f_c - \frac{v}{\lambda}\right] - k_0 d_0\right) \end{aligned} \quad (5.6)$$

其中, d_0 是 $t=0$ 时刻的距离,接收机振荡的频率因此减少了 v/λ 。换句话说,多普勒频移为

$$\nu = -\frac{v}{\lambda} = -f_c \cdot \frac{v}{c_0} \quad (5.7)$$

注意,在发射机和接收机相互远离时,多普勒频移是负值。由于移动的速度总是比光速小,所以多普勒频移相对也较小。

在上面的例子中,已经假设发射机的运动方向与波传播的方向成一条直线。如果不是这一情况,多普勒频移就由波传播方向的那个速度分量来决定,即分量 $v \cos(\gamma)$ (见图 5.5)。多普勒频移变为

$$\nu = -\frac{v}{\lambda} \cos(\gamma) = -f_c \cdot \frac{v}{c_0} \cos(\gamma) = -v_{\max} \cos(\gamma) \quad (5.8)$$

最大多普勒频移 $v_{\max}$ 的典型值为 1 Hz ~ 1 kHz。注意,通常关系式 $v_{\max} = f_c \cdot v/c_0$ 是基于几种假设的,如静止相互作用体,移动物体上没有双反射,等等。

既然多普勒频移这么小,那么很自然会问到它们会不会对无线链路产生显著影响。如果所有组成波有相同的多普勒频移,如 100 Hz,那么它对无线链路的影响实际上可以忽略,因为接收机的本地振动可以很容易地补偿这个频率偏移。重要的一点是,不同的多径分量有不同的多普勒频移。具有多普勒频移的几个波叠加产生了衰落深陷序列。再者,这可以用两径模型来演示。随着接收机的移动,它接收到的两个波形都产生了多普勒频移,只是偏移量不同。

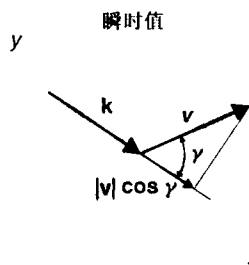


图 5.5 速度矢量 v 投影到传播方向 k 上

① 遗憾的是,快衰落经常用于表示两种完全不同的现象。一方面,它是小尺度衰落的同义词,独立于实际衰落的时间尺度(其取决于发射机、接收机和相互作用体的运动速度)。另一方面,它用于表示信道在一个符号长度持续时间内的变化(与“准静态”信道相对比)。这样,我们更愿意称干涉效应引起的衰落为小尺度衰落,因为它含义明确。

通过傅里叶变换到时域,得到了两个频率有轻微差别的振动引起的波拍效应(见图 5.6)。差拍包络的频率等于两个载波的频率差,即两个多普勒频移的差异。这样接收机看到一个振幅周期变化的信号,这正是在图 5.4 所示场强图中的“山峰和沟壑”上移动而产生的快衰落。

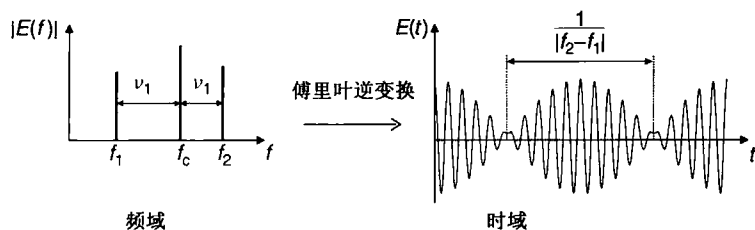


图 5.6 两个不同频率的载波相叠加(差拍)

总而言之,可以通过两个等价的因素来获得两径模型的衰落率。

1. 叠加两个入射波。绘出产生的干涉模式(场强的“山峰和沟壑”),数出当接收机沿着该模式移动时看到的每秒钟的衰落深陷的数目。
2. 另一种选择是考虑在接收天线处两个具有不同的多普勒频移的信号叠加,从波拍频率来确定衰落率,即两个波的多普勒频移的差异。

多普勒频率是信道的一个重要参数,即使它的数值很小。

- 多普勒频率是信道变化速率的一个度量,正如上面所讨论的。
- 很多轻微的多普勒频移信号的叠加导致了总的接收信号的相移,而且这一相移可以破坏角度调制信号的接收(见第 11 章和第 12 章)。这些相移导致了接收信号的随机频率调制(FM)(见 5.7 节),并且对于低比特率的信号尤为重要。

5.4 不含主导分量的小尺度衰落

遵循这些利用两径模型的基本考虑,现在研究多径传播的更一般性的情况。考虑有很多相互作用体和一个移动发射机的无线信道。由于相互作用体的数目比较大,对无线信道的确定性描述已经不再有效,这也是求助于统计描述方法的原因。这种统计描述对于无线通信的整个领域都是必不可少的,因而这里进行相当详细的解释。首先以 5.4.1 节的一个计算机实验开始,接着在 5.4.2 节给出一种更一般的数学推导。

5.4.1 一个计算机实验

考虑如下简单的计算机实验。信号从几个相互作用体射向在小范围内移动的接收机。相互作用体近似均匀分布在接收区域的周围。假设它们的距离足够远,从而使所有接收波都是均匀平面波,并且接收机在考察范围内移动时不会改变这些波的振幅。通过给每个波分配一个随机相位和随机振幅,来考虑相互影响的不同距离和强度。这里产生了 8 个组成波 E_i , 其绝对相位为 $|a_i|$, 入射角(关于 x 轴)为 ϕ_i , 相位为 φ_i 。

	$ a_i $	ϕ_i	φ_i
$E_1(x, y) = 1.0 \exp[-jk_0(x \cos(169^\circ) + y \sin(169^\circ))] \exp(j311^\circ)$	1.0	169	311
$E_2(x, y) = 0.8 \exp[-jk_0(x \cos(213^\circ) + y \sin(213^\circ))] \exp(j32^\circ)$	0.8	213	32
$E_3(x, y) = 1.1 \exp[-jk_0(x \cos(87^\circ) + y \sin(87^\circ))] \exp(j161^\circ)$	1.1	87	161
$E_4(x, y) = 1.3 \exp[-jk_0(x \cos(256^\circ) + y \sin(256^\circ))] \exp(j356^\circ)$	1.3	256	356
$E_5(x, y) = 0.9 \exp[-jk_0(x \cos(17^\circ) + y \sin(17^\circ))] \exp(j191^\circ)$	0.9	17	191
$E_6(x, y) = 0.5 \exp[-jk_0(x \cos(126^\circ) + y \sin(126^\circ))] \exp(j56^\circ)$	0.5	126	56
$E_7(x, y) = 0.7 \exp[-jk_0(x \cos(343^\circ) + y \sin(343^\circ))] \exp(j268^\circ)$	0.7	343	268
$E_8(x, y) = 0.9 \exp[-jk_0(x \cos(297^\circ) + y \sin(297^\circ))] \exp(j131^\circ)$	0.9	297	131

现在叠加这些组成波，再次使用复数基带表示。总的复数场强 E 就是各个组成波复数场强之和。也可以把它理解为随机相位复矢量之和。图 5.7 显示了在大小为 $5\lambda \times 5\lambda$ 的区域内 $t=0$ 时刻总场强的瞬时值，即 $\text{Re}\{E\}$ 。

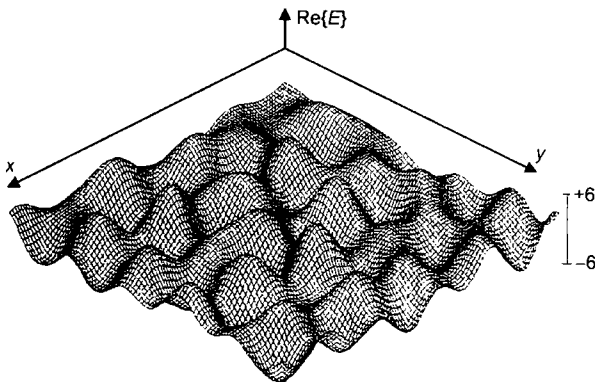


图 5.7 场强在 $t=0$ 时刻的瞬时值，即 $\text{Re}\{E\}$ 。8 个组成波在区域 $0 < x < 5\lambda, 0 < y < 5\lambda$ 的叠加

现在考虑发生在这一区域内的场强的统计量。从图 5.8 的直方图可看出， $\text{Re}\{E\}$ 的值很好地近似于零均值高斯分布。这是中心极限定律的结果：叠加 N 个统计独立的随机变量，其中没有一个占主导地位，与之相关的概率密度函数当 $N \rightarrow \infty$ 时趋于正态分布，更精确的表述参见附录 5. A (www.wiley.com/go/molisch)。中心极限定律的有效性条件近似地满足：8 个组成波有随机的入射角和相位，并且没有一个占主导地位。图 5.9 和图 5.10 示出了场强的虚部， $\text{Im}\{E\}$ ，也是正态分布的^①。

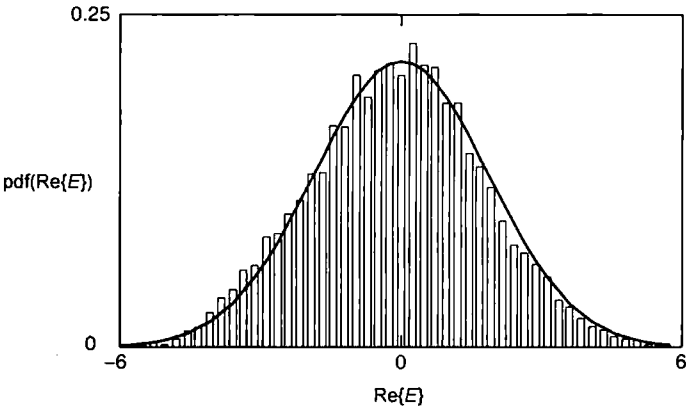


图 5.8 图 5.7 的场强的直方图。高斯概率密度函数用于进行对比

① 虚部表示在对应于无线频率振荡的 1/4 周期的时刻，即 $\omega t = \pi/2$ 时，场强的瞬时值。

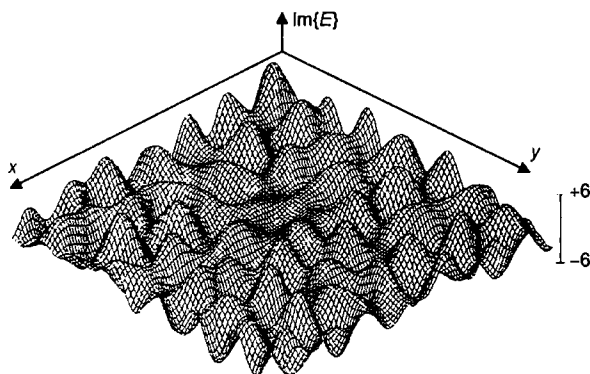


图 5.9 E 的虚部

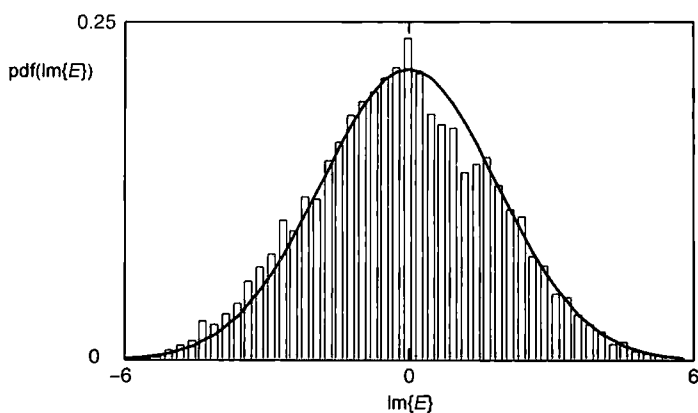


图 5.10 图 5.9 的场强的直方图。高斯概率密度函数用于进行对比

大多数接收机的性能取决于接收(绝对)振幅(幅度)。因而有必要研究接收信号包络的分布,对应于(复数的)场强相位矢量的幅度。图 5.11 显示了接收机沿着图 5.7 或图 5.8 中的 y 轴移动时的场强。图 5.12 的左侧显示了复数场强相位矢量,右侧是接收信号的绝对振幅。

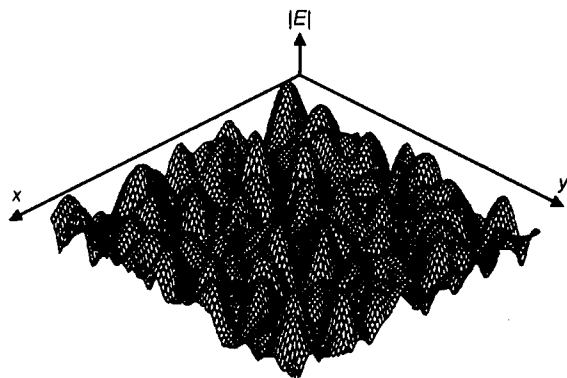


图 5.11 场强的振幅

图 5.11 和图 5.13 分别是振幅的三维表示及该区域振幅的统计量。图 5.13 也展示了瑞利概率密度函数的图形。瑞利分布描述了一个实部和虚部都是正态分布的复数随机变量的幅度,更多的细节在下面给出。图 5.14 说明了相位近似地服从均匀分布。

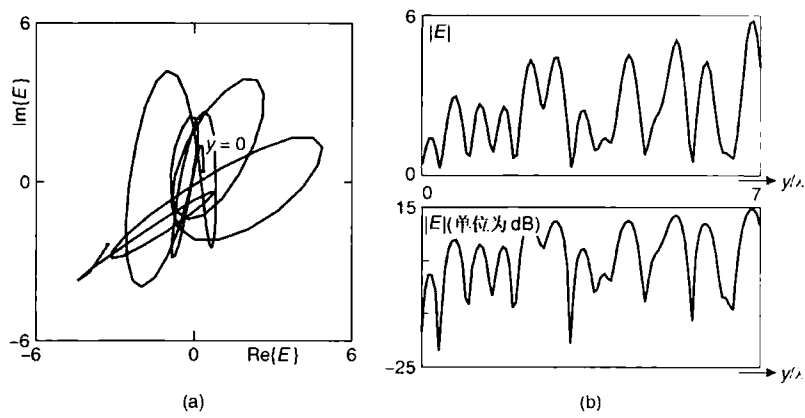


图 5.12 场强的复相位矢量(a)和接收机沿 y 轴移动时的接收振幅 $|E|$ (b)

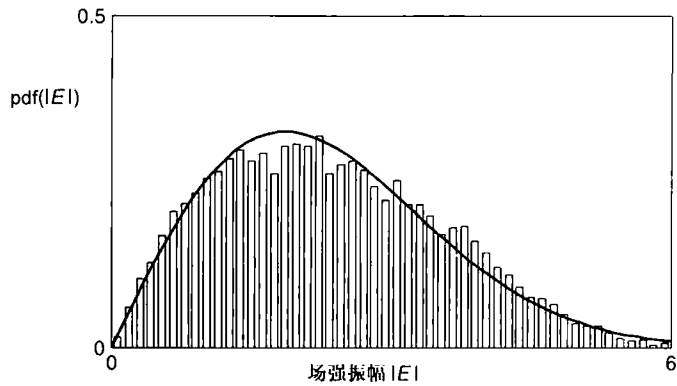


图 5.13 接收振幅的概率密度函数

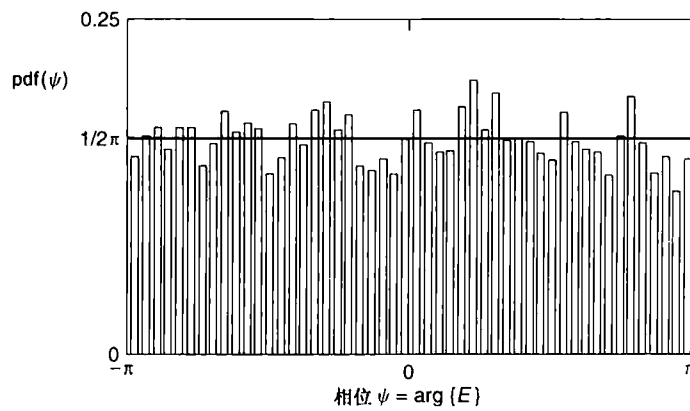


图 5.14 接收相位的概率密度函数

5.4.2 振幅与相位统计量的数学推导

在这些人为设计的实验之后，转向一个更详细并从数学意义上讲更彻底的瑞利分布的推导。考虑这样一种情形：由不同的相互作用体反射/散射而产生了 N 个均匀平面波(多径分量)。相互作用体和发射机静止不动，接收机以速度 v 移动。与前面一样，假设多径分量的绝对振幅在观察区域内不变。

这样, 振幅平方和为

$$\sum_{i=1}^N |a_i|^2 = C_P \quad (5.9)$$

其中, C_P 是常量, 然而相位 φ_i 变化很强烈, 可以近似为在 $[0, 2\pi]$ 均匀分布的随机变量。由第 i 个多径分量引起的接收机场强的实部就是 $|a_i| \cos(\varphi_i)$, 虚部是 $|a_i| \sin(\varphi_i)$ 。

而且, 为了计算总场强 $E(t)$, 要考虑到多普勒频移。如果考虑非调制载波, 则可得(采用实带通表示)

$$E(t) = \sum_{i=1}^N |a_i| \cos[2\pi f_c t - 2\pi v_{\max} \cos(\gamma_i) t + \varphi_i] \quad (5.10)$$

将该式用实带通符号的同相分量和正交分量来表示, 可得

$$E_{BP}(t) = I(t) \cdot \cos(2\pi f_c t) - Q(t) \cdot \sin(2\pi f_c t) \quad (5.11)$$

其中,

$$I(t) = \sum_{i=1}^N |a_i| \cos[-2\pi v_{\max} \cos(\gamma_i) t + \varphi_i] \quad (5.12)$$

$$Q(t) = \sum_{i=1}^N |a_i| \sin[-2\pi v_{\max} \cos(\gamma_i) t + \varphi_i] \quad (5.13)$$

$I(t)$ 和 $Q(t)$ 都是许多随机变量之和, 其中没有一个占主导地位(即 $|a_i| \ll C_P$)。遵循中心极限定律, 这种变量之和的概率密度函数是正态(高斯)分布, 与组成波振幅的具体概率密度函数无关, 即不需要 a_i 的知识或它们的分布!!! 更详细信息参见附录 5. A (www.wiley.com/go/molisch)。一个零均值的高斯随机变量的概率密度函数为

$$\text{pdf}_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (5.14)$$

其中, σ^2 表示方差。

由实部和虚部的统计量出发, 附录 5. B 推导出了接收信号振幅和相位的统计量。得出了 ψ 的概率密度函数, 即一个均匀分布:

$$\text{pdf}_\psi(\psi) = \frac{1}{2\pi} \quad (5.15)$$

以及 r 的概率密度函数, 即一个瑞利分布:

$$\text{pdf}_r(r) = \frac{r}{\sigma^2} \cdot \exp\left[-\frac{r^2}{2\sigma^2}\right], \quad 0 \leq r < \infty \quad (5.16)$$

$r < 0$ 时的概率密度为零, 因为振幅只能取正数。

5.4.3 瑞利分布的性质

一个瑞利分布有如下性质, 这些性质同时在图 5.15 中给出:

$$\left. \begin{aligned} \text{均值 } \bar{r} &= \sigma \sqrt{\frac{\pi}{2}} \\ \text{均方差 } \overline{r^2} &= 2\sigma^2 \\ \text{方差 } \overline{r^2} - (\bar{r})^2 &= 2\sigma^2 - \sigma^2 \frac{\pi}{2} = 0.429\sigma^2 \\ \text{中值 } r_{50} &= \sigma \sqrt{2 \cdot \ln 2} = 1.18\sigma \\ \text{最大值 } \max\{\text{pdf}(r)\} &\text{ 发生在 } r = \sigma \text{ 时} \end{aligned} \right\} \quad (5.17)$$

其中，字母上的横线代表期望值，不采用通常的符号 $E\{\cdot\}$ ，是为了避免与场强 E 相混淆。

累积分布函数 $\text{cdf}(x)$ 定义为随机变量小于 x 的概率。因而累积分布函数是概率密度函数的积分：

$$\text{cdf}(r) = \int_{-\infty}^r \text{pdf}(u) du \tag{5.18}$$

对瑞利概率密度函数应用这个公式，可得

$$\text{cdf}(r) = 1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{5.19}$$

当 r 值很小时，上式可以近似为

$$\text{cdf}(r) \approx \frac{r^2}{2\sigma^2} \tag{5.20}$$

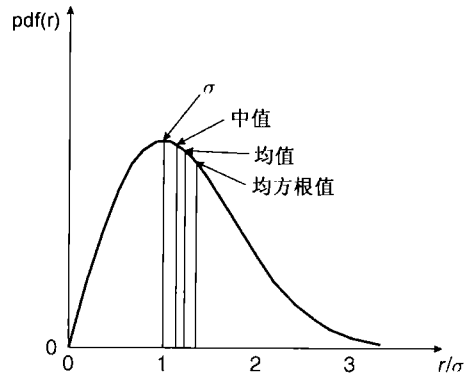


图 5.15 瑞利分布的概率密度函数

检查测量的场强值是否整体服从瑞利分布是很直接的：实验的累积概率分布值画在所谓的 Weibull 纸上(见图 5.16)。瑞利分布的累积概率分布是图上的那条直线。 r 值很小时， r 值增加 10 dB，则累积分布函数的值也增加 10 dB。通过变量变换，很容易看到平方根幅度，通过扩展得到的功率服从指数分布，其概率密度函数 $\text{pdf}_p(P) = \frac{1}{\bar{\Omega}} \exp(-P/\bar{\Omega})$ ，其中 $\bar{\Omega}$ 为平均功率^①。

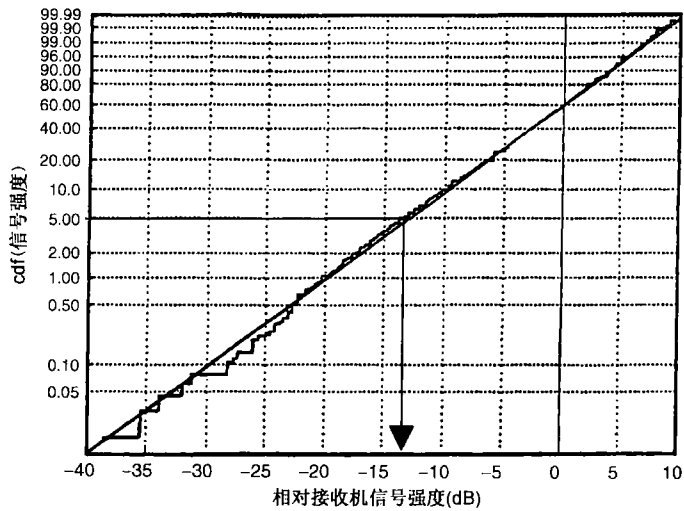


图 5.16 一个室内无视距情况下的(正态)接收功率的测量累积概率分布 [Gahleitner 1993]

瑞利分布广泛运用在无线通信中，有以下几方面原因。

- 在许多实际应用的情形中，它都是一个极好的近似，这已经被大量测量所证实。然而，值得注意的是仍然存在不适用的情形。这种情况可以发生在比如视距的情况下、一些室内的情况下和一些(超)宽带的情况下(见第 6 章和第 7 章)。
- 它描述了在没有主导信号分量的前提下的最差情形，因而存在大量的衰落深陷。这种最差情形假设对于设计一个稳健的系统很有用^②。

① 为简化讨论，本章不区分平方根幅度与功率，虽然它们实际上是通过一个比例常数相关的。
② 正如后面将看到的，存在一些衰落分布表现出更多的衰落深陷，例如 $m < 1$ 时的 Nakagami 分布。而且，对于某些特定系统而言，大量的多径分量也可能是一种优势(见 20.2 节)。

- 它仅取决于一个单一的参数, 平均接收功率。一旦知道了这个参数, 整个信号的统计特性就知道了。用测量或者确定的预测方法获得这一单一参数比在其他复杂的信道模型中获得多个参数容易得多, 出错的机会也少。
- 便于以数学方式实现。当场强分布为瑞利分布时, 差错概率和其他参数的计算通常可以用闭式完成。

5.4.4 瑞利分布场强的衰落余量

衰落统计量的知识对于无线系统的设计极其重要。在第3章中可看到, 对于噪声受限信号, 接收场强决定了性能。因为场强是随机变量, 即使是大的场强均值也不能保证所有时刻成功地通信。相反地, 仅在一定百分比的情况下, 场强才超过最小值。因此, 现在的任务就变成回答这样一个问题: “给定成功通信所需的最小功率或者场强, 平均功率有多大才能保证通信在所有情况下有不超过 $x\%$ 的失败率?” 换句话说, 衰落余量要留多大?

根据定义, 累积分布函数给出了某一场强电平不会被超过的概率。为了达到 $x\%$ 的中断概率, 可以确定:

$$x = \text{cdf}(r_{\min}) \approx \frac{r_{\min}^2}{2\sigma^2} \quad (5.21)$$

其中右边遵从式(5.20)。这样就能立即计算出场强的均方值 $2\sigma^2$, 因为 $2\sigma^2 = r_{\min}^2/x$ 。

例 5.1 对于一个振幅为瑞利分布的信号, 接收信号功率比平均功率至少低 20 dB, 6 dB 和 3 dB 的概率是多少? 比较精确结果和用近似公式(5.20)得出的结果。

解:

从瑞利信号包络 r :

$$\overline{r^2} = 2\sigma^2 \quad (5.22)$$

功率水平低于平均功率 20 dB 对应于 $\frac{r_{\min}^2}{2\sigma^2} = \frac{1}{100}$:

$$\Pr\{r < r_{\min}\} = 1 - \exp\left(-\frac{1}{100}\right) = 9.95 \times 10^{-3} \quad (5.23)$$

类似地, 6 dB 和 3 dB 时的精确结果分别为 0.221 和 0.393。

近似公式(5.20)分别给出 $\frac{r_{\min}^2}{2\sigma^2} = \frac{1}{100} = 0.01, 0.25$ 和 0.5 。因此这对于低于平均功率 6 dB 的功率水平是相当精确的, 但对于 $r_{\min}$ 的更高值就不准确了。

对于干扰受限的情形, 情况变得有点更复杂了: 不仅有有用信号有衰落, 干扰源也有衰落。为了计算信号和干扰幅度比的统计特性, 注意期望信号和干扰都是瑞利衰落的, 因此需要两个随机变量之比的概率密度函数, 其中每一个都是瑞利分布的 [Molisch et al. 1996] :

$$\text{pdf}(r) = \frac{2\tilde{\sigma}^2 r}{(\tilde{\sigma}^2 + r^2)^2} \quad (5.24)$$

其中, $\tilde{\sigma}^2 = \sigma_1^2/\sigma_2^2$ 是平均信号功率与平均干扰功率之比。相应的累积概率密度是

$$\text{cdf}(r) = 1 - \frac{\bar{\sigma}^2}{(\bar{\sigma}^2 + r^2)}$$

(5.25)

这个公式在计算重用距离时是必要的(见第 3 章和第 17 章)。

5.5 含有主导分量的小尺度衰落

5.5.1 一个计算机实验

当一个主导的多径分量(如一个视距分量或一个主导的镜面反射分量)存在时,衰落统计量会发生变化。重复 5.4.1 节的计算机实验,可以获得一些启示,但是现在加上一个(主导的)振幅为 5 的附加波:

	$ a_0 $	ϕ_0	φ_0
$E_0(x, y) = 5.0 \exp[-jk_0(x \cos(0^\circ) + y \sin(0^\circ))] \exp(j0^\circ)$	5.0	0°	0°

图 5.17 给出了 E 的实部,主导分量带来的贡献很明显,图 5.18 给出了 E 的绝对值。场强绝对值的直方图参见图 5.19。很明显,深衰落的概率比瑞利衰落情形中小得多。

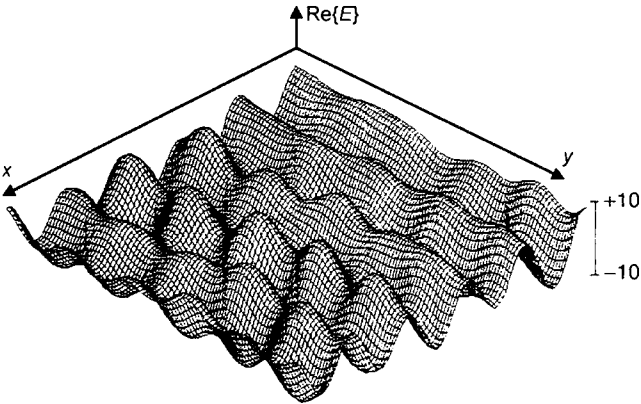


图 5.17 $\text{Re}\{E\}$, 即在一个主导的多径分量的存在情况下, $t=0$ 时刻的瞬时值

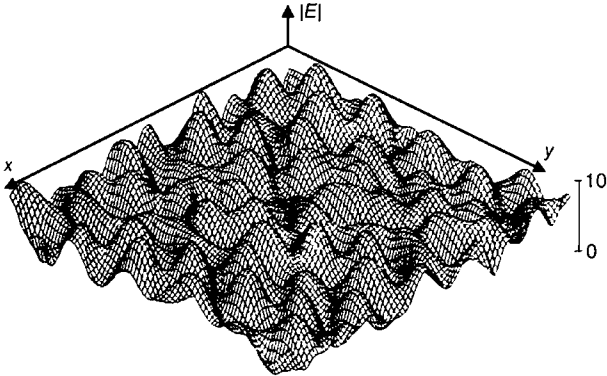


图 5.18 在例子的区域内, 场强的幅度 $|E|$, 即在一个主导的多径分量的存在情况下, $t=0$ 时刻的瞬时值

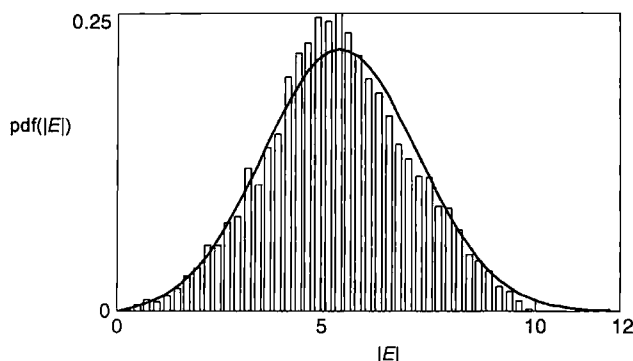


图 5.19 一个主导的多径分量的存在情况下振幅的直方图

5.5.2 振幅和相位分布的推导

振幅的概率密度函数可用类似于推导瑞利分布的方法来计算, 参见附录 5. B (www.wiley.com/go/molisch)。没有一般性的限制, 假设视距分量为零相位, 所以它是纯实数。因而实部是一个非零均值的高斯分布, 但虚部是一个零均值的高斯分布。执行附录 5. B 中给出的变量变换, 得到了振幅 r 和相位 ψ 的联合概率密度函数 [Rice 1947] :

$$\text{pdf}_{r,\psi}(r, \psi) = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2 + A^2 - 2rA\cos(\psi)}{2\sigma^2}\right) \quad (5.26)$$

其中, A 是主导分量的振幅。与瑞利情形相反, 这种分布是不可分离的。实际上, 必须对相位积分而得到振幅的概率密度函数, 反之亦然。

振幅的概率密度函数由莱斯分布给出 (图 5.19 中的实线):

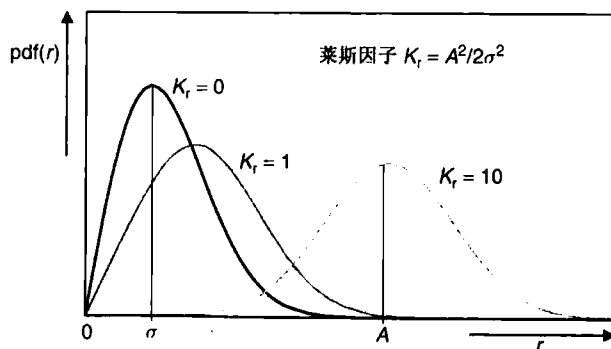
$$\text{pdf}_r(r) = \frac{r}{\sigma^2} \cdot \exp\left[-\frac{r^2 + A^2}{2\sigma^2}\right] \cdot I_0\left(\frac{rA}{\sigma^2}\right), \quad 0 \leq r < \infty \quad (5.27)$$

$I_0(x)$ 是第一类修正贝塞尔函数, 阶数为 0 [Abramowitz and Stegun 1965]。莱斯分布随机变量的均方值是

$$\overline{r^2} = 2\sigma^2 + A^2 \quad (5.28)$$

视距分量功率与漫射分量功率之比 $A^2/(2\sigma^2)$, 称为莱斯因子 K_r 。

图 5.20 给出了莱斯因子取 3 个不同值时的莱斯分布。视距分量越强, 发生深衰落的机会越少。当 K_r 趋于零时, 莱斯分布变为瑞利分布, 但对于大的 K_r 则近似成均值为 A 的高斯分布。

图 5.20 K_r (即视距分量功率与漫射分量功率之比) 取 3 个不同值时的莱斯分布

例 5.2 分别计算 $K_r=0.3$ dB, 3 dB 和 20 dB 时, 使中断概率小于 5% 的莱斯分布衰落余量。

解:

回顾前述内容, 中断概率可以用莱斯包络的累积分布函数表示:

$$P_{\text{out}} = \text{cdf}(r_{\min}) \tag{5.29}$$

对于莱斯分布, 累积分布函数如下:

$$\begin{aligned} \text{cdf}(r_{\min}) &= \int_0^{r_{\min}} \frac{r}{\sigma^2} \cdot \exp\left[-\frac{r^2 + A^2}{2\sigma^2}\right] \cdot I_0\left(\frac{rA}{\sigma^2}\right) dr, \quad 0 \leq r < \infty \\ &= 1 - Q_M\left(\frac{A}{\sigma}, \frac{r_{\min}}{\sigma}\right) \end{aligned} \tag{5.30}$$

其中, $Q_M(a, b)$ 是 Marcum Q 函数(见第 12 章), 即

$$Q_M(a, b) = e^{-(a^2+b^2)/2} \sum_{n=0}^{\infty} \left(\frac{a}{b}\right)^n I_n(ab) \tag{5.31}$$

$I_n(\cdot)$ 是 n 阶第一类修正贝塞尔函数。衰落余量为

$$\frac{\overline{r^2}}{r_{\min}^2} = \frac{2\sigma^2(1 + K_r)}{r_{\min}^2} \tag{5.32}$$

莱斯功率累积分布函数如图 5.21 所示, 为归一化包络的函数。不同的 K_r 所需的衰落余量在图中为 11.5 dB, 9.7 dB 和 1.1 dB, 分别对应于莱斯因子 0.3 dB, 3 dB 和 20 dB。

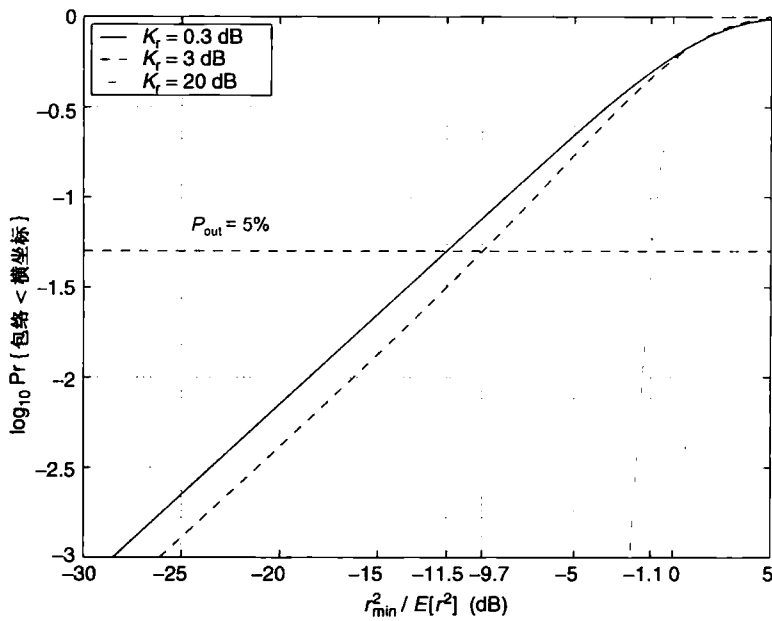


图 5.21 莱斯功率累积分布函数, $\sigma = 1$

主导分量的出现同样也改变了相位分布。设想一个非常强的主导分量, 总的信号相位一定非常接近主导分量的相位。换句话说, 相位分布收敛于 δ 函数, 这一点很直观。对于一般情况(已定义视距分量的相位 $\psi = 0$), 相位的概率密度函数可以由 r 和 ψ 的联合概率密度函数求得, 变为[Lustmann and Porrat 2010]

$$\text{pdf}(\psi) = \frac{1 + \sqrt{\pi K_r} e^{K_r \cos^2(\psi)} \cos(\psi) (1 + \text{erf}[\sqrt{K_r} \cos(\psi)])}{2\pi e^{K_r}} \quad (5.33)$$

其中 $\text{erf}(x)$ 是误差函数 [Abramowitz and Stegun 1965], 即

$$\text{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-t^2) dt$$

图 5.22 给出了 $\sigma = 1$, A 取不同值时的相位分布。

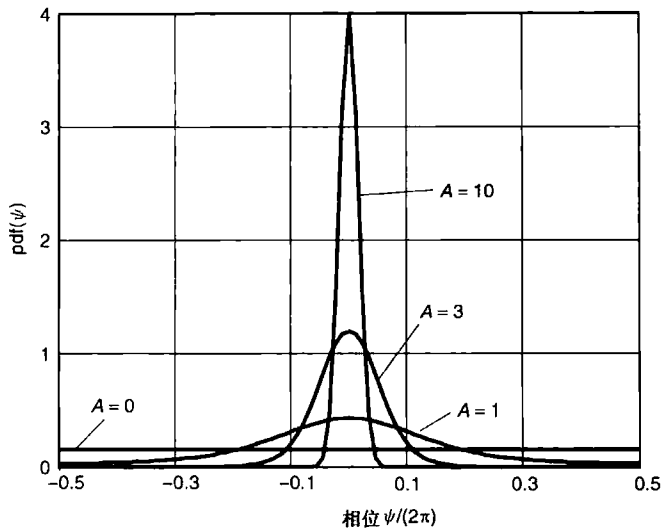


图 5.22 非零均值复高斯分布的概率密度函数, $\sigma = 1$, $A = 0, 1, 3, 10$

功率的概率密度函数为

$$\text{pdf}_P(P) = \frac{1 + K_r}{\bar{\Omega}} \exp\left(-K_r - \frac{(K_r + 1)P}{\bar{\Omega}}\right) I_0\left(2\sqrt{\frac{K_r(K_r + 1)P}{\bar{\Omega}}}\right), \quad P \geq 0 \quad (5.34)$$

从历史的角度看, 有意思的是, 有关莱斯分布的工作都是在未考虑无线信道的前提下进行的。经典的莱斯论文 [Rice 1947] 考虑了正弦波在加性高斯噪声下传播的问题。然而, 从数学的角度看, 这仅仅是一个确定的相位复矢量 (引起非零均值) 加上一个零均值复高斯分布的问题, 几乎与场强计算中的问题相同。现有的结果需要无线工程师来重新解释。这个事实令人感兴趣, 因为可能有其他无线问题也能通过这种“重新解释”的方法来解决。

5.5.3 Nakagami 分布

另一个广泛应用的场强概率分布是 Nakagami m 分布, 其概率密度函数如下:

$$\text{pdf}_r(r) = \frac{2}{\Gamma(m)} \left(\frac{m}{\bar{\Omega}}\right)^m r^{2m-1} \exp\left(-\frac{m}{\bar{\Omega}} r^2\right) \quad (5.35)$$

其中 $r \geq 0$, $m \geq 1/2$; $\Gamma(m)$ 是欧拉的伽马函数 [Abramowitz and Stegun 1965]。参数 $\bar{\Omega}$ 是均方值, $\bar{\Omega} = \overline{r^2}$, 参数 m 为

$$m = \frac{\bar{\Omega}^2}{(r^2 - \bar{\Omega})^2} \quad (5.36)$$

可以从测量值中直接提取这些参数。如果振幅是 Nakagami 衰落, 那么功率服从伽马分布:

$$\text{pdf}_P(P) = \frac{m}{\bar{\Omega}\Gamma(m)} \left(\frac{mP}{\bar{\Omega}}\right)^{m-1} \exp\left(-\frac{mP}{\bar{\Omega}}\right) \quad (5.37)$$

Nakagami 分布和莱斯分布的形状非常相似, 其中一个可以用来近似另一个。对于 $m > 1$, m 因子可以用 K_r 算出 [Stueber 1996] :

$$m = \frac{(K_r + 1)^2}{(2K_r + 1)} \quad (5.38)$$

而

$$K_r = \frac{\sqrt{m^2 - m}}{m - \sqrt{m^2 - m}} \quad (5.39)$$

虽然 Nakagami 和莱斯的概率密度表现了很好的“大体”一致, 但其趋于 $r=0$ 的斜率不同。这进而会影响可达到的分集阶数(见第 13 章)。

两个概率密度函数的主要差别在于, 莱斯分布给出了非零均值复高斯分布振幅的精确分布, 这意味着出现一个主导分量及大量非主导分量。Nakagami 分布以近似的方式描述了一个矢量过程的振幅分布, 在该矢量过程中, 中心极限定律不再是必须有效的(比如超宽带信道, 见第 7 章)。

5.6 多普勒谱

5.6.1 移动中的移动台的时变性

5.3 节给出了由于运动引起的频率偏移的物理解释, 即多普勒效应。如果移动台移动, 多径分量以不同方向到达移动台, 从而引起不同的频率偏移, 就会导致接收频谱的扩展。这一节的目标是假设发送的信号是正弦信号(即窄带情况)来推导这个频谱。宽带传输信号的更一般的情况在第 6 章中讨论。

当一个波仅从单一方向到来时, 重写多普勒频移的表达式。 γ 表示移动台的速度矢量 $\mathbf{v}$ 与移动台处波方向之间的夹角。如式(5.8)所示, 多普勒效应导致接收频率 f 偏移了大小 v , 因此接收频率为

$$f = f_c \left[1 - \frac{v}{c_0} \cos(\gamma) \right] = f_c - v \quad (5.40)$$

其中, $v = |\mathbf{v}|$ 。很显然, 频率偏移取决于波的方向, 而且一定在范围 $f_c - v_{\max} \sim f_c + v_{\max}$ 之内, 其中 $v_{\max} = f_c v / c_0$ 。

如果有多个多径分量, 就需要知道入射波功率随 γ 变化的函数分布。因为对接收信号的统计分布感兴趣, 所以考虑接收功率的概率密度函数 $\text{pdf}_\gamma(\gamma)$ 。作为略带滥用的表示法, 我们称之为入射波的概率密度函数。到达接收机的多径分量以移动台的天线模式加权。因此, 以方向 γ 到达的多径分量必须乘以模式 $G(\gamma)$ 。这样, 接收功率谱为方向的函数:

$$S(\gamma) = \bar{\Omega} [\text{pdf}_\gamma(\gamma)G(\gamma) + \text{pdf}_\gamma(-\gamma)G(-\gamma)] \quad (5.41)$$

其中, $\bar{\Omega}$ 是到达场的平均功率。在式(5.40)中也可看出从 γ 和 $-\gamma$ 方向来的波导致了同样的多普勒频移, 因此对推导多普勒频谱来讲, 没有必要进行区分。

在最后一步, 必须执行变量变换, $\gamma \rightarrow v$ 。雅可比矩阵可以确定为

$$\left| \frac{d\gamma}{d\nu} \right| = \left| \frac{1}{\frac{d\nu}{d\gamma}} \right| = \frac{1}{\left| \frac{\nu}{c_0} f_c \sin(\gamma) \right|} = \frac{1}{\sqrt{\left(f_c \frac{\nu}{c_0} \right)^2 - (f - f_c)^2}} = \frac{1}{\sqrt{\nu_{\max}^2 - \nu^2}} \quad (5.42)$$

因此多普勒频谱变为

$$S_D(\nu) = \begin{cases} \frac{1}{2} [\text{pdf}_\gamma(\gamma)G(\gamma) + \text{pdf}_\gamma(-\gamma)G(-\gamma)] \frac{1}{\sqrt{\nu_{\max}^2 - \nu^2}}, & -\nu_{\max} \leq \nu \leq \nu_{\max} \\ 0, & \text{其他} \end{cases} \quad (5.43)$$

为了进一步运用,同时定义:

$$\Omega_n = (2\pi)^n \int_{-\nu_{\max}}^{\nu_{\max}} S_D(\nu) \nu^n d\nu \quad (5.44)$$

为多普勒频谱的 n 阶矩。

特定的角度分布和天线模式对应其特定的公式。移动台处角度频谱的一个广泛采用的模型是波均匀地从所有方位角方向入射,并且都到达水平平面,因此

$$\text{pdf}_\gamma(\gamma) = \frac{1}{2\pi} \quad (5.45)$$

上式对应于无视距连接的情形,并且大量相互作用体均匀分布在移动台周围(见第7章)。进一步假设天线是垂直偶极子天线,天线模式 $G(\gamma) = 1.5$ (见第9章),多普勒频谱变为

$$S_D(\nu) = \frac{1.5\bar{\Omega}}{\pi \sqrt{\nu_{\max}^2 - \nu^2}} \quad (5.46)$$

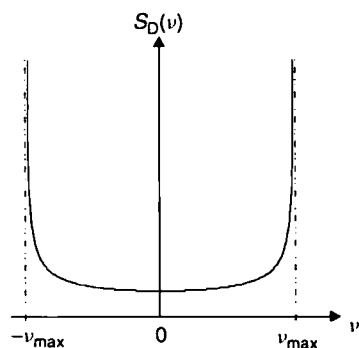


图 5.23 经典多普勒谱

这种频谱就是经典谱或 Jakes 频谱,如图 5.23 所示。它有一个很典型的“浴缸”(Jakes)形状,在最小和最大多普勒频率 $\nu = \pm \nu_{\max} = \pm f_c \nu / c_0$ 处存在(积分)奇点。这些奇点对应移动台的移动方向,或其反方向。值得注意的是,均匀的方位角分布导致了一个高度非均匀的多普勒谱。

自然地,测量的多普勒谱并不会出现奇点。即使模型和假设条件都严格有效,需要进行无穷多次测量样本,才能到达一个奇点。即使这样,经典多普勒频谱是使用最广泛的模型。其他模型还有:

- Aulin 模型[Parsons 1992],它在接近奇异点处限制 Jakes 谱的振幅;
- 高斯谱;
- 均匀谱(对应于所有波从全部三维空间均匀入射并且天线是各向同性模式)。

正如 5.2 节提到的,多普勒谱有如下两个重要的解释。

1. 它描述了频率色散。对于窄带系统和正交频分复用(OFDM),这种频率色散能够导致传输差错(将在第12章和第19章中详细讨论)。然而,它对大多数其他宽带系统没有直接影响,比如单载波时分多址(TDMA)或码分多址(CDMA)系统。
2. 它是信道时变性的量度标准,因此对所有系统都很重要。

衰落的时间依赖可以用衰落的自相关函数来很好地描述。 t 时刻同相分量与 $t + \Delta t$ 时刻同相分量的归一化相关为

$$\frac{\overline{I(t)I(t + \Delta t)}}{\overline{I(t)^2}} = J_0(2\pi \nu_{\max} \Delta t) \quad (5.47)$$

它与多普勒频谱 $S_D(\nu)$ 的傅里叶变换成反比, 而对于所有 Δt 值, $\overline{I(t)Q(t+\Delta t)} = 0$ 。因此, 包络的归一化相关函数为

$$\frac{\overline{r(t)r(t+\Delta t)} - \overline{r(t)}^2}{\overline{r(t)}^2} = J_0^2(2\pi\nu_{\max}\Delta t) \tag{5.48}$$

关于自相关函数的更详细内容, 参见第 6 章和第 13 章。

例 5.3 假设一个移动台位于衰落深陷处。平均来看, 这个移动台最少要移动多少距离才能不再受衰落深陷的影响?

解:

首先, 画出包络的相关函数[见式(5.48)]的图形, 如图 5.24 所示。如果现在定义“不再受影响”为“包络相关系数等于 0.5”, 则可以看出接收机(平均)必须移动 0.18λ 。如果想要与衰落深陷完全不相干, 则必须移动 0.38λ 。

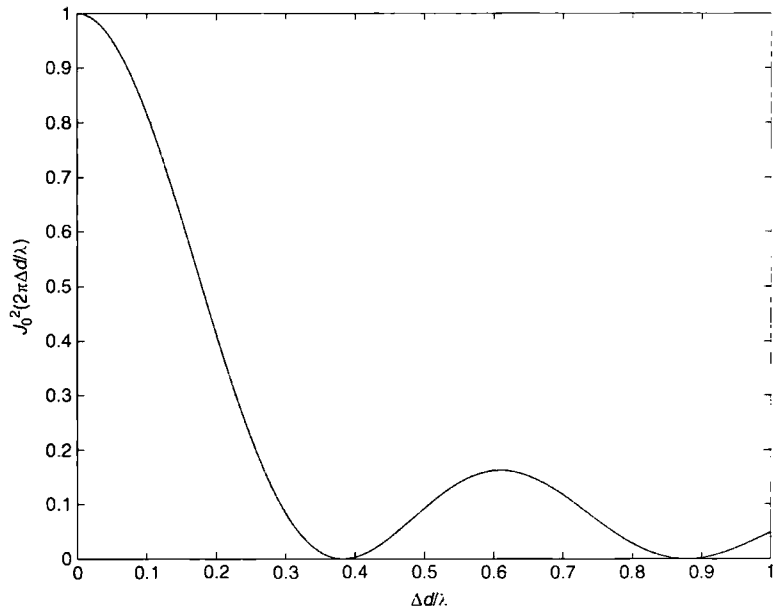


图 5.24 振幅相关与接收机位移的函数关系

5.6.2 固定无线系统中的时变性

在一个固定无线系统, 即移动台不动的系统中[Anderson 2003], 时变性可仅由相互作用体产生, 因而产生的衰落导致时域信道的考虑和不同的建模方式。在一个移动链路中, 当移动台移动时跟踪空间衰落, 即图 5.11 中的“地形”。这样, 空间和时域衰落可以互相映射, 比例系数为移动速度 v 。在固定无线链路中, 变化的原因在这里变成一些散射物体的移动, 最好的例子是汽车或者风吹的叶子。

测量表明, 在很多情况下, 衰落的幅度统计特性(取不同时间观测值的集合)为莱斯的。然而, 现在的莱斯系数表示时不变多径分量功率与时变多径分量功率之比。换句话说, 时域莱斯因子和视距(或者其他主导)分量没有关系: 在一个纯的非视距并有很多相同强度分量的情况下, 如果所有多径分量是时不变的, 则莱斯因子可趋于无穷大。

这样一个系统的多普勒谱包含一个在 $\nu=0$ 的 δ 脉冲, 和一个取决于相互作用体的运动和位置的连续谱。各种测量活动表明, 这个散开的谱是近似高斯的。

5.7 衰落的时间依赖性

5.7.1 电平通过率

多普勒谱完全刻画了衰落的时间统计量。然而, 通常仍需要一个能对系统性能更一目了然的不同表示方法。衰落深陷的发生率是一个能衡量系统的参数, 这个发生率称为电平通过率(LCR)。显然, 它取决于所考察的那个电平, 即衰落深陷是怎样定义的: 低于均值 30 dB 的情形要比低于均值 3 dB 的情形少得多。衰落深陷的允许深度取决于平均场强及所考察的系统, 我们希望获得任意电平(即衰落深陷的深度)的电平通过率。

假如给出了数学表达式, 电平通过率定义为接收场强在正方向穿过某一电平 r 的通过率的期望值, 也可写成

$$N_R(r) = \int_0^{\infty} \dot{r} \cdot \text{pdf}_{r,\dot{r}}(r, \dot{r}) d\dot{r}, \quad r \geq 0 \quad (5.49)$$

其中, $\dot{r} = dr/dt$ 是时间导数, $\text{pdf}_{r,\dot{r}}$ 是 r 和 $\dot{r}$ 的联合概率密度函数。

在附录 5. C (www.wiley.com/go/molisch) 中, 推导电平通过率为

$$N_R(r) = \sqrt{\frac{\Omega_2}{\pi\Omega_0}} \frac{r}{\sqrt{2\Omega_0}} \exp\left(-\frac{r^2}{2\Omega_0}\right) \quad (5.50)$$

注意, $\sqrt{2\Omega_0}$ 是振幅值的均方根。图 5.25 给出了瑞利衰落的振幅和 Jakes 多普勒谱的电平通过率。

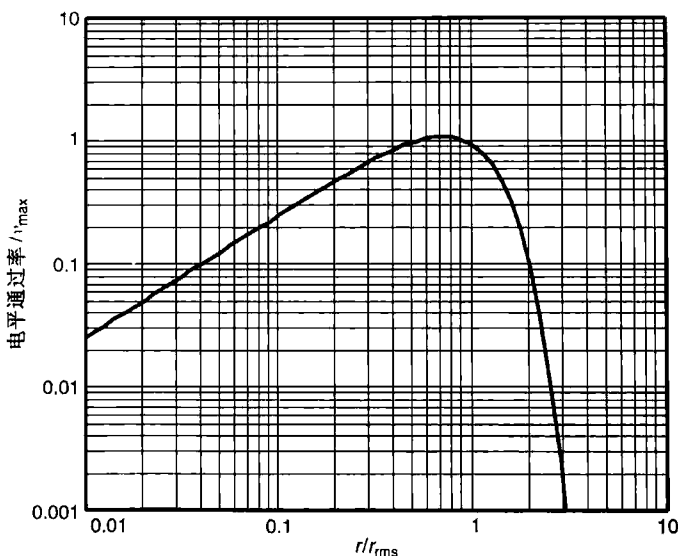


图 5.25 瑞利衰落振幅和 Jakes 谱情况下, 用最大多普勒频率归一化的电平通过率与归一化的电平 r/r_{rms} 的函数关系

5.7.2 平均衰落持续时间

另一个感兴趣的参数是平均衰落持续时间(ADF)。前几节已经推导出场强低于考察门限值时的速率(即电平通过率),以及场强低于这个门限值的时间百分比(即场强的累积分布函数)。平均衰落持续时间可以很容易地由两个量的商计算出来:

$$ADF(r) = \frac{cdf_r(r)}{N_R(r)} \quad (5.51)$$

平均衰落持续时间的图形如图 5.26 所示。

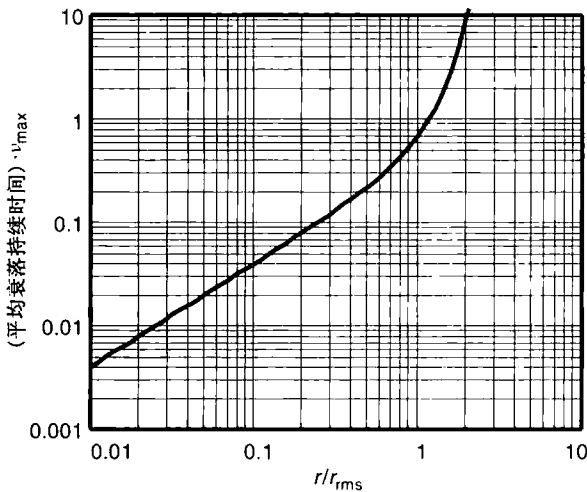


图 5.26 瑞利衰落振幅和 Jakes 谱情况下,用最大多普勒频率归一化的平均衰落持续时间与归一化的电平 r/r_{rms} 的函数关系

例 5.4 假设一个多径环境,接收信号服从瑞利分布,多普勒谱是经典浴缸(Jakes)形状。计算最大多普勒频率为 $\nu_{max} = 50$ Hz 时的电平通过率和平均衰落持续时间,振幅门限为

$$r_{min} = \frac{\sqrt{2\Omega_0}}{10}, \frac{\sqrt{2\Omega_0}}{2}, \sqrt{2\Omega_0}$$

解:

由式(5.50)计算电平通过率。对于 Jakes 情形,多普勒谱的二阶矩如下:

$$\Omega_2 = \frac{1}{2} \Omega_0 (2\pi \nu_{max})^2 \quad (5.52)$$

因此,

$$N_R(r_{min}) = \sqrt{2\pi} \cdot \nu_{max} \cdot \frac{r_{min}}{\sqrt{2\Omega_0}} \exp\left(-\frac{r_{min}^2}{2\Omega_0}\right) \quad (5.53)$$

平均衰落持续时间由式(5.51)计算,其中

$$cdf(r_{min}) = 1 - \exp\left(-\left(\frac{r_{min}}{\sqrt{2\Omega_0}}\right)^2\right) \quad (5.54)$$

代入不同的门限值 r_{min} , 所得结果如表 5.1 所示。

表 5.1 门限值对平均衰落持续时间和电平通过率的影响

$r_{\min}$	$N_R(r_{\min})$	$\text{cdf}(r_{\min})$	ADF (ms)
$\frac{\sqrt{2\Omega_0}}{10}$	12.4	0.01	0.8
$\frac{\sqrt{2\Omega_0}}{2}$	48.8	0.22	4.5
$\sqrt{2\Omega_0}$	46.1	0.63	13.7

5.7.3 随机频率调制

随机信道会导致接收信号的随机相位偏移。在一个时变信道中,这种相位偏移也是时变的。根据定义,时变的相偏就是频率调制。本节计算这种随机频率调制。

瞬时频率 $\dot{\psi}$ 的概率密度函数可以用附录 5. C (www.wiley.com/go/molisch) 中的联合概率密度函数 $\text{pdf}_{r,\dot{r},\dot{\psi}}$ 来计算。对变量 r , $\dot{r}$ 和 $\dot{\psi}$ 求积分,结果为

$$\text{pdf}_{\dot{\psi}}(\dot{\psi}) = \frac{1}{2} \sqrt{\frac{\Omega_0}{\Omega_2}} \left(1 + \frac{\Omega_0}{\Omega_2} \dot{\psi}^2\right)^{-3/2} \quad (5.55)$$

这是有两个自由度的“学生 t 分布”[Mardia et al. 1979] (见图 5.27)。累积分布为

$$\text{cdf}_{\dot{\psi}}(\dot{\psi}) = \frac{1}{2} \left[1 + \sqrt{\frac{\Omega_0}{\Omega_2}} \dot{\psi} \left(1 + \frac{\Omega_0}{\Omega_2} \dot{\psi}^2\right)^{-1/2} \right] \quad (5.56)$$

值得注意的是,瞬时频率值遍布于 $-\infty$ 到 $+\infty$ 之间,它们并没有被限制在可能的多普勒频率范围内!

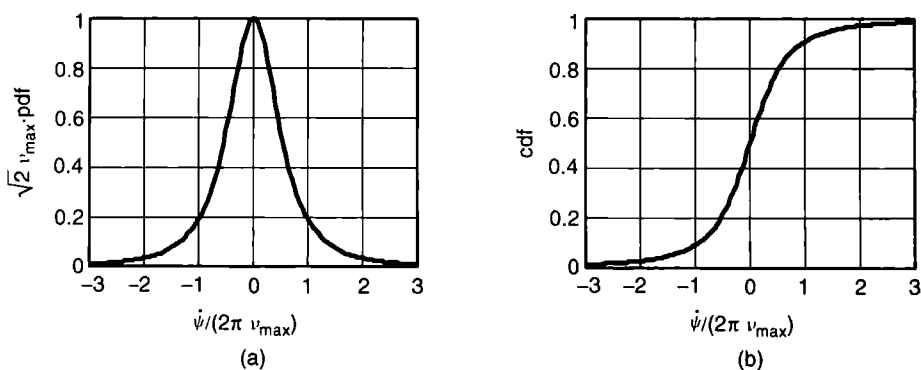


图 5.27 随机频率调制的归一化概率密度函数(a)和累积分布函数(b)

强的频率调制很可能发生在衰落深陷中:如果信号电平低,可能发生非常强的相对变化。振幅水平为 r_0 的条件下[Jakes 1974],瞬时频率的概率密度的数学表达式展示了一些直观上合意的结果:

$$\text{pdf}(\dot{\psi}|r_0) = \frac{r_0}{\sqrt{2\pi\Omega_2}} \exp\left(-r_0^2 \frac{\dot{\psi}^2}{2\Omega_2}\right) \quad (5.57)$$

这是方差为 Ω_2/r_0^2 的零均值高斯分布。信号电平越高,方差越小。

可以看出,衰落深陷以两种方式产生误差:一方面,信号电平越低,对噪声的敏感度越高。另一方面,它们增加了强随机调频的概率,这就会在任何依靠传输信号相位来传递信息的系统中引入误差。另外,第 12 章会讨论到衰落深陷也与符号间干扰有关。

5.8 大尺度衰落

由不同的多径分量叠加产生的小尺度衰落,在几个波长的空间范围内快速变化。如果场强在一个小的区域内(比如,10个波长 $\times$ 10个波长)进行平均,则可得到小尺度平均(SSA)场强^①。前面几节把小尺度平均场强当成常数。然而,正如本书第一部分“引言”所解释的,当我们在一个大的空间尺度内考虑问题时,小尺度平均场强由于相互作用体造成的多径分量的阴影而变化。

许多实验研究表明,小尺度平均场强 F 用对数刻度来描绘时,是均值为 μ 的高斯分布。这样的分布是对数正态分布,它的概率密度函数是

$$\text{pdf}_F(F) = \frac{20/\ln(10)}{F\sigma_F\sqrt{2\pi}} \cdot \exp\left[-\frac{(20\log_{10}(F) - \mu_{\text{dB}})^2}{2 \cdot \sigma_F^2}\right] \quad (5.58)$$

其中, σ_F 是 F 的标准差, μ_{dB} 是 F 以dB为单位的平均值。类似地,功率也是对数正态分布。然而,有一个非常重要的细节:将小尺度平均功率的对数拟合到一个正态分布时,可发现该分布的中值 $\mu_{P,\text{dB}}$ 与场强分布的中值 μ_{dB} 相关, $\mu_{P,\text{dB}} = \mu_{\text{dB}} + 10\log(4/\pi)$ dB。该功率的概率密度函数因此为

$$\text{pdf}_P(P) = \frac{10/\ln(10)}{P\sigma_P\sqrt{2\pi}} \cdot \exp\left[-\frac{(10\log_{10}(P) - \mu_{P,\text{dB}})^2}{2 \cdot \sigma_P^2}\right] \quad (5.59)$$

其中, $\sigma_P = \sigma_F$ 。 σ_F 的典型值为4~10 dB。

对数正态分布的另一个有趣特性是:对数正态分布值的总和也近似服从对数正态分布。根据合成分布参数计算的均值和方差有各种方法。适合各种特殊情况的方法可以参见 Stueber [1996];最近由 Mehta et al. [2007]导出了一个一般的灵活方法。

F 的对数正态方差通常要归因于阴影效应。考察图 5.28 中的情形。如果移动台移动,它就改变了角度 γ ,继而改变绕射系数 v_F 。如果边缘与移动台之间的距离很大,那么移动台必然要移动一个大的距离(比如,几十个波长),才能使场强发生显著变化。注意,这种效应改变了多径分量的绝对振幅 $|a|$,并且与干涉效应没有关系。现在考虑这样一种情况,一个多径分量在从发射机到接收机的路程中,经历了若干个这样或类似的过程,每一个过程都引发了一定的衰减。接收场强就取决于这些衰减的乘积。在对数刻度下,就是这些衰减之和。因此可以把这种效果建模为对数刻度上的随机变量之和,即一个对数正态分布。然而,刚才描述的这种机制并不是在所有物理状态都必然有效^②。

现在考虑基于大范围内的样本(即,包括对数正态衰落和干涉效应)的场强的统计量。这些样本的概率密度函数由所谓的 Suzuki 分布给出,它遵循条件统计直接推导的结果。小区域内场强的均值为 $\bar{r} = \sigma \sqrt{\pi/2}$,而且场强的局部值的分布是 σ 的条件分布:

$$\text{pdf}(r|\bar{r}) = \frac{\pi r}{2\bar{r}^2} \exp\left(-\frac{\pi r^2}{4\bar{r}^2}\right) \quad (5.60)$$

局部均值(即上面用的期望值)依据对数正态统计量分布。无条件化得到场强的概率密度函数:

① 严格地讲,这只是小尺度平均场强的一个近似,因为在用来取平均的有限区域内,可能只有有限个统计独立的样本值。

② 另一种解释,基于双散射函数由 Andersen [2002]提出。但是,独立于其产生机制,大量测量证实, F 的概率密度函数由对数正态函数很好地近似。

$$\text{pdf}_r(r) = \int_0^\infty \frac{\pi r}{2\bar{r}^2} \exp\left(-\frac{\pi r^2}{4\bar{r}^2}\right) \frac{2 \cdot 10 \ln(10)}{\bar{r} \sigma_F \sqrt{2\pi}} \exp\left(-\frac{(20 \log_{10}(\bar{r}) - \mu_{\text{dB}})^2}{2\sigma_F^2}\right) d\bar{r} \quad (5.61)$$

这个概率密度函数也称为 Suzuki 分布。图 5.29 给出了一个例子。如果小尺度统计是莱斯分布或者 Nakagami 分布, 则可以定义一个类似的函数。

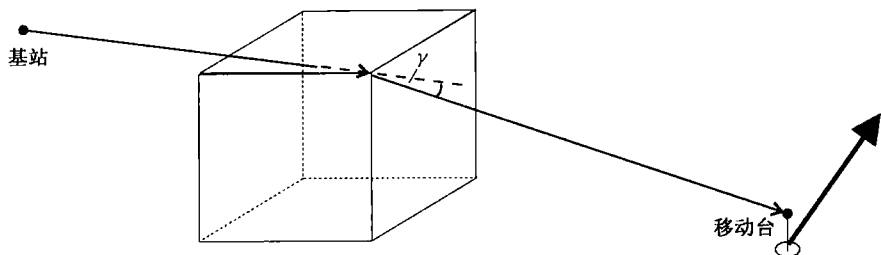


图 5.28 被建筑物遮挡

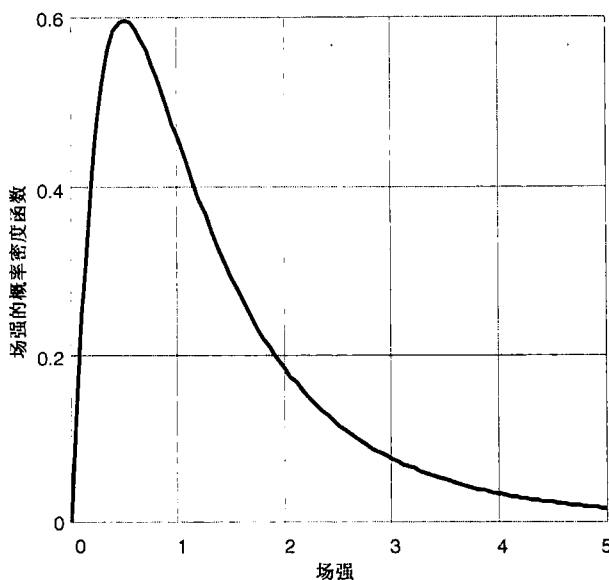


图 5.29 Suzuki 分布, $\sigma_F = 5.6 \text{ dB}$, $\mu = 1.4 \text{ dB}$

功率的概率密度函数为

$$\text{pdf}_P(P) = \int_0^\infty \frac{1}{\Omega} \exp\left(-\frac{P}{\Omega}\right) \frac{10/\ln(10)}{\Omega \sigma_P \sqrt{2\pi}} \exp\left(-\frac{(10 \log_{10}(\Omega) - \mu_P, \text{dB})^2}{2\sigma_P^2}\right) d\Omega \quad (5.62)$$

既然实际情况中大尺度和小尺度衰落都会发生, 衰落余量必须要考虑两个效果的结合(见第3章)。一种可能性是将瑞利分布的衰落余量与对数正态分布的衰落余量相加。由于其简单性, 这种方法经常采用, 但是它过高地估计了所需的衰落余量。更准确的方法基于 Suzuki 分布的累积概率函数, 可以令式(5.59)从 $-\infty$ 到 x 的积分得到。这样就可以计算出, 在给定允许中断概率(比如 0.05)的情况下, 必要的平均场强 r_0 , 具体过程参见下面的例题。

例 5.5 考虑一个信道, 其 $\sigma_F = 6 \text{ dB}$, $\mu = 0 \text{ dB}$ 。计算为使中断概率小于 5% 的相对于阴影的 dB 均值的 Suzuki 分布的衰落余量。再分别计算瑞利分布和阴影分布的衰落余量。

解:

对于 Suzuki 分布:

$$\begin{aligned} P_{\text{out}} &= \text{cdf}(r_{\min}) = \int_0^{r_{\min}} \text{pdf}_r(r) dr \\ &= \int_0^{\infty} \left(1 - \exp\left(-\frac{\pi r_{\min}^2}{4\bar{r}^2}\right) \right) \frac{20/\ln(10)}{\bar{r}\sigma_F\sqrt{2\pi}} \exp\left(-\frac{(20\log_{10}\bar{r} - \mu_{\text{dB}})^2}{2\sigma_F^2}\right) d\bar{r} \end{aligned} \quad (5.63)$$

代入 $\sigma_F = 6 \text{ dB}$, $\mu_{\text{dB}} = 0 \text{ dB}$, 累积分布函数可以表示为

$$\text{cdf}(r_{\min}) = \frac{20/\ln 10}{\sqrt{2\pi} \cdot 6} \int_0^{\infty} \frac{1}{\bar{r}} \exp\left(-\frac{(20\log_{10}\bar{r} - 0)^2}{2 \cdot 36}\right) \left(1 - \exp\left(-\frac{\pi}{4} \cdot \frac{r_{\min}^2}{\bar{r}^2}\right) \right) d\bar{r} \quad (5.64)$$

衰落余量定义为

$$M = \frac{\mu^2}{r_{\min}^2} \quad (5.65)$$

对于 $r_{\min}$ 的不同值, 计算式(5.65), 可以得到累积分布函数的图形, 如图 5.30 所示。中断概率为 5% 时, 需要的衰落余量为 15.5 dB。

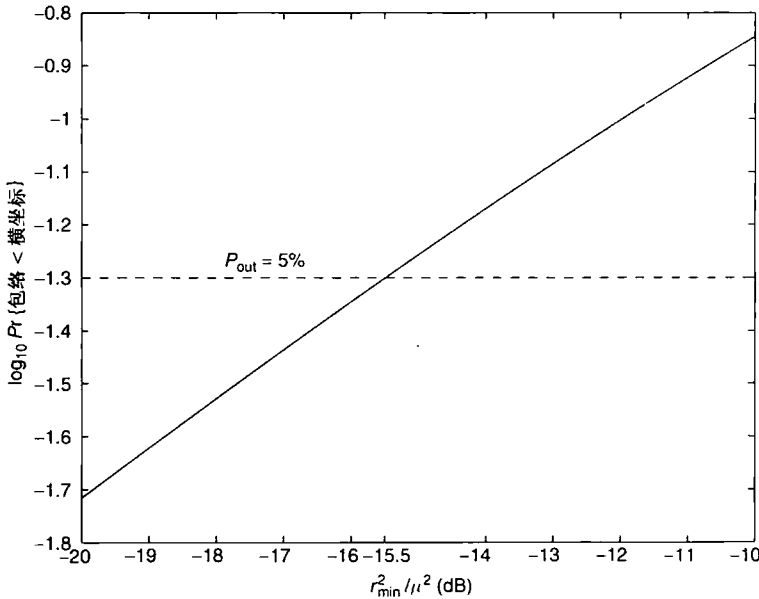


图 5.30 Suzuki 累积概率分布, $\sigma_F = 6 \text{ dB}$, $\mu_{\text{dB}} = 0$

只有瑞利衰落时的中断概率计算如下:

$$\begin{aligned} P_{\text{out}} &= \text{cdf}(r_{\min}) = 1 - \exp\left(-\frac{r_{\min}^2}{2\sigma^2}\right) \\ &= 1 - \exp(-1/M) \end{aligned} \quad (5.66)$$

经过一些操作, 可得

$$\begin{aligned} M_{\text{Rayleigh, dB}} &= -10\log_{10}(-\ln(1 - P_{\text{out}})) \\ &= 12.9 \text{ dB} \end{aligned} \quad (5.67)$$

对于阴影衰落, 以 dB 为单位的场强值的概率密度函数是标准的高斯分布。因此 cdf 的补 (即 1 减去 cdf) 由 Q 函数定义为

$$Q(a) = \frac{1}{\sqrt{2\pi}} \int_a^{\infty} \exp\left(-\frac{x^2}{2}\right) dx$$

中断概率为

$$P_{\text{out}} = Q\left(\frac{M_{\text{large-scale, dB}}}{\sigma_F}\right) \quad (5.68)$$

将 $\sigma_F = 6$ dB 和 $P_{\text{out}} = 0.05$ 代入式(5.68), 可得

$$\begin{aligned} M_{\text{large-scale, dB}} &= 6 \cdot Q^{-1}(0.05) \\ &= 9.9 \text{ dB} \end{aligned} \quad (5.69)$$

将瑞利衰落与阴影衰落余量之和作为计算的衰落余量, 可得

$$M_{\text{dB}} = 12.9 + 9.9 \text{ dB} = 22.8 \text{ dB} \quad (5.70)$$

与通过 Suzuki 分布计算的结果相比, 可以看到这是一种更保守的估计。

结论就是, Suzuki 分布可以合理地由对数正态分布很好地近似, 其中近似对数正态分布的参数仅与阴影的参数(如瑞利衰落不相关)通过下式相关:

$$\mu_{\text{approx, dB}} = \mu_{\text{shadow, dB}} - 1.5 \quad (5.71)$$

$$\sigma_{\text{dB}}^2 = \sigma_{\text{shadow, dB}}^2 + 13 \quad (5.72)$$

最后变成单独阴影的空间相关(与小尺度衰落无关)。阴影的值仅随着移动台的位置缓慢变化。这样, 在相距 Δx 的距离测量的阴影场强(或者功率)的实现是相关的。最一般的模型是指数相关模型, 即 $E_x\{F(x)F(x+\Delta x)\} = \exp(-\Delta x/\bar{x})$, 其中去相关距离典型地在 5 ~ 50 m 的范围, 具体值取决于环境。

5.9 附录

请访问 www.wiley.com/go/molisch。

深入阅读

本章推导的数学基础在许多统计和随机过程的权威教材中都进行了论述, 特别是经典教材 Papoulis[1991]。振幅分布的统计模型和多普勒频谱首先在 Clarke[1968]中描述。信道统计的全面阐述, 瑞利情况下的多普勒频谱、电平通过率和平均衰落持续时间的推导可以在 Jakes[1974]中找到。因为瑞利衰落基于 I 分量和 Q 分量的高斯衰落, 关于高斯多元分析的大量文献是很合适的[Muirhead 1982]。Nakagami 分布的推导及其统计特性可以在 Nakagami[1960]中找到; 物理解释在 Braun and Dersch[1991]中给出; 莱斯分布在 Rice[1947]中给出推导。Suzuki 分布在 Suzuki[1977]中给出推导。对数正态分布的更详细信息, 特别是几个对数正态分布的变量之和, 可以在 Stueber[1996]和 Cardieri and Rappaport[2001]和 Mehta et al.[2007]中找到。Nakagami 小尺度衰落和对数正态阴影的结合在 Thjung and Chai[1999]中进行了处理。衰落统计量, 包括 Nakagami 和莱斯分布的平均衰落持续时间和电平通过率, 在 Abdi et al.[2000]中进行了总结。统计信道描述的另一个重要方面是根据指定的多普勒频谱产生随机变量。该领域的一个广泛的描述可以在 Paetold[2002]中找到。

第6章 宽带和方向性信道的特性

6.1 概述

在上一章中,传输信号为纯正弦信号时,考虑了多径传播对接收场强和时变性的影响。这些考虑对于所有发送信号的带宽“非常小”(更精确的定义见下文)的系统都是正确的。然而,大多数现在和未来的无线系统采用很大的宽带,或者是因为它们要满足高的数据速率,或者是因为它们的多址接入技术方案(见第17章至第19章),因此必须要能描述在大宽带信道的变化。本章讨论的就是这些描述方法。

多径传播在宽带系统中的影响可以采用两种不同的方式来解释:(i)信道传输函数随带宽而变化(这称为信道的频率选择性);(ii)信道的冲激响应不是 δ 函数,换句话说,到达信号的持续时间比发送信号的长(又称为时延色散)。这两种解释是等价的,可以通过时延(时)域转换和频域之间的傅里叶变换来说明。

这一章首先再次采用最简单的信道,即二径信道,解释宽带信道的基本概念。然后给出宽带时变信道的最一般化的统计描述公式(见6.3节),并讨论它最普遍的特殊形式,即广义平稳非相关散射(WSSUS)模型(见6.4节)。由于这些描述方法相当复杂,经常采用精简参数进行更简洁的描述(见6.5节)。6.6节考虑这样一种情况,即信道不仅足够宽,能够表示传输函数可测量的变化,甚至宽到带宽和载波频率相当的程度,这种情况称为“超宽带”。

在宽带信道上工作的系统有如下一些重要的性质。

- 存在符号间干扰。很容易从时延色散的角度来理解这一点。如果发送一个长度为 T_s 的符号,则该符号相应的接收信号有较长的持续时间,因而会干扰下一个符号(见第2章)。12.3节描述在没有采取任何措施的情况下,这种符号间干扰对误比特率的影响。第16章将描述均衡器的结构,它可以有效地抵消这种符号间干扰的有害影响。
- 可以减少衰落的有害影响。这个作用在频域很容易理解:即使发送信号的部分频谱衰落很严重,其他频率并没有衰减。适当的编码和信号处理可以利用这一事实(将在第16章、第18章和第19章中详述)。

信道特性的变化不仅取决于所考察的频率,也取决于位置。后者的作用与信道的方向性有关,即多径分量是从哪个方向入射的。6.7节讨论这些方向特性的随机描述方法,这对天线的分集(见第13章)和多元天线特别重要(见第20章)。

6.2 时延色散的成因

6.2.1 两径模型

为什么信道表现出时延色散,或者在给定的频率范围内总是变化的?正如第5章开始所介绍的,最简单的情况是两径模型。发送信号通过两个不同的传播路径到达接收机,这两条路

径的时延分别为

$$\tau_1 = d_1/c_0 \quad \text{和} \quad \tau_2 = d_2/c_0 \quad (6.1)$$

现在假设这两个时延不随时间变化, 这种情况发生在发射机、接收机和相互作用体都不移动时。因此, 信道是线性和时不变的, 其冲激响应为

$$h(\tau) = a_1 \delta(\tau - \tau_1) + a_2 \delta(\tau - \tau_2) \quad (6.2)$$

这里再次采用复数幅度 $a = |a| \exp(j\varphi)$ 。很明显, 这样的信道表现出时延色散特性; 冲激响应的支撑(持续时间)有一个有限的扩展范围, 即 $\tau_2 - \tau_1$ 。冲激响应的傅里叶变换可以得到传输函数 $H(j\omega)$:

$$H(f) = \int_{-\infty}^{\infty} h(\tau) \exp[-j2\pi f\tau] d\tau = a_1 \exp[-j2\pi f\tau_1] + a_2 \exp[-j2\pi f\tau_2] \quad (6.3)$$

传输函数的幅度为

$$|H(f)| = \sqrt{|a_1|^2 + |a_2|^2 + 2|a_1||a_2|\cos(2\pi f \cdot \Delta\tau - \Delta\varphi)} \quad (6.4)$$

其中, $\Delta\tau = \tau_2 - \tau_1$, $\Delta\varphi = \varphi_2 - \varphi_1$ 。

图 6.1 是一个典型情况的传输函数。可以看到, 传输函数取决于频率, 因而有频率选择性衰落信道。还可以看到, 传输函数在所谓的槽口频率处有深陷点(凹点)。在两径模型中, 当接收到的两个到达波相位差 180° 时就会出现槽口频率。两个相邻槽口频率的频率差为

$$\Delta f_{\text{Notch}} = \frac{1}{\Delta\tau} \quad (6.5)$$

两个波形的幅度越相近, 它们之间的相消干涉就越严重。

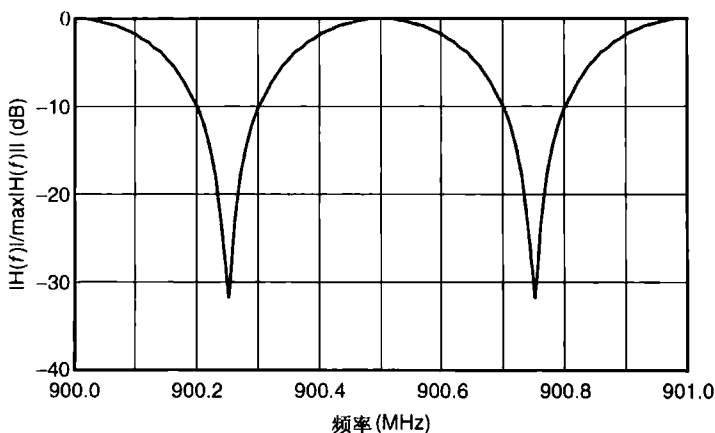


图 6.1 载波频率 900 MHz 处的归一化传输函数, 其中 $|a_1| = 1.0$, $|a_2| = 0.95$, $\Delta\varphi = 0$, $\tau_1 = 4 \mu\text{s}$, $\tau_2 = 6 \mu\text{s}$

有衰落深陷的信道不仅幅度失真, 信号的相位也会失真。这可以通过考察群延迟清楚地看到, 群延迟定义为信道传输函数相位 $\phi_H = \arg(H(f))$ 的导数:

$$\tau_{Gr} = -\frac{1}{2\pi} \frac{d\phi_H}{df} \quad (6.6)$$

从图 6.2 中可以看到, 在衰落深陷处群延迟变得很大。后面的章节将会讨论到群延迟与符号间干扰有关。

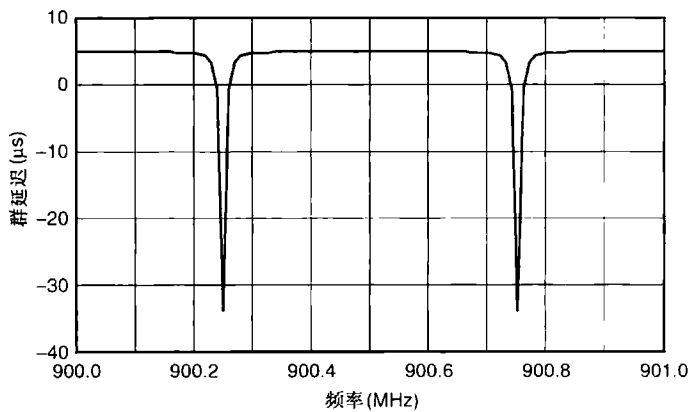


图 6.2 群延迟和频率的关系(参数与图 6.1 相同)

6.2.2 一般的情况

讨论了简单的两径模型之后,现在讨论更一般的情况,相互作用体可以在平面的任何地方。再次假设场景是静态的,发射机、接收机和相互作用体都不动。现在画一个椭球,其焦点为发射机和接收机,它的离心率决定了时延^①。与同一特定椭球上的相互作用体发生一次相互作用的所有射线同时达到接收机。与不同椭球上的相互作用体发生相互作用的射线,将在不同时刻到达接收机。因此,如果环境中的相互作用体不是全部位于同一个椭球上,信道就是时延色散的。

很显然,在实际环境中相互作用体不可能都精确地位于同一个椭球上。下一个问题是:这种“单个椭球”条件必须怎样严格满足,才能保证信道仍然是“实际上”非时延色散的?答案取决于系统的带宽。当 $\Delta\tau \ll 1/W$ 时,带宽为 W 的接收机无法区分在 τ 和 $\tau + \Delta\tau$ 时刻到达的回波(对很多定性的考虑,结合 $\Delta\tau = 1/W$ 考虑上述条件是充分的)。因此,在圆环形区域内反射的回波,到达时刻在 τ 和 $\tau + \Delta\tau$ 之间时,可认为“实际上”同时到达(见图 6.3)。

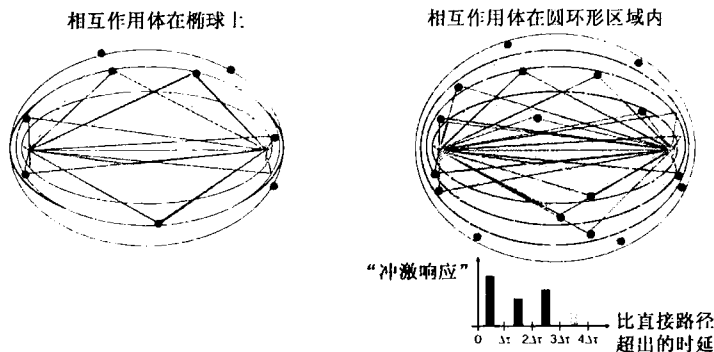


图 6.3 位于同一椭球上的散射物导致同样的时延

宽带信道的冲激响应的时域离散近似可以通过把冲激响应分成宽度为 $\Delta\tau$ 的小块,然后计算在每个小块内的回波之和得到。如果在每一个圆环形区域内有足够多的非主导的相互作用体,则落在每一个时延小块内的多径分量满足中心极限定理。在这种情况下,每个小块的幅度

① 这些椭球和第 4 章描述的非涅耳球很相似。不同的是第 4 章引入 $i \cdot \pi$ 的相位偏移的超出时延,这里关注的时延通常大得多。

可以统计地描述,而且这个幅度的概率密度函数是瑞利分布或莱斯分布。因此,第5章中的所有公式仍然成立,但是它们适用于一个时延小块内的场强。进一步定义最小时延为基站和移动台之间的直接路径的时延 d/c_0 , 定义最大时延为基站和移动台之间通过最远的“有效”相互作用体时的时延,即能够对冲激响应产生可测量的贡献的最远的作用体的时延。最大超出时延 $\tau_{\max}$ 定义为最小时延和最大时延之间的差值。

以上考虑同时给出了时域角度的窄带和宽带的数学表述:如果系统带宽的倒数 $1/W$ 远远大于 $\tau_{\max}$, 则系统是窄带系统。在这种情况下,所有的回波都落在同一个时延块内,时延块的幅度为 $\alpha(t)$ 。在其他所有情况下,系统是宽带的。在宽带系统中,接收到的信号形状和持续时间与发送的信号是不同的。在窄带系统中,它们是相同的。

如果冲激响应是有限长度的,根据傅里叶变换理论,可以得出传输函数是频域依赖的 $\mathcal{F}\{h(\tau)\} = H(f)$ 。这样,时延色散等同于频率选择性。一个频率选择性信道不能用一个简单的衰减系数来描述,而是必须为传输函数的细节建模。注意,如果在一个足够大的带宽上分析,任何实际信道都是频率选择性的。在实际的系统中,的确是这样吗? 这等价于比较最大超出时延和系统带宽的倒数。图6.4描绘了这些关系,展示了宽带系统在时延和频域的变化。

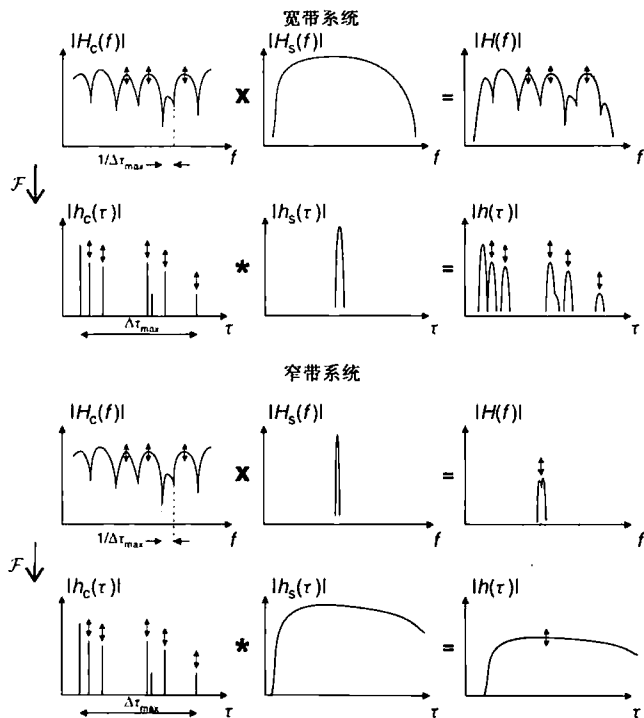


图6.4 窄带和宽带系统。 $H_c(f)$ 为信道传输函数; $h_c(\tau)$ 为信道冲激响应。引自 Molisch [2000] © Prentice Hall

要强调的是,宽带无线系统的定义与射频工程师定义的“宽带”有根本的区别。宽带的射频定义意味着系统带宽与载波频率相当^①。而在无线通信中,将信道的性质和系统的性质相比较。因此,一个系统对一个信道来说是宽带系统,而同一个系统对另一个信道来说是窄带系统。

① 在无线通信中,通常把带宽大于载波频率的20%的系统称为超宽带(UWB)系统(见6.6节)。这个定义和宽带的射频定义相似。

6.3 无线信道的系统理论描述

如前面的章节所述,无线信道可以描述为冲激响应,因而可以解释为线性滤波器。如果基站、移动台和相互作用体是静态的,那么信道是时不变的,冲激响应为 $h(\tau)$ 。在这种情况下,可以应用熟知的线性时不变(LTI)系统[Oppenheim and Schaffer 1985]的理论。然而,在一般情况下,无线信道是时变的,冲激响应 $h(t, \tau)$ 是随时间变化的,我们必须区分开时间 t 和时延 τ 。因此必须用到线性时变系统(LTV)理论。这并不仅仅是线性时不变系统的简单扩展,而是有了大量的理论上的挑战,并破坏了很多直观上的概念。幸运的是,许多无线信道可以归到慢时变系统,也称为准静态信道。在这种情况下,线性时不变系统的许多概念可以保留,只需做一些小的修改。

6.3.1 确定性线性时变系统的特性

由于时变系统的冲激响应 $h(t, \tau)$ 取决于两个变量 τ 和 t ,可以关于两者之一(或两者)进行傅里叶变换,得到4种不同但等价的表示方法。这一节研究这些表示方法及其优缺点。

从系统理论角度讲,最简单明了的表示就是将系统的输入信号(发送信号) $x(t)$ 和输出信号(接收信号) $y(t)$ 之间的关系写成下式:

$$y(t) = \int_{-\infty}^{\infty} x(\tau)K(t, \tau) d\tau \tag{6.7}$$

其中 $K(t, \tau)$ 是积分方程的核,它与冲激响应有关。对于线性时不变系统,关系 $K(t, \tau) = h(t - \tau)$ 成立。一般地,定义时变冲激响应为

$$h(t, \tau) = K(t, t - \tau) \tag{6.8}$$

因此有

$$y(t) = \int_{-\infty}^{\infty} x(t - \tau)h(t, \tau) d\tau \tag{6.9}$$

当冲激响应随时间变化比较慢时,准确地说,当冲激响应(信号)的持续时间比信道发生明显变化的时间短时,就可以进行直观的解释。可以认为在时刻 t ,系统的行为就像一个线性时不变系统。变量 t 可以认为是用于参数化冲激响应的“绝对”时间,即表明哪个(从一个大的集合中)冲激响应 $h(\tau)$ 现在是有效的。这样的系统又称为准静态系统。

关于变量 τ 的冲激响应的傅里叶变换可以得到时变传输函数 $H(t, f)$:

$$H(t, f) = \int_{-\infty}^{\infty} h(t, \tau) \exp(-j2\pi f\tau) d\tau \tag{6.10}$$

输入输出关系可以表示为

$$y(t) = \int_{-\infty}^{\infty} X(f)H(t, f) \exp(j2\pi ft) df \tag{6.11}$$

对准静态系统的解释很直观:输入信号的频谱乘以“当前有效”的传输函数的频谱,可以得到输出信号的频谱。然而,如果信道随时间变化较快,则式(6.11)只是纯粹的数学关系式。输出信号的频谱是一个二重积分:

$$Y(\tilde{f}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X(f)H(t, f) \exp(j2\pi ft) \exp(-j2\pi \tilde{f}t) df dt \tag{6.12}$$

而不能简化为 $Y(f) = H(f)X(f)$ [Matz and Hlawatsch 1998]。

冲激响应关于变量 t 的傅里叶变换则得到一种不同的表示方法,也就是多普勒变化冲激响

应,更一般地称为扩展函数:

$$s(\nu, \tau) = \int_{-\infty}^{\infty} h(t, \tau) \exp(-j2\pi \nu t) dt \quad (6.13)$$

这个函数描述了输入信号在时延和多普勒域的扩展。

最后,函数 $s(\nu, \tau)$ 可以关于变量进行变换,得到多普勒变化的传输函数 $B(\nu, f)$:

$$B(\nu, f) = \int_{-\infty}^{\infty} s(\nu, \tau) \exp(-j2\pi f \tau) d\tau \quad (6.14)$$

图 6.5 总结了系统函数之间的相互关系,图 6.6 给出了一个测量的冲激响应的例子,图 6.7 给出了基于图 6.6 给出的数据计算得到的扩展函数。

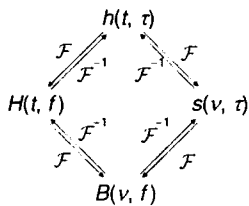


图 6.5 确定系统函数之间的关系

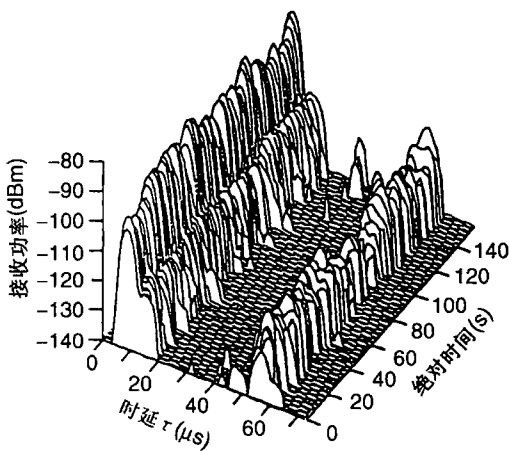


图 6.6 德国达姆施塔特附近的多山地区测量的冲激响应的平方幅度 $|h(t, \tau)|^2$ 。测量持续时间 140 s, 中心频率 900 MHz, τ 表示超出时延。引自 Liebenow and Kuhlmann [1993] © U. Liebenow

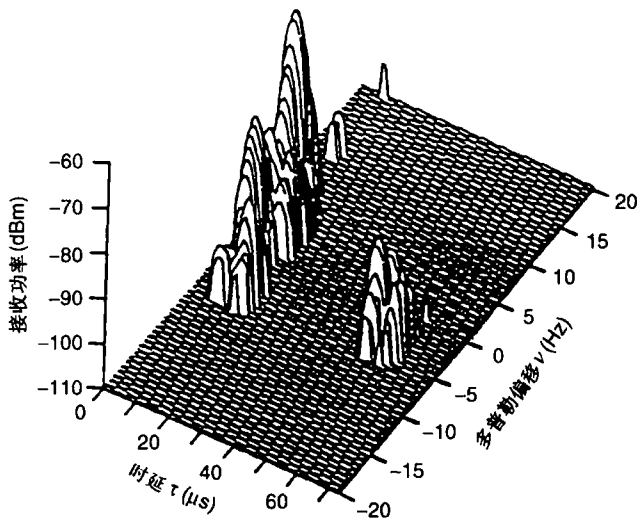


图 6.7 根据图 6.6 给出的数据计算得到的扩展函数。U. Liebenow 授权使用

6.3.2 随机系统函数

现在转到无线信道的随机描述。把它们解释为时变的随机系统,一个完整的描述需要冲激

响应的多维概率密度函数,比如任何可能时刻和时延下的复数幅度的联合概率密度函数。然而,这在实际中往往过于复杂。因此这里只限于二阶描述,也就是自相关函数(ACF)。

首先回顾一维随机过程的自相关函数(即只取决于一个参数 t 的过程)。随机过程 y 的自相关函数定义为

$$R_{yy}(t,t') = E\{y^*(t)y(t')\} \tag{6.15}$$

其中,期望值在所有可能实现的集合上求取。该集合的定义见附录 6. A (www.wiley.com/go/molisch)。自相关函数描述了信号 y 在不同时刻的幅度的概率密度函数的二阶矩之间的关系。如果概率密度函数是零均值的高斯过程,那么二阶描述包括了要求的所有信息。如果概率密度函数是非零均值的高斯过程,则均值

$$\bar{y}(t) = E\{y(t)\} \tag{6.16}$$

和自相关函数

$$\tilde{R}_{yy}(t,t') = E\{[y(t) - \bar{y}(t)]^*[y(t') - \bar{y}(t')]\} \tag{6.17}$$

就构成一个完整的描述。如果信道是非高斯信道,那么一阶和二阶统计量并不是信道的完整描述。以下主要研究零均值的高斯信道。

现在回到信道的随机描述这一问题上。把输入输出关系代入式(6.15)中,可以得到接收信号的自相关函数的如下表达式:

$$R_{yy}(t,t') = E\left\{\int_{-\infty}^{\infty} x^*(t-\tau)h^*(t,\tau) d\tau \int_{-\infty}^{\infty} x(t'-\tau')h(t',\tau') d\tau'\right\} \tag{6.18}$$

系统是线性的,那么期望与积分可以互换。而且,发送信号可以理解为与信道独立的随机过程,因此发送信号的期望和信道的期望可以独立完成。所以,接收信号的自相关函数为

$$\begin{aligned} R_{yy}(t,t') &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\{x^*(t-\tau)x(t'-\tau')\} E\{h^*(t,\tau)h(t',\tau')\} d\tau d\tau' \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{xx}(t-\tau,t'-\tau') R_h(t,t',\tau,\tau') d\tau d\tau' \end{aligned} \tag{6.19}$$

即发送信号的自相关函数和信道的自相关函数相结合,而后者为

$$R_h(t,t',\tau,\tau') = E\{h^*(t,\tau)h(t',\tau')\} \tag{6.20}$$

注意,信道的自相关函数由于内在的随机过程是二维的,因而取决于 4 个变量。

可以看出信道的自相关函数和确定信道的冲激响应在形式上的相似性:通过傅里叶变换可以得到随机系统函数。与确定情况下不同,现在必须进行关于变量对 t, t' 或者 τ, τ' 的双重傅里叶变换,从而用基本方式得到了输入和输出信号的不同自相关函数表达式之间的关系,如 $R_s(v, v', \tau, \tau') = E\{s^*(v, \tau)s(v', \tau')\} = \iint R_h(t, t', \tau, \tau') \cdot \exp(+j2\pi vt)\exp(-j2\pi v't') dt dt'$ 。

6.4 WSSUS 模型

相关函数取决于 4 个变量,对信道的刻画来说是相当复杂的形式。通过进一步假设信道的物理特性,可以简化相关函数。最常用的假设是所谓的广义平稳(WSS)假设和非相关散射(US)假设。基于这两个假设的一种模型称为广义平稳非相关散射(WSSUS)模型。

6.4.1 广义平稳

广义平稳的数学定义为:自相关函数与变量 t, t' 无关,而是取决于两者的差 $t - t'$ 。因此,

二阶幅度的统计特性不再随时间变化^①。因而有下式：

$$R_h(t, t', \tau, \tau') = R_h(t, t + \Delta t, \tau, \tau') = R_h(\Delta t, \tau, \tau') \quad (6.21)$$

按一般规律来讲，广义平稳意味着信道的统计特性不随时间变化。不能将其与静态信道相混淆，在静态信道中，信道衰落是不随时间变化的。对于平坦瑞利衰落这一简单情况，广义平稳意味着平均功率和多普勒谱都不随时间变化，而瞬时幅度是可以随时间变化的。

根据数学定义，广义平稳必须对所有的任意时刻 t 都是满足的。在实际中，这是不可能的，因为移动台移动的距离较大，由于阴影效应和路径损失的变化，平均接收功率也在变化。当然，广义平稳典型的情况是指直径约为 10λ (可与 5.1 节比较) 的区域内是满足的。这样可以定义有限时间间隔内的准静态 (与移动台移动的距离有关)，在这一时间内，统计特性没有明显的变化。

广义平稳也意味着不同多普勒频移的分量经历的衰落。这可以通过考虑多普勒变化冲激响应 $s(\nu, \tau)$ 来证明。把式 (6.21) 代入 R_s 的定义中，可得到

$$R_s(\nu, \nu', \tau, \tau') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_h(t, t + \Delta t, \tau, \tau') \exp[2\pi j(\nu t - \nu'(t + \Delta t))] dt d\tau' \quad (6.22)$$

可以重写为

$$R_s(\nu, \nu', \tau, \tau') = \int_{-\infty}^{\infty} \exp[2\pi j t(\nu - \nu')] dt \int_{-\infty}^{\infty} R_h(\Delta t, \tau, \tau') \exp[-2\pi j \nu' \Delta t] d\Delta t \quad (6.23)$$

第一个积分是 $\delta(\nu - \nu')$ 的积分表达式，因此可以分解为

$$R_s(\nu, \nu', \tau, \tau') = \tilde{P}_s(\nu, \tau, \tau') \delta(\nu - \nu') \quad (6.24)$$

这表明，如果它们有不同的多普勒频移，则所经历的衰落是不相关的。下文将详细讨论式 (6.24) 中定义的函数 $\tilde{P}_s(\cdot)$ 。

类似地，将 R_h 记为

$$R_B(\nu, \nu', f, f') = P_B(\nu, f, f') \delta(\nu - \nu') \quad (6.25)$$

6.4.2 非相关散射

非相关散射假设的定义是：具有不同时延的贡献是不相关的。用数学公式可以表示为

$$R_h(t, t', \tau, \tau') = P_h(t, t', \tau) \delta(\tau - \tau') \quad (6.26)$$

或者对于 R_s 有

$$R_s(\nu, \nu', \tau, \tau') = \tilde{P}_s(\nu, \nu', \tau) \delta(\tau - \tau') \quad (6.27)$$

当一个多径分量不包含不同时延的另一个多径分量的任何相位信息时，就满足非相关散射条件。如果散射体在空间中是随机分布的，即使移动台移动了很小的距离，相位也是在以不相关的方式变化的。

对于传输函数来说，非相关散射条件意味着 R_h 与频率的绝对值无关，而只与频率差有关^②：

$$R_H(t, t', f, f + \Delta f) = R_H(t, t', \Delta f) \quad (6.28)$$

6.4.3 WSSUS 假设

非相关散射 (US) 假设和广义平稳 (WSS) 假设是双重的：非相关散射假设定义散射体的不

① 狭义平稳 (严平稳) 指任意阶的衰落统计量都不随时间变化。对于高斯信道，广义平稳即狭义平稳。

② 证明作为练习留给读者。

同时延之间是不相关的,而广义平稳假设散射体的不同多普勒频移之间是不相关的。另外,我们认为非相关散射假设的 R_H 只与频率差有关,而广义平稳假设的 R_H 仅与时间差有关。

自然,结合这两种定义就是 WSSUS 条件,因此,自相关函数必须满足如下条件:

$$R_H(t, t + \Delta t, \tau, \tau') = P_H(\Delta t, \tau) \delta(\tau - \tau') \quad (6.29)$$

$$R_H(t, t + \Delta t, f, f + \Delta f) = R_H(\Delta t, \Delta f) \quad (6.30)$$

$$R_S(\nu, \nu', \tau, \tau') = P_S(\nu, \tau) \delta(\nu - \nu') \delta(\tau - \tau') \quad (6.31)$$

$$R_B(\nu, \nu', f, f + \Delta f) = P_B(\nu, \Delta f) \delta(\nu - \nu') \quad (6.32)$$

与取决于 4 个变量的自相关函数不同,公式右边的 P 函数仅取决于两个变量。在进一步的推导中可以大大简化其形式化描述、参数和应用。由于它们的重要性,因此定义不同的名字。根据 Kattenbach[1997],定义:

- $P_h(\Delta t, \tau)$ 为时延互功率谱密度;
- $R_H(\Delta t, \Delta f)$ 为时频相关函数;
- $P_s(\nu, \tau)$ 为散射函数;
- $P_B(\nu, \Delta f)$ 为多普勒互功率谱密度。

散射函数有特殊的重要性,因为它的物理解释比较容易。如果只有单个相互作用发生,那么散射函数的每一个微分元对应于一个物理存在的相互作用体。从多普勒频移可以确定到达方向角(DOA),时延可以确定散射存在的椭圆的半径。

WSSUS 假设非常普遍,但在实际中难以实现。附录 6. A (www.wiley.com/go/molisch) 详细讨论了该假设及其有效性。

6.4.4 抽头延迟线模型

WSSUS 信道可以用抽头延迟线模型来表示,其中与每个抽头的输出相乘的系数随时间变化。这样,冲激响应可写为

$$h(t, \tau) = \sum_{i=1}^N c_i(t) \delta(\tau - \tau_i) \quad (6.33)$$

其中, N 为抽头数, $c_i(t)$ 是抽头的取决于时间的复系数, τ_i 是第 i 个抽头的延迟。对于每一个抽头,多普勒谱确定了系数随时间的变化情况。每个抽头的谱可以是不同的,虽然许多模型认为所有抽头的多普勒谱是相同的(见第 7 章)。

抽头延迟线的一种解释可作为信道中的多径传播的物理表示。 N 个分量中的每一个对应于一组相距很近的多径分量;只有当到达信号是由来自离散的相互作用体的完全可分辨的回波组成的时,这个模型才是完全确定的。然而,在多数的实际情况下,接收机的分辨率还不能分辨所有的多径分量。因此,冲激响应可写成

$$h(t, \tau) = \sum_{i=1}^N \sum_k a_{i,k}(t) \delta(\tau - \tau_i) = \sum_{i=1}^N c_i(t) \delta(\tau - \tau_i) \quad (6.34)$$

注意,只有在带宽受限的系统中,公式中第二部分才有意义。在这种情况下,每个复系数 $c_i(t)$ 代表几条衰落多径分量之和。WSSUS 意味着所有抽头的衰落是独立的,且其功率与时间无关。

抽头延迟线的另一种解释基于采样定理。我们感兴趣的任何无线系统及其信道都是带限的。因此,冲激响应可以表示为连续冲激响应的采样形式 $\tilde{h}_{\text{in}}(t, \tau) = \sum A_\ell(t) \delta(\tau - \tau_\ell)$ 。类似地,散射函数和相关函数等都可以用其采样形式来表示。通常采样是等距离的, $\tau_\ell = \ell \cdot \Delta$

$\tilde{\tau}$, 其中的采样间隔 $\Delta\tilde{\tau}$ 由奈奎斯特定理确定。冲激响应的连续形式可以通过插值恢复为

$$h_{bl}(t, \tau) = \sum_{\ell} A_{\ell}(t) \operatorname{sinc}(W(\tau - \tau_{\ell})) \quad (6.35)$$

其中, W 为带宽。注意, 如果物理的相互作用体满足广义平稳非相关散射条件, 但不是等间隔的, 那么抽头权重 $A_{\ell}(t)$ 不必是广义平稳非相关散射的。

许多无线信道的标准模型(见第7章)都是专门为特定系统及系统带宽而开发的。为了进行离散仿真, 经常有必要将抽头的位置调整到不同的采样格, 也就是说, 离散仿真需要信道表示 $h(t, \tau) = \sum A_{\ell}(t) \delta(\tau - \ell T_s)$, 但 τ_{ℓ}/T_s 是非整数。以下的方法得到了广泛采用。

1. 取最接近的整数。这个方法导致误差, 新采样率越高, 误差越小。
2. 分割抽头能量。划分两个相邻抽头 $kT_s < \tau < (k+1)T_s$ 之间的能量, 大小可用与原始的抽头之间的距离加权。
3. 重采样。这可以通过内插公式完成, 即以需要的速率对式(6.35)中的 $h_{bl}(t, \tau)$ 重采样。另一种方法是在频域描述信道, 然后以需要的抽头间隔变换回来(采用离散傅里叶变换)。

6.5 精简参数

用相关函数来描述无线信道有些麻烦。即使采用广义平稳非相关散射假设, 仍然是两个变量的函数。一种更可取的表示应该是一个变量的函数, 甚至更好, 仅用一个参数。显然, 这样表述信道会严重丢失一些信息, 但是为了简化表述, 这样的代价有时是可接受的。

6.5.1 相关函数的积分

把两个变量变成一个变量的最直接方法就是对其中一个进行积分。对散射函数的多普勒频移 ν 进行积分, 得到时延多普勒功率谱密度 $P_h(\tau)$, 更常用的称法是功率时延谱(PDP)。功率时延谱表示在 $[\tau, \tau + d\tau]$ 内到达接收机的信号功率强度(发送单位能量 δ 脉冲时), 与多普勒频移无关。功率时延谱可以从复冲激响应 $h(t, \tau)$ 得到:

$$P_h(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |h(t, \tau)|^2 dt \quad (6.36)$$

当然, 需要具备各态历经性。注意, 在实际应用中, 积分不是对无限时间上的积分, 而是在一定的准静态有效的时间间隔内(见上文)。

类似地, 散射函数关于 τ 的积分可以得到多普勒功率谱密度 $P_B(\nu)$ 。

令时频相关函数的 $\Delta t = 0$, 就可以得到频率相关函数, 即 $R_H(\Delta f) = R_H(0, \Delta f)$ 。很明显, 频率相关函数是功率时延谱的傅里叶变换。时间相关函数 $R_H(\Delta t) = R_H(\Delta t, 0)$ 是多普勒功率谱密度的傅里叶逆变换。

6.5.2 功率时延谱的矩

功率时延谱是一个函数, 但为了快速得到测量结果的概貌, 最好将每一个测量活动都用单个参数描述。当有很多可能的参数时, 最常用的方法是功率时延谱的归一化矩。

以计算零阶矩开始, 即功率在时间上的积分:

$$P_m = \int_{-\infty}^{\infty} P_h(\tau) d\tau \quad (6.37)$$

归一化一阶矩(平均时延)为

$$T_m = \frac{\int_{-\infty}^{\infty} P_h(\tau) \tau d\tau}{P_m} \quad (6.38)$$

归一化二阶中心矩,也就是均方根时延扩展为

$$S_\tau = \sqrt{\frac{\int_{-\infty}^{\infty} P_h(\tau) \tau^2 d\tau}{P_m} - T_m^2} \quad (6.39)$$

均方根时延扩展在所有参数中有着特殊的地位。已经证明,在某些特定的环境下,由于时延色散引起的差错概率仅与均方根时延扩展成正比(见第12章),而功率时延谱的形状对其没有明显影响。在这种情况下, S_τ 即为需要知道的环境的所有信息。不得不强调,这在特定条件下是成立的,而且均方根时延扩展不是“万能钥匙”。值得注意的是,不是任何可实现物理信号的 S_τ 都是有限值。信道 $P_h(\tau) \propto 1/(1+\tau^2)$ 是物理上可实现的,也不与能量守恒相矛盾,但是 $\int_{-\infty}^{\infty} P_h(\tau) \tau^2 d\tau$ 不收敛。

例6.1 计算功率时延谱为两个尖峰的均方根时延扩展 $P_h(\tau) = \delta(\tau - 10 \mu s) + 0.3\delta(\tau - 17 \mu s)$ 。

解:

功率在时间上的积分可以由式(6.37)得到:

$$\begin{aligned} P_m &= \int_{-\infty}^{\infty} (\delta(\tau - 10^{-5}) + 0.3\delta(\tau - 1.7 \cdot 10^{-5})) d\tau \\ &= 1.30 \end{aligned} \quad (6.40)$$

平均时延可以由式(6.38)得到:

$$\begin{aligned} T_m &= \int_{-\infty}^{\infty} (\delta(\tau - 10^{-5}) + 0.3\delta(\tau - 1.7 \cdot 10^{-5})) \tau d\tau / P_m \\ &= (10^{-5} + 0.3 \cdot 1.7 \cdot 10^{-5}) / 1.3 = 1.16 \cdot 10^{-5} s \end{aligned} \quad (6.41)$$

最后,根据式(6.39)计算均方根时延扩展:

$$\begin{aligned} S_\tau &= \sqrt{\frac{\int_{-\infty}^{\infty} (\delta(\tau - 10^{-5}) + 0.3\delta(\tau - 1.7 \cdot 10^{-5})) \tau^2 d\tau}{P_m} - T_m^2} \\ &= \sqrt{\frac{(10^{-5})^2 + 0.3(1.7 \cdot 10^{-5})^2}{1.3} - (1.16 \cdot 10^{-5})^2} = 3 \mu s \end{aligned} \quad (6.42)$$

6.5.3 多普勒谱的矩

多普勒谱的矩的计算完全类似于功率时延谱的矩。积分功率为

$$P_{B,m} = \int_{-\infty}^{\infty} P_B(\nu) d\nu \quad (6.43)$$

其中,显然 $P_{B,m} = P_m$ 。平均多普勒频移为

$$\nu_m = \frac{\int_{-\infty}^{\infty} P_B(\nu) \nu d\nu}{P_{B,m}} \quad (6.44)$$

均方根多普勒扩展为

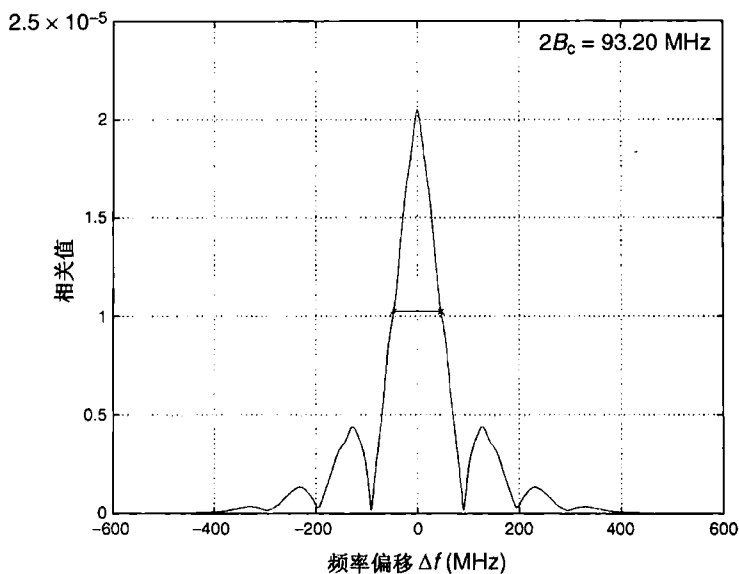
$$S_\nu = \sqrt{\frac{\int_{-\infty}^{\infty} P_B(\nu) \nu^2 d\nu}{P_{B,m}} - \nu_m^2} \quad (6.45)$$

6.5.4 相干带宽和相干时间

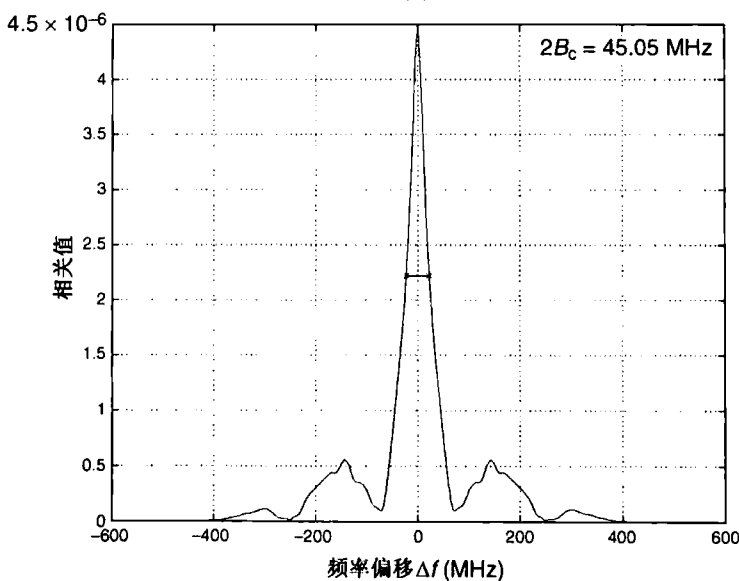
在频率选择性信道中,不同的频率分量衰落不同。显然,不同频率之间的距离越远,衰落的相关性就越小。相干带宽 B_{coh} 定义为相关系数小于一定门限的频率差。

在广义平稳非相关散射假设下, B_{coh} 的准确数学定义可以由频率相关函数 $R_H(0, \Delta f)$ 得到(见图 6.8 的例子), 相干带宽定义为

$$B_{\text{coh}} = \frac{1}{2} \left[\arg \max_{\Delta f > 0} \left(\frac{|R_H(0, \Delta f)|}{R_H(0, 0)} = 0.5 \right) - \arg \min_{\Delta f < 0} \left(\frac{|R_H(0, \Delta f)|}{R_H(0, 0)} = 0.5 \right) \right] \quad (6.46)$$



(a)



(b)

图 6.8 典型的频率相关函数。引自 Kattenbach[1997] © Shaker Verlag

这本质上是相关函数的半峰半宽带宽。这稍显复杂的公式表示源于相关函数不一定单调递减。更确切地讲,可能存在局部最大值超过门限值的情况。因此用包含了超过相关函数门限值的所有部分的带宽来准确地定义^①。

均方根时延扩展 S_r 和相干带宽 B_{coh} 显然是相关联的: S_r 是由功率时延谱 $P_h(\tau)$ 得到的,而 B_{coh} 是由频率相关函数得到的。频率相关函数是功率时延谱的傅里叶变换。基于这一认识, Fleury[1996] 导出一个“不确定性关系式”:

$$B_{coh} \geq \frac{1}{2\pi S_r}$$

(6.47)

式(6.47)是一个不等式,因此不可能从一个变量得到另一个变量。这样就引出一个问题:是 B_{coh} 还是 S_r 能更好地反映信道特性? 这个问题的答案要考虑特定的系统。对于一个没有均衡器的频分多址或时分多址系统,其均方根时延扩展是我们关心的,因为它与误比特率有关(见第12章),虽然一般来说它过于强调大的时延回波。在正交频分复用系统中(见第19章),信息在不同的并行子载波上传输,相干带宽是一个很好的度量标准。

时间相关函数是用来度量信道变化快慢的。相干时间 T_{coh} 的定义类似于相干带宽,它也与均方根多普勒扩展存在不确定性关系。

图 6.9 总结了系统函数、相关函数和特殊参数之间的关系,整个图中假设各态历经性成立。

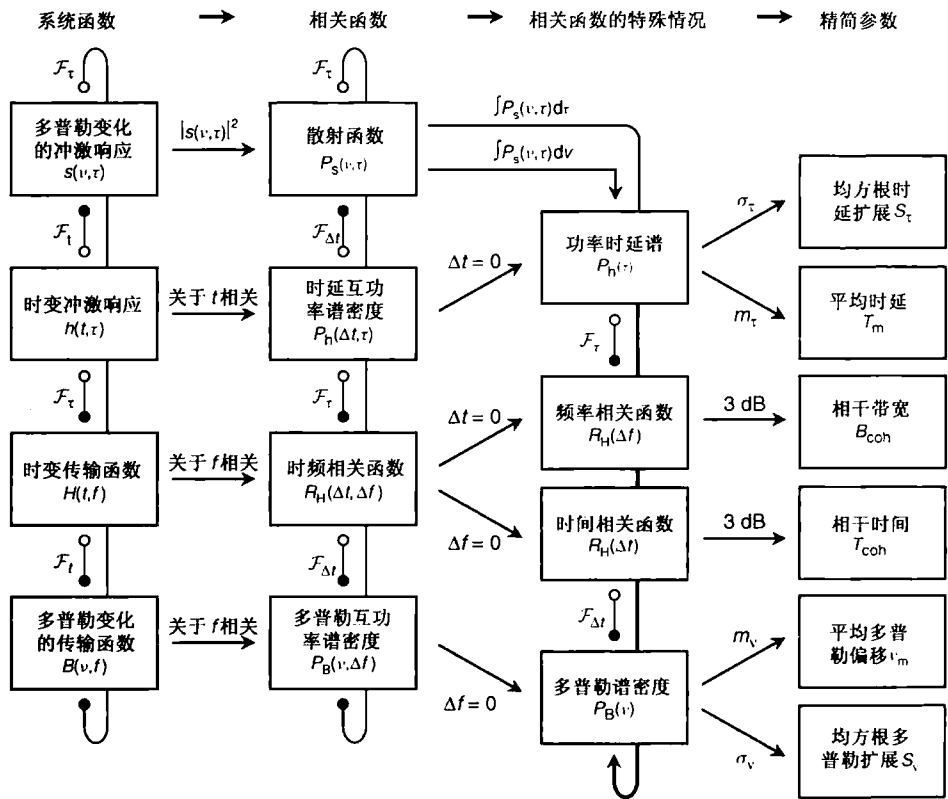


图 6.9 各态历经信道冲激响应的系统函数、相关函数和精简参数之间的关系。引自 Kattenbach[1997] © Shaker Verlag

① 另一种定义把相干带宽定义为相关函数的二阶中心矩,避开了局部最大值的所有问题。遗憾的是,这个二阶矩在一种很重要的实际情况下变为无限值,即相关函数为 Lorentzian 时,也就是 $1/(1 + \Delta f^2)$, 这对应着一个指数功率时延谱。

6.5.5 窗参数

另一组有用的参数称为窗参数[de Weck 1992],更确切地说是干扰因子 Q_T 和时延窗 W_Q 。它们用来度量在一定时延间隔内到达的平均功率时延谱占总能量的百分比。与时延扩展和相干带宽不同,窗函数需要在特定系统的上下文中进行定义。

干扰因子 Q_T 是在持续时间为 T 的窗内到达的信号功率和到达窗外的信号功率之比。时延窗表征了由于时延色散引起的自干扰特性。例如,如果一个系统有一个均衡器可以处理时延为 T 的多径分量,那么该窗内的每个多径分量都是“有用的”,而窗外的能量则会引起干扰,因此这些分量携带的比特信息就不能得到处理,从而成为不可避免的干扰^①。对于全球移动通信系统(见第24章),其规范就要求采用长度为4个符号持续时间的均衡器,也就是对应 $16\text{ }\mu\text{s}$ 的时延。因此,通常定义一个参数 Q_{16} ,即当 $T=16\text{ }\mu\text{s}$ 时的干扰因子。

干扰因子的数学定义为

$$Q_T = \frac{\int_{t_0}^{t_0+T} P_h(\tau) d\tau}{P_m - \int_{t_0}^{t_0+T} P_h(\tau) d\tau} \quad (6.48)$$

这个因子不仅与功率时延谱和持续时间 T 有关,而且与窗的起始时延 t_0 有关。通常可以通过设置起始时延为最小超出时延(即第一个决定窗的起始的多径分量) $t_0 = \tau_{\min}$ 的方式,消除与窗起始时延的依赖关系。或者,选择使得 Q_T 最大的 t_0 :

$$Q_T = \max_{t_0} \left\{ \frac{\int_{t_0}^{t_0+T} P_h(\tau) d\tau}{P_m - \int_{t_0}^{t_0+T} P_h(\tau) d\tau} \right\} \quad (6.49)$$

由于接收机可以经常调整均衡器的定时以优化性能,所以上述定义是有意义的。

另一个参数是时延窗 W_Q (见图6.10)。它定义了窗的长度应该为多长才能使窗内的功率是窗外功率的 Q 倍,其数学定义公式和干扰因子的定义相同。区别在于现在 T 是变量,而 Q 成为定值。

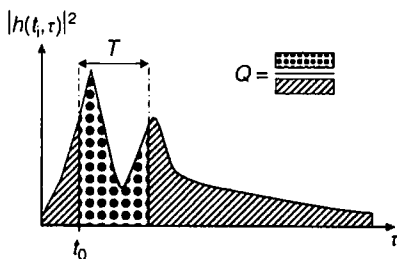


图6.10 窗参数的定义。引自 Molisch[2000]© Prentice Hall

例6.2 对于一个指数的功率时延谱, $P_h(\tau) = \exp(-\tau/2\text{ }\mu\text{s})$, 计算时延窗, 使干扰因子分别为 10 dB 和 20 dB, 即 91% 和 99% 的能量包含在窗内。对于两个尖峰谱 $\delta(\tau - 10\text{ }\mu\text{s}) + 0.3\delta(\tau - 17\text{ }\mu\text{s})$ 重新计算以上问题。

解:

令起始时延为最小超出时延。对于指数的功率时延谱, 由式(6.48)可以得到干扰因子为

^① 这只是一种近似的解释。当多径分量的时延超过均衡器的长度时, 不存在从“有用”到“干扰”的急剧“跳跃”。更确切地讲, 是一种平滑过渡, 与非均衡系统中的时间色散很相似, 关于这一点可参考第12章。

$$Q_T = \frac{\int_0^T e^{-\tau/2 \cdot 10^{-6}} d\tau}{\int_0^\infty e^{-\tau/2 \cdot 10^{-6}} d\tau - \int_0^T e^{-\tau/2 \cdot 10^{-6}} d\tau}$$

求得 T 为

$$T = 2 \cdot 10^{-6} \ln(Q_T + 1)$$

对于 91% 的窗和 99% 的窗, 有 $T_{91\%} = 4.8 \mu s$ 和 $T_{99\%} = 9.2 \mu s$ 。对于两个尖峰谱, 起始时延为 $t_0 = 10^{-5}$, 在窗内的能量为

$$\int_{10^{-5}}^{T+10^{-5}} (\delta(\tau - 10^{-5}) + 0.3\delta(\tau - 1.7 \cdot 10^{-5})) d\tau = \begin{cases} 1, & 0 < T < 7\mu s \\ 1.3, & T > 7\mu s \\ 0, & \text{其他} \end{cases}$$

因此, 对于 $T > 7 \mu s$ 的情况, 干扰因子分别大于 10 dB 和/或 20 dB。

6.6 超宽带信道

6.6.1 具有大相对带宽的超宽带信号

从多径传输引起的时延色散的角度讲, 以上的模型是宽带的。然而, 它们还基于下面的假设:

- 1. 认为相互作用体(IO)的反射、透射和绕射系数在所研究的带宽内是常数。
- 2. 系统的相对带宽(带宽除以载波频率)远小于 1。

注意, 这些条件对目前使用的大多数无线系统是满足的。然而, 超宽带(UWB)传输技术(见第 18 章)在近几年备受关注。超宽带系统的相对带宽大于 20%。这样, 发送信号中的不同频率分量能“看到”不同的传输环境。例如, 在建筑物拐角处频率为 100 MHz 的绕射系数与频率为 1 GHz 的不同。同样地, 墙壁和家具的反射系数也随频率的不同而不同。于是, 信道冲击实现表示为

$$h(\tau) = \sum_{i=1}^N a_i \chi_i(\tau) \otimes \delta(\tau - \tau_i) \tag{6.50}$$

其中, $\chi_i(\tau)$ 表示由相互作用体的频率选择性引起的第 i 条多径分量的失真。这些失真的表达式见 Molisch [2005] 和 Qiu [2002]。图 6.11 给出一个短脉冲由一个屏引起的绕射产生失真的例子。

对于超宽带系统, 传播效应也会表现出频率相关性。正如第 4 章所解释的, 如果天线的增益是常数, 那么路径损耗是频率的函数。类似地, 绕射和反射也只与频率相关。因此, 由于天线和信道的结合, 发送信号中越高的频率分量通常衰减得越严重。这种效应导致各自多径分量的失真, 因为任何频率相关的传输函数都导致时延色散, 从而引起一个多径分量的失真。

作为频率依赖失真的一个结果, 统计信道模型也需要改变。首先, 路径损耗需要根据 $G_{pr}(d, f) = E\left\{\int_{f-\Delta/2}^{f+\Delta/2} |H(f, d)|^2 df\right\}$ 重新定义, 其中数学期望包含小尺度和大尺度, Δ 为足够小的带宽, 以便其中所有传播效应保持为常数。而且, 当冲激响应表示为抽头延迟线时, 不同抽头的衰落变得相关。多径分量失真将使一个多径影响到接下来的几个抽头。这样, 一个分量的衰落影响了几个抽头的幅度, 因此引起相关性。

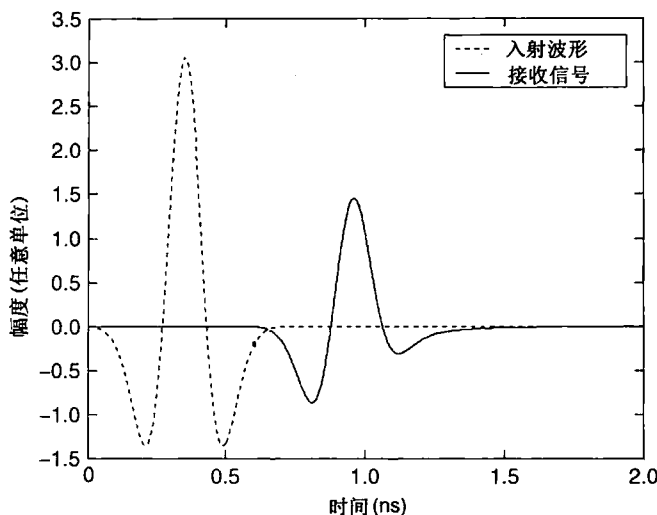


图 6.11 超宽带脉冲经一个半无限的屏绕射。引自 Qiu[2002]© IEEE

6.6.2 具有大绝对带宽的超宽带信号

超宽带信号的另一种定义是其具有超过 500 MHz 的绝对带宽。除了超宽带系统的高时间分辨率，仍存在一定的概率使得多个多径分量落在一个可分辨时延的间隔内并相加；换句话说，即使在超宽带中，也存在衰落。与传统系统的不同主要在于一个间隔内的多径分量数。该数量受环境的影响：环境中物体越多，多径分量越多。例如，住宅环境倾向于比工业环境有更少的多径分量。而且，所考虑间隔的时延起着重要作用：对大的超出时延，有更多的可能路径产生该特定时延。这样，衰落深度随着时延的增加而增加。依赖于这些因素，幅度的瑞利分布可能适用也可能不适用。建议采用 Nakagami 分布、莱斯分布或者对数正态分布。

在高绝对带宽的情况下，不是每个可分辨间隔都包含多径分量，因此包含多径分量的时延间隔不断被“空”时延间隔分开。这导致功率时延谱(PDP)是“稀疏的”。发生这些现象所需的带宽取决于环境。这样的冲激响应经常用 Saleh-Valenzuela 模型描述(详见 7.3.3 节)。

6.7 方向性描述

现在转向考虑多径分量的方向(除了它们的幅度和时延之外)的信道描述。由于如下两个原因，这样的描述是有益的。

- 对于空间分集(见第 13 章)和多元天线(见第 20 章)来说，方向性是非常重要的。
- 它容许从天线的影响中分离出传播效应。注意，在传统的宽带表示中，如前面各节所采用的，冲激响应包含了加权多径分量之和，其中权重取决于所用的天线。因此，改变了天线就改变了冲激响应，即使真正的传输信道没有改变。

由于这些原因，采用双向冲激响应(DDIR)^①是有益的，它是由多径分量的贡献之和组成的：

^① 要达到完全的一般化，应该包含关于极化情况的描述。但为了避免麻烦的矩阵符号，这里省去了这种情况，只在 7.4.4 节进行简要讨论。

$$h(t, \tau, \Omega, \Psi) = \sum_{l=1}^{N(t)} h_l(t, \tau, \Omega, \Psi) \quad (6.51)$$

DDIR 取决于特定时刻的时间 t , 时延 τ , 离开方向角(DOD) Ω , 到达方向角(DOA) Ψ , 和多径分量数 $N(t)$; 对位置的依赖没有显式写出。 $h_l(t, \tau, \Omega, \Psi)$ 是第 l 条多径分量的贡献, 可以建模为

$$h_l(t, \tau, \Omega, \Psi) = |a_l| e^{j\varphi_l} \delta(\tau - \tau_l) \delta(\Omega - \Omega_l) \delta(\Psi - \Psi_l) \quad (6.52)$$

它本质上是在抽头延迟线模型中加上方向特性。除了绝对幅度和时延, 到达方向角和离开方向角变化也很慢(跨越很多波长), 而相位 φ 变化很快。

单向冲激响应可以通过对双向冲激响应(以发送天线的模式加权)在离开方向角上的积分得到。在到达方向角上对单向冲激响应(以接收天线的模式加权)的积分可以得到传统的冲激响应。

方向性信道的随机描述与无方向信道的情况类似。冲激响应的自相关函数(ACF)可以推广为与方向相关, 因而它取决于 6 个或 8 个变量。也可以引入一种“一般化的广义平稳非相关散射条件”, 以使来自不同方向衰减的贡献是独立的。注意, 移动台处的多径分量的方向和多普勒扩展是相关联的, 因此 ν 和 Ψ 不再是独立的变量(以下假设 Ψ 为移动台处的方向)。

类似于无方向的情况, 可以定义无线信道的精简描述。首先定义:

$$E\{s^*(\Omega, \Psi, \tau, \nu) s(\Omega', \Psi', \tau', \nu')\} = P_s(\Omega, \Psi, \tau, \nu) \delta(\Omega - \Omega') \delta(\Psi - \Psi') \delta(\tau - \tau') \delta(\nu - \nu') \quad (6.53)$$

从而可以得到双向时延功率谱(DDDPS)为

$$\text{DDDPS}(\Omega, \Psi, \tau) = \int P_s(\Psi, \Omega, \tau, \nu) d\nu \quad (6.54)$$

由此, 可以建立从基站天线角度看到的角时延功率谱(ADPS):

$$\text{ADPS}(\Omega, \tau) = \int \text{DDDPS}(\Psi, \Omega, \tau) G_{MS}(\Psi) d\Psi \quad (6.55)$$

其中, G_{MS} 是移动台的天线功率模式。角时延功率谱经常归一化为

$$\int \int \text{ADPS}(\tau, \Omega) d\tau d\Omega = 1 \quad (6.56)$$

角功率谱(APS)由下式给出:

$$\text{APS}(\Omega) = \int \text{ADPS}(\Omega, \tau) d\tau \quad (6.57)$$

注意, 在 Ω 上对角时延功率谱积分可以恢复原来的功率时延谱。

如果所有的多径分量在水平面入射, 以至于 $\Omega = \phi$, 则方位角扩展定义为角功率谱的二阶中心矩。在许多论文中, 它的定义类似于式(6.39)的形式, 即

$$S_\phi = \sqrt{\frac{\int \text{APS}(\phi) \phi^2 d\phi}{\int \text{APS}(\phi) d\phi} - \left(\frac{\int \text{APS}(\phi) \phi d\phi}{\int \text{APS}(\phi) d\phi}\right)^2} \quad (6.58)$$

然而, 由于方位角的周期性使得它的定义是模糊的: 根据定义, $\text{APS} = \delta(\phi - \pi/10) + \delta(\phi - 19\pi/10)$ 将有一个不同的角度扩展形式 $\text{APS} = \delta(\phi - 3\pi/10) + \delta(\phi - \pi/10)$, 即使这两个角功率谱仅仅相差一个常数偏差。在 Fleury[2000]中有更好的定义:

$$S_\phi = \sqrt{\frac{\int |\exp(j\phi) - \mu_\phi|^2 \text{APS}(\phi) d\phi}{\int \text{APS}(\phi) d\phi}} \quad (6.59)$$

其中,

$$\mu_\phi = \frac{\int \exp(j\phi) \text{APS}(\phi) d\phi}{\int \text{APS}(\phi) d\phi} \quad (6.60)$$

例 6.3 对 $0^\circ < \phi < 90^\circ$ 和 $340^\circ < \phi < 360^\circ$, 考虑将角功率谱定义为 $\text{APS} = 1$, 分别根据式(6.58)和式(6.59)的定义计算角度扩展。

解:

根据式(6.58)有

$$S_\phi = \sqrt{\frac{\int_0^{\pi/2} \phi^2 d\phi + \int_{17\pi/9}^{2\pi} \phi^2 d\phi}{\int_0^{\pi/2} d\phi + \int_{17\pi/9}^{2\pi} d\phi} - \left(\frac{\int_0^{\pi/2} \phi d\phi + \int_{17\pi/9}^{2\pi} \phi d\phi}{\int_0^{\pi/2} d\phi + \int_{17\pi/9}^{2\pi} d\phi} \right)^2} \quad (6.61)$$

$$= 2.09 \text{ rad} = 119.7^\circ$$

另一方面, 从式(6.59)和式(6.60)可得

$$\left. \begin{aligned} \mu_\phi &= \frac{18}{11\pi} \int_{-\pi/9}^{\pi/2} \exp(j\phi) d\phi = 0.7 + 0.49j \\ S_\phi &= \sqrt{\frac{18}{11\pi} \int_{-\pi/9}^{\pi/2} (\cos(\phi - \text{Re}(\mu_\phi))^2 + (\sin(\phi - \text{Im}(\mu_\phi)))^2) d\phi} \\ &= 0.521 \text{ rad} = 29.9^\circ \end{aligned} \right\} \quad (6.62)$$

这两种方法得到的结果完全不同。很容易看出第二个值 29.9° 更有道理: 角功率谱是 -20° 到 90° 之间的连续值。一个角功率谱范围为 0° 到 111° 的扩展也应该与之相同。把修正的角功率谱代入式(6.58)中, 可得到角扩展为 32° , 而从式(6.59)得到的值仍为 29.9° 。有意思的是这个值接近于 $(\phi_{\max} - \phi_{\min})/(2\sqrt{3})$, 而对于一个矩形功率时延谱而言, 它的均方根时延扩展和最大超出时延之间的关系也为 $S_\tau = (\tau_{\max} - \tau_{\min})/(2\sqrt{3})$ 。

与时延扩展类似, 角度扩展也只是角度色散的部分描述。已经证明, 一个标准线性阵列单元处信号的相关性仅与均方根角度扩展有关, 而与角功率谱的形状无关。然而, 这只在一些特定的假设下有效。

方向性信道描述在多天系统是非常有价值的(见第20章)。这种情况下希望得到不同天线元处的联合冲激响应。如果在链路两端都有天线阵, 冲激响应就变为一个矩阵。分别表示发送元和接收元的坐标为 $\mathbf{r}_{\text{TX}}^{(1)}, \mathbf{r}_{\text{TX}}^{(2)}, \dots, \mathbf{r}_{\text{TX}}^{(N_t)}$ 和 $\mathbf{r}_{\text{RX}}^{(1)}, \mathbf{r}_{\text{RX}}^{(2)}, \dots, \mathbf{r}_{\text{RX}}^{(N_r)}$, 则从第 i 个发送元到第 j 个接收元的冲激响应为

$$\begin{aligned} h_{ij} &= h(\mathbf{r}_{\text{TX}}^{(j)}, \mathbf{r}_{\text{RX}}^{(i)}) \\ &= \sum_\ell h_\ell(\mathbf{r}_{\text{TX}}^{(1)}, \mathbf{r}_{\text{RX}}^{(1)}, \tau, \Omega_\ell, \Psi_\ell) \tilde{G}_{\text{TX}}(\Omega_\ell) \tilde{G}_{\text{RX}}(\Psi_\ell) \exp(j(\mathbf{k}(\Omega_\ell) \cdot (\mathbf{r}_{\text{TX}}^{(j)} - \mathbf{r}_{\text{TX}}^{(1)}))) \\ &\quad \times \exp(-j(\mathbf{k}(\Psi_\ell) \cdot (\mathbf{r}_{\text{RX}}^{(i)} - \mathbf{r}_{\text{RX}}^{(1)}))) \end{aligned} \quad (6.63)$$

其中, 明确写出冲激响应对位置的依赖, 利用了位置 $\mathbf{r}_{\text{TX}}^{(j)}$ 的第 ℓ 条多径分量与参考天线元位置 $\mathbf{r}_{\text{TX}}^{(1)}$ 的冲激响应通过一个相移相关这一事实。而且, 令 h_ℓ 依赖于参考天线元 $\mathbf{r}_{\text{TX}}^{(1)}$ 和 $\mathbf{r}_{\text{RX}}^{(1)}$ 的位置而不是绝对时间 t 。这里的 $\tilde{G}_{\text{TX}}$ 和 $\tilde{G}_{\text{RX}}$ 分别为发送和接收天线元的复(幅度)方向图。 $\{\mathbf{k}\}$ 是第 ℓ 个到达方向角或离开方向角方向的单位波向量。 $\langle \cdot \rangle$ 和 $\cdot$ 表示点乘。这样就总是可以由双向冲激响应来得到冲激响应矩阵。

如果接收阵是标准线性阵, 则可以把式(6.63)写为

$$\mathbf{H} = \int \int h(\tau, \Omega, \Psi) \tilde{G}_{\text{TX}}(\Omega) \tilde{G}_{\text{RX}}(\Psi) \alpha_{\text{RX}}(\Psi) \alpha_{\text{TX}}^*(\Omega) d\Psi d\Omega \quad (6.64)$$

其中用到的导向矢量为

$$\alpha_{\text{TX}}(\Omega) = \frac{1}{\sqrt{N_t}} \left[1, \exp(-j2\pi \frac{d_a}{\lambda} \sin(\Omega)), \dots, \exp(-j2\pi(N_t - 1) \frac{d_a}{\lambda} \sin(\Omega)) \right]^T$$

类似地, 可定义 $\alpha_{\text{RX}}(\Psi)$ 。 Ω 和 Ψ 是从天线宽边测量得到的。

6.8 附录

请访问网站 www.wiley.com/go/molisch。

深入阅读

线性时变系统的理论可以参阅经典文章 Bello[1963], 进一步的细节参见 Kozek[1997] 和 Matz and Hlawatsch[1998]。WSSUS 系统理论在 Bello[1963] 中建立, 其进一步的研究可以参考 Hoeher[1992], Kattenbach [1997], Molnar et al. [1996] 和 Paetzold [2002]; 无线通信中 WSSUS 的有效性的更多考虑可以参考 Fleury [1990], Kattenbach [1997], Kozek [1997] 和 Molisch and Steinbauer [1999]。刻画非 WSSUS 信道特性的一种方法在 Matz [2003] 中做了描述。Molisch and Steinbauer [1999] 综述了包括时延扩展在内的精简参数。超宽带信道的一般描述可以在 Molisch[2005], Molisch[2009] 和 Qiu[2004] 中找到。Durgin[2003], Ertel et al. [1998], Molisch[2002] 和 Yu and Ottersten [2002] 中讨论了空间信道的描述方法。将 WSSUS 方法扩展到方向性模型在 Fleury [2002] 和 Kattenbach [2002] 中进行了讨论。

第7章 信道模型

7.1 概述

我们需要建立传播信道的模型来进行无线系统的设计、仿真和规划。前面几章已经讨论了无线信道的一些基本性质及其数学描述,如幅度衰落统计、散射函数和时延扩展等。这一章更具体地讨论如何将这些数学描述转变成一般的仿真模型,以及如何确定这些模型的参数。

信道模型有如下两个主要应用。

1. 无线系统的设计、测试和定型,需要简单的信道模型反映信道传播的一些重要性质,即会影响系统性能的性质。这一般通过对以参数形式描述的冲激响应的统计特性的信道模型进行简化得到,参数个数较少,且与特定位置无关。由于信道参数和系统性能之间的闭式关系,这些模型有时能得到关于系统的内在认识。而且,这种模型对系统设计人员而言很容易实现,以达到测试的目的。
2. 无线网络的设计人员对优化某一地理区域内的特定系统比较感兴趣。基站的位置及其他网络设计参数都通过计算机而不是实地试验进行优化。对于这样的应用,需要一种能够充分利用已知地理信息和形态信息的位置特定的信道模型。但是,模型必须稳健,因为在地理数据库中可能存在小的误差。

以下是用于上述应用的3种建模方法。

1. 存储信道冲激响应。信道探测仪(见第8章)对冲激响应 $h(t, \tau)$ 测量、数字化并存储。这种方法的主要优点是所得的冲激响应比较接近实际,而且用存储数据进行系统仿真具有再生性,因为数据可以无限期地再利用,甚至可用于不同系统的仿真。这是不同于整个系统实地试验的一个重要方面,在实地试验中不能保证冲激响应随时间保持不变。使用存储冲激响应的缺点是:(i) 获取和存储数据需要大量工作;(ii) 数据只能表征某个区域,一个传播环境不必具备典型性。
2. 确定信道模型。这种模型利用数据库的地理和形态信息确定麦克斯韦方程组的解或其近似解。其基本思想和存储冲激响应相同:确定某个地理位置的冲激响应。因此这两种方法都属于特定区域的模型。与存储冲激响应(测量得到)相比,确定性信道模型(计算得到)的缺点是:(i) 计算量大;(ii) 由于原始数据的不精确性及计算方法的近似性,使得结果不是很精确。这种模型的主要优点在于计算机仿真比测量容易得多,而且某些估计方法(如射线跟踪,见7.5节)允许不同传播机理的影响彼此独立。
3. 随机信道模型。这些模型建模信道冲激响应的概率密度函数。这种方法不是要准确地预知一个特定位置的信道冲激响应,而是要预测一个大范围内的概率密度函数。最简单的例子就是瑞利衰落模型,它不是要准确地得到每个位置的场强,而是试图准确地描述大范围内场强的概率密度函数。同理可以建立随机宽带模型。

一般而言,随机模型更多地用于系统的设计和比较,而特定区域模型则更适合于网络的规

划和系统的开发。确定性和随机方法可以结合起来提高模型的效率。例如,通过确定性模型可以得到大尺度平均功率,而求平均的区域内的变化则用随机模型描述。

显然,上述模型没有一个可以达到完全精确,因此建立一个“令人满意的精确度”的标准很重要。

- 从纯科学角度看,任何误差都不能令人满意。但是,从工程角度看,模型准确度在达到一定的值之前,提高模型的精度是没有意义的^①。
- 对于确定性建模方法,原始数据的错误导致不可避免的误差。对于由测量导出的随机模型,测量点数有限及测量误差都会限制精确度。理想情况下,由特定建模方法导致的误差应该小于那些不可避免的误差。
- 以下准则可以进一步降低对建模精确度的要求:模型误差不应该“明显地”改变系统设计或开发规划。根据这个定义,系统设计人员必须确定什么是“明显的”。

7.2 窄带模型

7.2.1 小尺度和大尺度衰落模型

对于一个窄带信道,其冲激响应是一个时变衰减的函数,对于慢时变信道:

$$h(t, \tau) = \alpha(t)\delta(\tau) \quad (7.1)$$

正如第5章提到的,小范围内的幅度变化一般建模为一个随机过程,其自相关函数由多普勒谱确定。复数幅度建模为一个零均值、循环对称的复高斯随机变量。由此产生了服从瑞利分布的绝对幅度,因此简单地称这种情况为“瑞利衰落”。

当考虑更大范围内的变化时,小尺度内的平均幅度 F 服从标准差为 σ_F 的对数正态分布, σ_F 的典型值为 4 ~ 10 dB。对数正态阴影的空间自相关函数一般假设为双边指数分布,其相关距离为 5 ~ 100 m,取决于具体的环境。

7.2.2 路径损耗模型

接下来考虑对小尺度和大尺度衰落取平均的接收场强的模型。这类模型的最简单例子是自由空间路径损耗模型和“断点”模型($n=2$ 对距离 $d < d_{\text{break}}$ 有效, $n=4$ 对 $d > d_{\text{break}}$ 有效,见第4章)。在后面所描述的更复杂的模型中,路径损耗不仅取决于距离,还取决于一些附加的外部参数,如建筑物高度和测量环境(如郊区环境)等。

Okumura-Hata 模型

Okumura-Hata 模型是到目前为止应用最广泛的模型,其路径损耗(以 dB 为单位)为

$$PL = A + B \log(d) + C \quad (7.2)$$

其中, A , B 和 C 为取决于频率和天线高度的因子。 A 随着载波频率的增加而增加,随着基站和移动站高度的增加而减小。路径损耗指数(与 B 成正比)随着基站高度的增加而减小。附录

^① 很多研究论文采用“令人满意的精度”作为修饰语强调新模型的价值。如果一个方法把理论与测量值之间的偏差从 12 dB 降低到 9 dB,就会认为 9 dB 是“令人满意的”。后来的一篇论文把误差降低到 6 dB,将会认为同一个 9 dB 是“令人不满意的”。

7. A (www.wiley.com/go/molisch) 具体给出了这些修正因子。这种模型仅适用于大区制系统, 其基站高度超过周围的屋顶。

COST 231-Walfish-Ikegami 模型

COST (European COoperation in the field of Scientific and Technical research, 欧洲科技领域研究合作组织) 231-Walfish-Ikegami 模型也适合于微小区和小的宏小区, 因为它对基站和移动台之间的距离和天线高度有限制 (见图 7.1)。

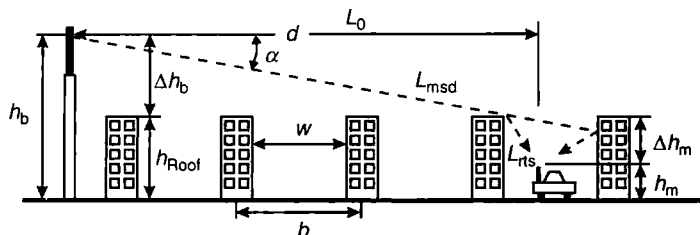


图 7.1 COST 231-Walfish-Ikegami 模型中的参数。引自 Damosso and Correia[1999] © European Union

在这个模型中, 总路径损耗包括自由空间路径损耗 PL_0 、沿着传播路径的多屏损耗 L_{msd} , 以及最后一个屋顶边沿到移动台的衰减 L_{ts} (屋顶到街道的绕射和散射损失) (见图 7.1)。自由空间损耗取决于载波频率和距离, 而屋顶到街道的绕射损耗取决于频率、街道宽度、移动台的高度, 以及街道相对于基站-移动台连线的方位。多屏损耗取决于建筑物之间的距离、基站和移动台之间的距离及载波频率、基站高度和屋顶高度。该模型假设 Manhattan 街道呈栅格状 (街道以直角相交)、建筑物高度恒定且地形平坦。而且, 该模型不包括通过街道峡谷 (street canyons) 的波导效应, 而这将会导致低估接收场强。有关该模型的细节可参见附录 7. B (www.wiley.com/go/molisch)。

Motley-Keenan 模型

对于室内环境, 墙壁的衰减起着非常重要的作用。考虑到这个因素, Motley-Keenan 模型假设路径损耗 (以 dB 表示) 可以写为 [Motley and Keenan 1988]

$$PL = PL_0 + 10n \log(d/d_0) + F_{wall} + F_{floor}$$

其中, F_{wall} 为多径分量从发送端到接收端需要穿过的墙壁所引起的衰减总和。同理, F_{floor} 表示位于基站与移动台之间的地面引起的衰减总和。墙壁的衰减与建筑物材料有关, 在 300 MHz ~ 5 GHz 频率范围内衰减可达 1 ~ 20 dB, 频率越高则衰减更大。

Motley-Keenan 模型是一种特定点模型, 因为它需要基站和移动台的位置信息及建筑物的设计图。但它并不很准确, 因为忽略了沿墙“走来走去”的传播路径。例如, 两个远远分开的办公室之间的传播或者通过很多墙壁 (准视距路径), 或者通过走廊 (信号离开办公室, 沿着走廊传播, 并进入办公室内的接收端)。后者形式的传播路径往往更加有效, 但是在 Motley-Keenan 模型中并未考虑。

7.3 宽带模型

7.3.1 抽头延迟线模型

最常采用的宽带模型是 N 抽头瑞利衰落模型。这是一个相当普遍的结构, 它基本上是第 6 章介绍的抽头延迟线结构, 只是加上了所有抽头的幅度服从瑞利衰落的限制。增加一个视距

路径分量并不困难, 冲激响应变为

$$h(t, \tau) = a_0 \delta(\tau - \tau_0) + \sum_{i=1}^N c_i(t) \delta(\tau - \tau_i) \quad (7.3)$$

其中, 视距分量 a_0 不随时间变化, $c_i(t)$ 为零均值的复高斯随机过程, 其自相关函数取决于相应的多普勒谱(如 Jakes 谱)。在大多数情况下, $\tau_0 = \tau_1$, 因此第一个抽头的幅度是服从莱斯分布的。

当抽头数限制为 $N=2$ 且没有视距路径时, 该模型进一步简化, 这是表现时延色散^①的最简单的随机衰落信道, 因此经常用于理论分析。它也可以称为 2 径信道, 2 时延信道, 或者 2 峰信道。

另一个常见的信道模型包括一个纯的确定性的视距分量加上一个衰落抽头($N=1$), 该抽头的延迟 τ_0 可以与 τ_1 不同。这个模型广泛用于卫星信道, 在这些信道中, 几乎总有一个视距连接, 且接收端附近的建筑物的反射会产生一个时延衰落分量。当 $\tau_0 = \tau_1$ 时信道变为平坦衰落的莱斯信道。

7.3.2 功率时延谱模型

通过大量的测量发现, 功率时延谱可以近似地表示为单边指数函数:

$$P_h(\tau) = P_{sc}(\tau) = \begin{cases} \exp(-\tau/S_\tau), & \tau \geq 0 \\ 0, & \text{其他} \end{cases} \quad (7.4)$$

在更一般的模型中(见 7.3.3 节), 功率时延谱是几个延迟指数函数的和, 对应于多簇相互作用体, 即

$$P_h(\tau) = \sum_l \frac{P_l^c}{S_{\tau,l}^c} P_{sc}(\tau - \tau_{0,l}^c) \quad (7.5)$$

其中, P_l^c , $\tau_{0,l}^c$ 和 $S_{\tau,l}^c$ 分别为第 l 簇的功率、时延和时延扩展, 所有簇的功率之和为 7.2 节描述的窄带功率。

对于式(7.4)的功率时延谱形式, 均方根时延扩展表征为时延色散。对于式(7.5)中的多个簇的情况, 均方根时延扩展是数学上的定义, 但经常存在有限的物理意义。现有的大多数测量还是用这个参数来表征时延色散。

不同的环境下, 时延扩展的典型值如下所示, 详细内容可参见 Molisch and Tufvesson [2004]。

- 室内住宅楼。典型值为 5 ~ 10 ns, 最大曾测到过 30 ns。
- 室内办公环境。典型时延扩展值介于 10 ~ 100 ns 之间, 但是也测到过 300 ns。房间大小对时延扩展有明显的影响, 建筑物的大小和形状也对其有影响。
- 工厂和机场大厅。时延扩展范围为 50 ~ 200 ns。
- 微小区。时延扩展范围约为 5 ~ 100 ns(视距情况)或 100 ~ 500 ns(非视距情况)。
- 隧道和矿井。空隧道的时延扩展一般比较小(20 ns 的量级), 而车辆较多的隧道的时延扩展较大(可达 100 ns)。
- 典型市区和郊区环境。时延扩展介于 100 ~ 800 ns 之间, 尽管也有过 3 μ s 这么大的值。

① 注意, 这个信道的每一个抽头表现出瑞利衰落, 因此该信道与第 5 章、第 6 章为纯粹教学目的采用的两径模型不同

- 差的市区和丘陵地区环境显然是多个簇的例子，导致其时延扩展较大，达到 $18 \mu\text{s}$ 。在欧洲各城市测量过的簇延迟高达 $50 \mu\text{s}$ ，美国城市的测量值会稍小一些，山区地形存在高达 $100 \mu\text{s}$ 的簇延迟。

时延扩展是基站与移动台之间距离的函数，随距离近似为以 d^ε 增加，其中市区和郊区环境 $\varepsilon=0.5$ ，山区 $\varepsilon=1$ 。时延扩展的变化也是很大的，有些论文指出时延扩展服从对数正态分布，在城区和郊区环境下其方差的典型值为 $2 \sim 3 \text{ dB}$ 。Greenstein et al. [1997] 首次提出了包括所有效应的完整模型。

7.3.3 射线和簇到达时间模型

前一节讨论了功率时延谱的模型，所建立的功率时延谱都是关于延迟的连续函数，这意味着接收端的带宽非常小，以至于离散的多径分量不能分离，而是成了一个连续的功率时延谱。对于带宽较大的系统，多径分量可以分离。在这种情况下，用多径分量到达时间描述功率时延谱更为有利，加上一个“包络”函数，用于描述多径分量的功率，其为时延的函数。

为了建立多径分量到达时间的统计模型，一阶近似假定在城区内引起反射的物体的空间位置是随机的，这样将使超出时延呈泊松分布。但是，测量显示多径分量倾向于分组（“簇”）到达。已经研究出两种模型来反映这种情况： Δ -K 模型和 Saleh-Valenzuela (SV) 模型。

Δ -K 模型有两个状态： S_1 和 S_2 ，前者平均到达率为 $\lambda_0(t)$ ，后者平均到达率为 $K\lambda_0(t)$ 。整个过程从状态 S_1 开始。假设一个多径分量在 t 时刻到达，在间隔 $[t, t+\Delta]$ 内转换到状态 S_2 。如果在这个间隔内没有其他路径到达，则在间隔结束时返回到 S_1 ，注意 $K=1$ 或 $\Delta=0$ 时，上述过程回到标准的泊松过程。

Saleh-Valenzuela (SV) 模型的过程略有不同。它先验假设存在簇。在每个簇内，多径分量按照泊松分布到达，簇自身的到达时间也服从泊松分布（但时间间隔常量不同）。而且，簇内多径分量的功率随时延呈指数下降，簇的功率服从另一个（不同的）指数分布（见图 7.2）。

时域冲激响应的数学描述如下：

$$h(\tau) = \sum_{l=0}^L \sum_{k=0}^K c_{k,l}(\tau) \delta(\tau - T_l - \tau_{k,l})$$

其中，簇到达时间和射线到达时间的分布如下：

$$\text{pdf}(T_l | T_{l-1}) = \Lambda \exp[-\Lambda(T_l - T_{l-1})], \quad l > 0$$

$$\text{pdf}(\tau_{k,l} | \tau_{k-1,l}) = \lambda \exp[-\lambda(\tau_{k,l} - \tau_{k-1,l})], \quad k > 0$$

其中， T_l 为第 l 个簇的第一条路径的到达时间， $\tau_{k,l}$

为第 l 个簇内的第 k 条路径相对于第一条路径到达

时间的延迟（根据定义有 $\tau_{0,l}=0$ ）， Λ 为簇到达率， λ 为射线到达率，即每个簇内路径的到达率。在上面的公式中，参数对绝对时间的依赖已经被抵消。

每个簇内的功率时延谱为

$$E[|c_{k,l}|^2] \propto P_l^c \exp(-\tau_{k,l}/\gamma) \quad (7.6)$$

其中， P_l^c 为第 l 簇的能量， γ 为簇内衰减时间常量，簇功率也呈指数下降：

$$P_l^c \propto \exp(-T_l/\Gamma) \quad (7.7)$$

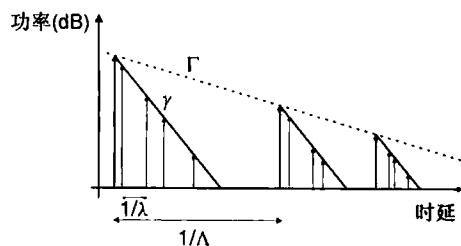


图 7.2 Saleh-Valenzuela 模型

7.3.4 标准化的信道模型

抽头延迟线模型的一个特殊情况是 COST 207 模型, 在 4 种典型环境下指定了功率时延谱或抽头权值及多普勒谱。这些功率时延谱是依据在欧洲进行的大量野外测量推导出来的。该模型分 4 种不同的大蜂窝环境: 典型市区(TU)、差的市区(BU)、乡村地区(RA)和丘陵地区(HT)。根据环境不同, 功率时延谱可能是单指数衰减或包含两个彼此有延迟的单指数函数(簇)(见图 7.3)。第二个簇对应着远处的充当有效相互作用体的高层建筑或山体, 因此会产生一组具有相当大功率时延的多径分量。详细内容可参考附录 7. C(www.wiley.com/go/molisch)。

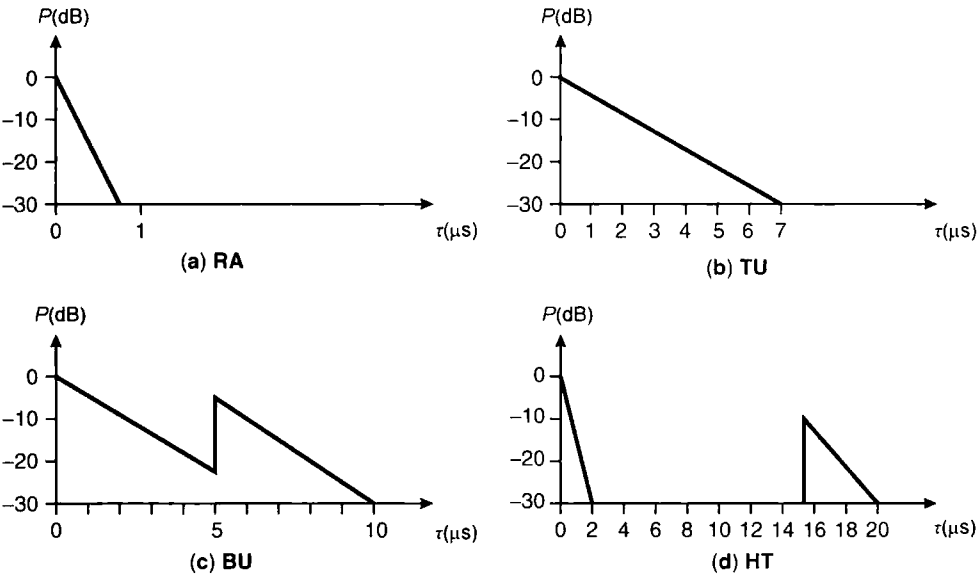


图 7.3 COST 207 功率时延谱。引自 Molisch[2000]© Prentice Hall

COST 207 模型是基于在相当窄的带宽下测量的, 且仅适于 200 kHz 带宽或更窄的系统。对于第三代蜂窝系统(带宽为 5 MHz)的仿真, 国际电信联盟指定了另一组考虑到大带宽的模型。该模型区分步行、车载和室内环境。详细内容可参考附录 7. D(www.wiley.com/go/molisch)。另外, 室内无线局域网和个域网也建立了抽头延迟线模型, 综述可参见 Molisch and Tufvesson[2004]。

7.4 方向性模型

7.4.1 一般模型结构及因子分解

第 6 章已经讨论过, 一个相当一般化的模型是基于双向时延功率谱(DDDPS)的, 它取决于 3 个变量: 离开方向角、到达方向角和时延。如果双向时延功率谱可以分解为 3 个函数, 每个函数仅取决于 1 个参数, 则可以得到一个重要的简化形式:

$$DDDPS(\Omega, \Psi, \tau) = APS^{BS}(\Omega)APS^{MS}(\Psi)P_h(\tau) \tag{7.8}$$

这表明基站和移动台处的角功率谱(APS)与时延无关, 移动台处的角功率谱与基站发送的方向也无关, 反之亦然。

这种因子分解大大简化了理论计算和信道模型参数化。然而,这不一定总与实际情况相符。更一般化的一种模型是,假设双向时延功率谱包括几个簇,每个簇有可分离的双向时延功率谱:

$$\text{DDPS}(\Omega, \Psi, \tau) = \sum_l P_l^c \text{APS}_l^{c, \text{BS}}(\Omega) \text{APS}_l^{c, \text{MS}}(\Psi) P_{h,l}^c(\tau) \quad (7.9)$$

其中,上标“c”表示“簇”, l 为第 l 个簇的编号。显然,当且仅当只有一个簇存在时,该模型变为式(7.8)所描述的模型。

在本节余下部分假设因子分解是可能的,只描述角功率谱模型 $\text{APS}_l^{c, \text{BS}}$ 和 $\text{APS}_l^{c, \text{MS}}(\Psi)$,即因子分解的组成部分(功率时延谱已经在7.3节讨论过)。

7.4.2 基站处的角度色散

基站处的角功率谱的最常用模型是其方位角服从拉普拉斯分布[Pedersen et al. 1997]:

$$\text{APS}(\phi) \propto \exp \left[-\sqrt{2} \frac{|\phi - \phi_0|}{S_\phi} \right] \quad (7.10)$$

其中, ϕ_0 是方位角的均值。仰角谱的模型一般是 δ 函数(即所有发射线在水平面入射),因此有 $\Omega = \phi$ 。也可以建模为拉普拉斯函数。

均方根角度扩展(见6.7节)和簇角度扩展的典型值如下所示[Molisch and Tufvesson 2004]。

- 室内办公环境。对于非视距情况,均方根簇角度扩展在 $10^\circ \sim 20^\circ$ 之间,视距情况下的典型值约为 5° 。
- 工业环境。非视距情况下的均方根角度扩展为 $20^\circ \sim 30^\circ$ 。
- 微小区。视距情况下的均方根角度扩展为 $5^\circ \sim 20^\circ$ 。非视距情况下的均方根角度扩展为 $10^\circ \sim 40^\circ$ 。
- 典型城区和郊区环境。测得人口密集城区的均方根角度扩展为 $3^\circ \sim 20^\circ$ 的量级。在郊区环境,由于经常出现视距情况,所以其值会小一些。
- 差的市区和丘陵地区环境。由于存在多个簇,均方根角度扩展为 20° 或者更大。
- 乡村地区。均方根角度扩展为 $1^\circ \sim 5^\circ$ 。

在室外环境,角度扩展在大区域内的分布服从对数正态分布且与时延扩展有关。这使得扩展的对数可以看成相关的高斯随机变量。角度扩展对距离的依赖关系仍需讨论。

7.4.3 移动台处的角度色散

对于室外环境,一般假设射线是从所有方位入射到移动台的,因为移动台被“本地相互作用体”(汽车、人群和房屋等)所包围。这个模型早在20世纪70年代就提出来了,但是最近的研究表明方位角的扩展可以非常小,尤其在街道峡谷处,所以角功率谱又可以近似为拉普拉斯函数,而簇角度扩展一般认为是 20° 的量级。而且,角度分布是时延的函数。对于无视距的位于街道峡谷处的移动台,屋顶上方的传播时延较小,从而有较大的角度扩展,较晚的分量沿街道传播使得角度范围更小。在存在(准)视距的室内环境,较早的分量具有很小的角度扩展,而较大时延分量有几乎均匀的角功率谱。

对于室外仰角谱,屋顶上方传播的多径分量的仰角谱服从零到能看到屋顶的最大角度之间的均匀分布。较晚到达的分量穿过街道峡谷,仰角服从拉普拉斯分布。

7.4.4 极化

大部分信道模型仅仅分析垂直极化的传播，对应于发射和接收都使用垂直极化天线的情况。但是，人们对极化分集的兴趣越来越高，即天线是同址的，但是接收波的极化方式不同。为了仿真这样的系统，需要双极化辐射的传播模型。

一个垂直极化的天线的传输会经历相互作用，导致到达接收端之前，能量会泄漏到水平极化分量，反之亦然。多径分量的衰落系数可写成 2×2 的极化矩阵，复数幅度 $\mathbf{a}_i$ 为

$$\mathbf{a}_i = \begin{pmatrix} a_i^{VV} & a_i^{VH} \\ a_i^{HV} & a_i^{HH} \end{pmatrix} \tag{7.11}$$

其中， V 和 H 分别表示垂直极化和水平极化。

最常用的极化矩阵信道模型假设矩阵中的元素是统计独立的复高斯衰落变量。假设 VV 和 HH 分量的平均功率相同，类似地， VH 和 HV 分量的平均功率也相同。互极化率 (XPD) 是 VV 和 VH 平均功率之比 (用 dB 表示)，建模为高斯随机变量。互极化率的均值和方差可能与传播环境有关，甚至还与考察分量的时延有关，其均值的典型值为 $0 \sim 12$ dB，方差约为 $3 \sim 6$ dB [Shafi et al. 2006]。

7.4.5 模型实现

前几节讨论了角度谱的连续模型，为了利用计算机实现，通常需要离散模型。方向性信道模型 (DCM) 的实现方法之一是广义抽头延迟线。与 6.4 节描述的原理相同，在这种方法中，双向时延功率谱是离散的。

另一种方法是所谓的几何随机信道模型 (GSCM)，这种方法并不是指多径分量的方向和强度具有随机性，而是指相互作用体的位置和作用过程的强度具有随机性 (见图 7.4)。另外，假设只发生一个相互作用过程。方向分解后的冲激响应可以通过如下两个步骤得到。

- 1. 根据相互作用体所在处的概率密度函数指定其位置。
- 2. 在单一作用的前提下，确定相互作用体对双向冲激响应的贡献，每个多径分量 (对应一个相互作用体) 有唯一的到达方向角、离开方向角、幅度延迟和相位偏移。

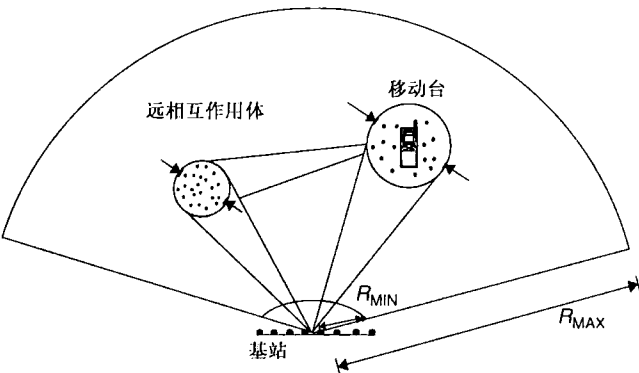


图 7.4 几何随机信道模型建模原理

最简单的模型是假设所有相关的相互作用体都在移动台附近。这种情况是存在的，例如郊区环境中建筑物结构比较规则的宏小区。在这种情况下，来自移动台的射线与其周围的相

相互作用体相互作用，但是不需要与基站附近的那些物体发生进一步相互作用。如果移动台周围的相互作用体分布不同，则模型也不同。

- 有些论文将相互作用体呈圆形分布于移动台周围，参见 Lee[1973]。
- 还有些论文建议在一个圆盘内均匀分布。当移动台移动时，移动台周围的圆盘随之移动。这样，有些相互作用体会“掉出”相互作用体圆盘，而另一些新的相互作用体则会进入圆盘(见图 7.5)。这与物理现实中远离移动台的相互作用体的贡献较小相对应(虽然它们仍然物理地存在着)。
- 也有人提出单边高斯分布 $\text{pdf}(r) = \exp(-r^2/2\sigma^2)$, $r \geq 0$ 。根据该分布计算功率时延谱和角功率谱，则可得到图 7.6 和图 7.7 的结果。可以发现这些结果与指数功率时延谱和拉普拉斯角功率谱很相似。

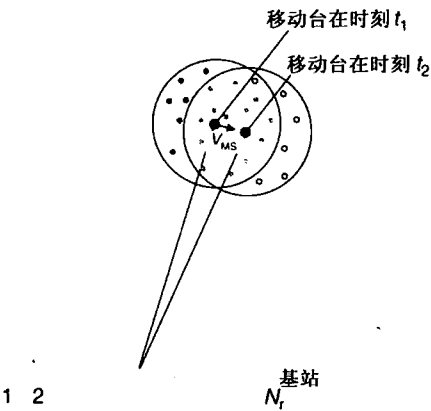


图 7.5 当移动台移动时，相互作用体的“消失”和“出现”。假定所有的相互作用体都在围绕移动台的一个圆盘内。散射体只有在 t_1 (t_2)起作用时标为黑(空)圈，散射体在两个时刻都起作用时标为灰圈。引自 Fuhl et al. [1998] © IEE

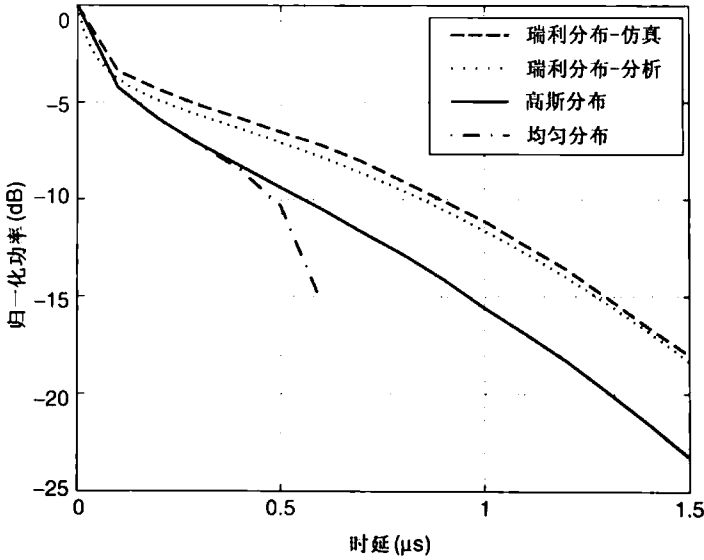


图 7.6 散射体不同分布下的功率时延谱。引自 Laurila et al. [1998] © IEEE

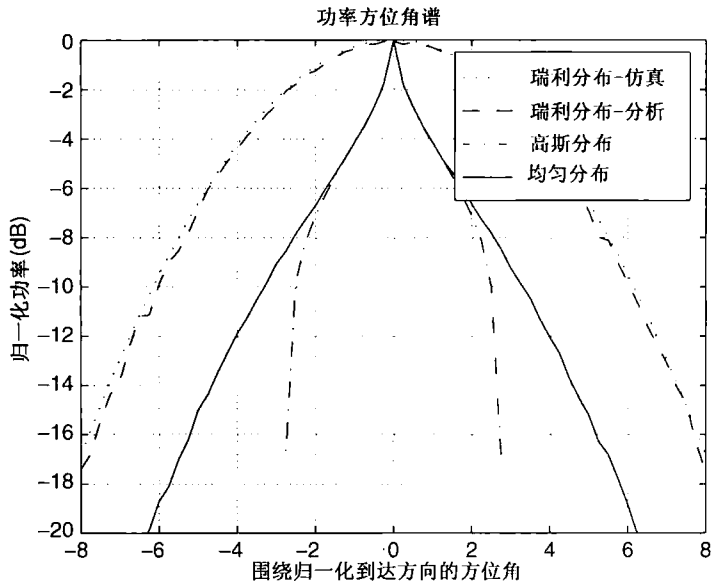


图 7.7 散射体不同分布下的角功率谱。引自 Laurila et al. [1998] © IEEE

当所有相互作用体都在移动台附近时，这种情况属于“单簇”情况，此时角度时延功率谱 (ADPS) 近似为

$$ADPS(\tau, \phi) \propto \exp(-\tau/S_\tau) \exp(-\sqrt{2}|\phi - \phi_0|/S_\phi)$$

其推广包括所谓的远相互作用体 (也称为远散射体)，对应于高层建筑或山体。一个远相互作用体可以建模为一个单面反射镜 (对应于带有光滑玻璃面的高层建筑) 或者一簇相互作用体。与移动台周围的相互作用体不同，远相互作用体的位置在整个仿真过程中是保持不变的。

几何信道模型在进行运动仿真时更有优势。无论什么时候，只要移动台移动，多径分量的参数调整都是自动进行的。因此，从移动过程中可以自动得到正确的衰落相关性，同时也考虑了移动台处到达方向角与多普勒频移之间的相关性。由移动台大尺度运动引起的到达方向角、离开方向角和时延的任何改变都会自动包括在内，而这些在抽头延迟线模型中是很难建模的。

7.4.6 标准化方向性模型

欧洲的研究倡导者 COST 259 建立了一个方向性信道模型 (DCM) 并得到普遍接受。该模型非常切合实际，结合了很多不同环境下的大量效应及其相互影响。由于该模型非常复杂，本节仅给出一些基本特征。在 Steinbauer and Molisch [2001] 中有该模型第一版的具体描述，完整的说明在 Asplund et al. [2006] 和 Molisch et al. [2006] 中可以找到。

COST 259 DCM 包括信道的小尺度变化和连续的大尺度变化。这通过区分 3 种不同层来有效实现。

- 1. 在顶层，不同无线环境 (RE)，即具有相似传播特征的环境类别 (如“典型城区”) 之间是有差别的。总而言之，一共有 13 种无线环境，其中包括 4 个宏蜂窝无线环境 (即基站高度在屋顶之上)、4 个微蜂窝 (室外，基站高度低于屋顶) 无线环境和 5 个微微蜂窝 (室内) 无线环境。
- 2. 大尺度效应由概率密度函数描述，不同的无线环境的参数也不同。例如，当移动台在大

距离内移动时,时延扩展、角度扩展、阴影和莱斯因子也随着改变。每一个大尺度衰落参数的实现确定一个双向时延功率谱。

3. 在第三层,双向冲激响应为双向时延功率谱的实现,由小尺度衰落产生。

大尺度效应的描述采用几何-随机的混合方式,并应用了上述相互作用体簇的概念。在仿真开始时,相互作用体簇(围绕移动台的一个本地簇和几个远距离的相互作用体簇)随机分布在覆盖区域内,这是随机部分。在仿真过程中,簇之间的延迟和角度可根据它们的位置及基站和移动台的位置,采用确定的方式得到,这是几何部分。每个簇有个小尺度平均的双向时延功率谱,其延迟服从指数分布,基站处的方位角和仰角服从拉普拉斯分布,移动台处的方位角和仰角服从均匀分布或拉普拉斯分布。可以直接从平均角时延迟功率谱得到双向复冲激响应,也可以将它映射到相互作用体分布,再用几何方法得到冲激响应。

在宏小区里,簇的位置是随机的。而在微小区或微微小区内,利用虚拟蜂窝部署区域(VCDA)的概念,簇的位置是确定的。一个VCDA是一幅虚拟城镇或办公楼的地图,规定了移动台的路径。这种方法与射线跟踪方法类似,但是有两个重要区别:(i)“城市地图”不需要反映实际城市,这样可以制成对很多城市具有“代表性”;(ii)只有簇的位置由射线跟踪确定,而簇内的行为采用随机的方式处理。

其他标准化的模型在“深入阅读”节及附录中描述。

7.4.7 多输入多输出矩阵模型

前面几节已经描述了包含多径分量方位信息的模型。多天线系统的另一个常用概念是用随机的方法建模多输入多输出(MIMO)信道的冲激响应矩阵(见6.7节)。在这种情况下,信道的特点不仅仅表现为矩阵元素(一般服从瑞利或莱斯分布)的幅度统计特性,还包括这些元素之间的相关性。相关矩阵(对每个延迟抽头)的定义如下:首先将信道矩阵中所有元素“堆叠”成一个矢量, $\mathbf{h}_{\text{stack}} = [h_{1,1}, h_{2,1}, \dots, h_{N_t,1}, h_{1,2}, \dots, h_{N_t,N_t}]^T$, 然后计算相关矩阵 $\mathbf{R} = E\{\mathbf{h}_{\text{stack}} \mathbf{h}_{\text{stack}}^\dagger\}$, 其中上标 $\dagger$ 表示厄米特共轭转置。一种流行的简行模型假设相关矩阵可以写成克罗内克(Kronecker)乘积 $\mathbf{R} = \mathbf{R}_{\text{TX}} \otimes \mathbf{R}_{\text{RX}}$, 其中 $\mathbf{R}_{\text{TX}} = E\{\mathbf{H}^\dagger \mathbf{H}^\star\}$, $\mathbf{R}_{\text{RX}} = E\{\mathbf{H} \mathbf{H}^\dagger\}$ 。该模型表明接收端的相关矩阵独立于传输的方向。这等价于假设双向时延功率谱可以分解为基站处和移动台处的独立的角功率谱。在这种情况下,得到信道传输函数的矩阵为

$$\mathbf{H} = \frac{1}{E\{\text{tr}(\mathbf{H} \mathbf{H}^\dagger)\}} \mathbf{R}_{\text{RX}}^{1/2} \mathbf{G}_G \mathbf{R}_{\text{TX}}^{1/2} \quad (7.12)$$

其中, $\mathbf{G}_G$ 为一个具有独立同分布(iid)复高斯元素的矩阵。

7.5 确定性信道建模方法

从原理上讲,无线传输信道可以看成确定性信道。根据麦克斯韦方程组及电磁场的边界条件(包括所有物体的位置、形状、介电和传导性质),可以确定所有点在所有时刻的场强。对于室外环境,这种纯确定性信道模型必须考虑传输环境的所有地理特征和形态特征;对于室内环境,建筑物结构、墙壁特性甚至家具都应考虑。本节在这种确定性观点基础上概述信道模型的基本原理。

为了使确定性模型成为一种可行选择,必须克服两个主要挑战:

(i) 大量的计算时间;

(ii)需要确切的边界条件知识。

- 由于计算时间和存储量极高而不能实际应用,这种状况一直持续到大约1990年。之后就发生了变化。一方面,计算机技术发展得如此之快,以至于有些在20世纪90年代就算使用超型计算机也不可能实现的任务,现在用个人计算机就能实现。另一方面,更为有效的确定性算法的发展也改善了条件。
- 边界条件的准确知识对于确定性模型的成功应用是必须的。这意味着需要知道整个“相关”环境的位置和电磁性质(后面会讨论“相关”的意思)。最近几年,在卫星图像和建筑规划的基础上产生的数字地形图、城市规划等也取得了很大的进展。

最准确的解(给定环境数据库)是麦克斯韦方程组的“强力”解,用的是积分方程或微分方程。积分方程大多是著名的矩方法的变形,相互作用体内引起的未知电流可由一组基本函数表示。其最简单的形式是矩形函数,延续一个波长的几分之一。微分方程包括有限元法(FEM)和越来越普遍的时域有限差分法(FDTD)。

所有这些方法都具有很高的精确度,但是在大多数环境下计算量需求受限制。因此更多的是使用麦克斯韦方程组的近似值作为基本解。应用最广泛的近似解是高频近似(又称射线近似)^①。在该近似中,电磁波建模为遵循几何光学准则的射线(反射和透射的斯涅尔定律)。更进一步的改进允许采用近似方法将绕射、散射也包括在内。在本节的其余部分将重点讨论基于射线方案的各种实现。

7.5.1 射线发射

在射线发射过程中,发射天线向不同方向发送(发射)射线。典型地,空间总角度 4π 可以分为 N 个相同大小的单元,每条射线的发射方向对应一个这样的单元中心(即空间角度均匀采样)(见图7.8)。射线的数目应该是该方法精确度和计算时间的折中。

该算法跟随每条射线的传播,直到它抵达接收端或者变得太弱而不重要(如降到噪声强度之下)。跟随射线时需要考虑如下很多影响。

- 自由空间衰减。由于每条射线代表某一空间角度,单位面积的能量沿射线路径按 d^{-2} 下降;
- 反射改变射线的方向,并导致额外的衰减。反射系数可以由斯涅尔定律计算得到(见第4章),与入射角有关,或许还与入射线的极化方式有关。
- 绕射和散射包含在更高级的模型中。这些情况下,一条射线入射到一个相互作用体后,会产生几条新的射线。绕射射线的幅度一般可以根据几何学和绕射理论计算得到,这在第4章中已经讲过。

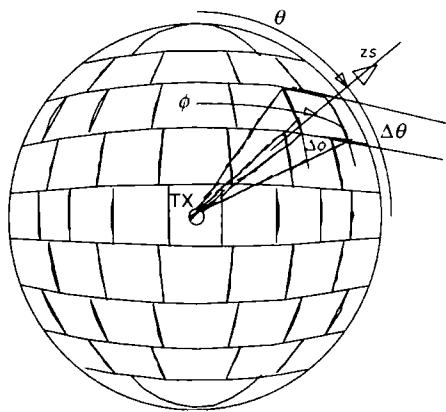


图 7.8 射线发射原理。引自 Damossio and Correia [1999] © European Union

射线分裂算法是该方法在精确度方面的一个重要改进。该算法假设射线的有效交叉部分

^① 文献中,这些方法经常称为射线跟踪。但是,“射线跟踪”也常用于表示一种特定实现方法(在后面描述)。对于一般的算法类别,我们将坚持“高频近似”的称法

不超过一定尺寸(例如典型相互作用体的尺寸)。因此,如果射线传播得离发射端太远,它可以分裂为两条射线。下面结合图 7.9 具体介绍其原理。为了简化讨论,仅考虑二维情况。每条射线并不仅代表一个角度,而是一个宽为 ϕ 的角度范围(对应于两条发射射线的角度)。该夹角与半径为 d 的圆的交叉部分的长度约为 ϕd (对于三维情况,设想用“横截面”代替“长度”)。因此,离发射端越远,射线覆盖的区域越大。为了模拟的高精确度,这个长度不能太大。当长度达到 $\bar{L}$ 时,射线就被分裂(从而长度减小为 $\bar{L}/2$),得到的子射线(重新代表一个宽为 ϕ 的角度范围)继续传输,直到其长度达到 $\bar{L}$,如此继续。

射线发射给出了整个环境(即很多不同的发射端位置和一个给定的接收端位置)的信道特征。换句话说,一旦确定了基站的位置,就可以计算出整个小区预想范围内的覆盖、时延扩展和其他信道特征。而且,预处理技术允许包括多个发射端位置。环境(相互作用体)又可以分解为“片”(大小有限的

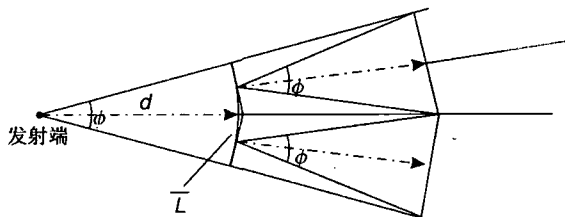


图 7.9 射线分裂原理

区域,一般与射线的最大有效面积相同),所有片的相互作用可以计算出来。那么,对于每个发射端位置,只需要计算接收端与作为最初的相互作用体的片的相互作用[Hoppe et al. 2003]。

7.5.2 射线跟踪

经典的射线跟踪决定所有可以从一个发射端位置到一个接收端位置的射线,该方法包括如下两个步骤。

1. 首先,确定可以把能量从发射端传输到接收端的所有射线。这一般通过映像原理实现。通过反射到达接收端的射线,表现出与虚拟源发出的射线相同的行为,该虚拟源位于原始源的映像(相对于反射面)所在的位置(见图 7.10)。
2. 第二步,计算出自由空间传播和有限反射系数引起的衰减,从而给出所有多径分量的参数。

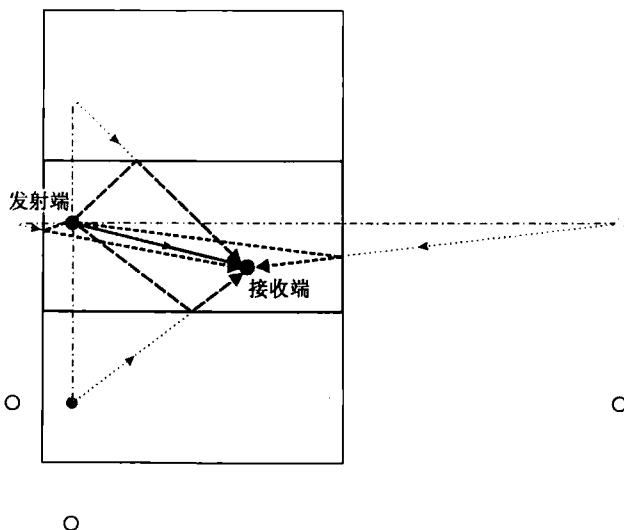


图 7.10 映像原理。灰圈对应于单个反射的虚拟源,白圈对应于双反射的虚拟源。点线代表从虚拟源到接收端的射线,虚线代表实际反射线,实线代表视距线

射线跟踪可以实现单反射和双反射过程的快速运算,也不需要射线分裂。但不利因素是,仿真中所付出的努力随着包括的反射阶数呈指数增长。而且,散射和绕射也并非无关紧要的最后,该模型对大范围内信道特征的计算不如射线发射有效。

7.5.3 有效性考虑

无论是射线发射还是射线跟踪,都不可能准确地预测到达射线的相位。这种预测需要一个在几分之一波长内准确的地理和建筑物数据库。因此,不妨假设所有射线有均匀分布的相位。在这种情况下,仅仅有可能预测出信道特征在小范围内的统计特性,冲激响应的实现可以通过将随机相位赋给多径分量得到。这是在本章开头提到过的确定-随机混合方法的另一种形式。

进一步减少计算量的方法是射线跟踪不在所有三维上进行,而仅在二维上进行。这种简化方法是否可用取决于如下传输环境。

- 室内。室内环境总是需要考虑实际三维情况,甚至当基站和移动台在同一平面上时,地板和天花板的反射也是重要的传播路径。
- 宏小区。根据定义,基站天线比屋顶高出很多。这样,传输大多通过屋顶上方到达移动台附近的点。从这些点出发经过绕射或对面房屋墙壁的反射到达移动台。有些情况下,只要垂直面的射线跟踪就足够了,尤其是当射线跟踪仅需预测接收功率和时延扩展时。另一方面,这种只有垂直射线跟踪的方法不能准确地预测移动台处射线的方向。
- 微小区,小距离基站-移动台。由于基站和移动台的天线均低于屋顶,屋顶上方传播的绕射损耗很大。穿过街道峡谷这种水平面上的传输过程更有效。在这些情况下,仅仅水平面上的射线跟踪就足够了。
- 微小区,大距离基站-移动台。在这种情况下,水平面上的射线传播(与屋顶上方部分相比)的相关功率很小,水平分量经历了多个绕射和反射过程,而通过屋顶上方的分量的损耗大部分由基站和移动台附近的绕射损耗决定,因此受距离影响较小。这种情况下,可以使用所谓的 2.5 维模型。一方面只有水平面的传播,另一方面只有垂直面的仿真,两部分的贡献加在一起。

2.5 维模型也可用于宏小区。但是,当宏小区和微小区内的基站天线与屋顶高度相近时,2.5 维的射线跟踪不能准确地为某些传播过程建模。例如,在这种方法中无法考虑一个远处的相互作用体簇(高层建筑)(见图 7.11)。

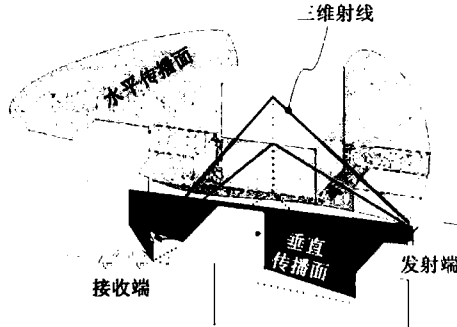


图 7.11 二维和三维建模

7.5.4 地理数据库

所有确定性方法的基础是关于环境的地理和形态信息,信息的准确性决定了确定性信道模型能达到的精确度。

对于室内环境,这些信息可以从建筑物设计图得到,现在的建筑物设计图经常可以得到其数字形式。

在乡村地区,可以得到 10 ~ 100 m 分辨率的地理数据库,这些数据库往往通过卫星观测的手段产生。在很多国家,形态信息(土地利用)也是可以得到的。但是,采用自动的和一致的方式获取该信息具有很大的挑战性。

在城区,数字数据库使用两种类型的数据:矢量数据和像素数据。对于矢量数据,建筑物端点的实际位置被存储起来。对于像素数据,有规则的栅格点叠加到区域上,并对每个像素点说明它是落入“自由空间”(街道、公园等)还是被建筑物覆盖。在两种情况下,建筑物高度和材料都可能包含在数据库中。

7.6 附录

请访问 www.wiley.com/go/molisch。

深入阅读

关于信道模型方面的文献有很多。除了文中已提到的原创性著作,Andersen et al. [1995] 和 Molisch and Tufvesson [2004] 对不同模型进行了综述。对于抽头延迟线模型的实现,我们推荐的是 Paetzold [2002]。Xu et al. [2002] 提出了将抽头延迟线模型推广到 MIMO 信道的方法。Turin et al. [1972] 研究了到达时间的泊松近似,后来 Suzuki [1977] 和 Hashemi [1979] 又做了进一步的研究和扩展。Saleh and Valenzuela [1987] 首次提出了 Saleh-Valenzuela 模型。Delta-K 模型在 Hashemi [1993] 中给出了描述。各种模型的参数化在 Molisch and Tufvesson [2004] 中进行了综述;MIMO 信道模型的参数见 Almers et al. [2007];车辆间传播信道模型的参数见 Molisch et al. [2009]。

Okumura-Hata 模型基于 Okumura et al. [1968] 在日本的大量测量,并由 Hata [1980] 提出适用于计算机仿真的形式。利用地形轮廓的扩展在 Badsberg et al. [1995] 中进行了讨论。

COST 231-Walfish-Ikegami 模型是在 Walfish and Bertoni [1988] 和 Ikegami et al. [1984] 研究的基础上,由 COST 231 [Damosso and Correia 1999] 研究与标准化组织提出的。

Blanz and Jung [1998], Fuhl et al. [1998], Norklit and Andersen [1998], Petrus et al. [2002] 和 Libert and Rappaport [1996] 以各种方式提出了几何随机信道模型。Molisch et al. [2003] 具体介绍了几何随机信道模型的有效实现,Molisch [2004] 和 Molisch and Hofstetter [2006] 进行了到多相互作用过程的推广。功率方位谱的拉普拉斯结构首先在 Pedersen et al. [1997] 中提出,虽然对其有效性存在一些争论,但它在今天还是获得了广泛的应用。

利用转移函数矩阵描述 MIMO 信道的思想来源于经典作品 Foschini and Gans [1998] 和 Winters [1987]。Kermoal et al. [2002] 提出了 Kronecker 假设;Weichselberger et al. [2006] 引入

了包含到达方向角和出发方向角之间的相关性的更一般化的模型; Gesbert et al. [2002] 和 Sayeed[2002]提出了其他广义模型。方向性信道建模方法及典型的参数化在 Almers et al. [2007]中给出综述。

矩方法在 Harrington[1993]中有经典描述。可以增加其效率的特殊方法包括自然基集 [Moroney and Cullen 1995]、快速多极算法 [Rokhlin 1990] 及列表相互作用法 [Brennan and Cullen 1998]。Zienkiewicz and Taylor[2000]中描述了有限元法, Kunz and Luebbers[1993]中描述了时域有限差分法。

射线跟踪最初来自于计算机图形学领域,且有很多可参考的书,如 Glassner[1989];它在无线中的应用在 Valenzuela[1993]中进行了描述。射线发射算法的描述可以在 Lawton and McGeehan[1994]中找到。射线分裂在 Kreuzgruber et al. [1993]中引入。也有很多利用射线跟踪进行无线信道预测的商用软件程序。

至于标准化模型,最近发展了很多包括方向信息或有更大带宽的模型。COST 259 模型在 Molisch et al. [2006b]和 Asplund et al. [2006]中描述,而 IEEE 802.11n 模型包括了室内环境的空间模型 [Erceg et al. 2004]。IEEE 802.15.3a 和 IEEE 802.15.4a 信道模型 [Molisch et al. 2003a, 2004a]描述了超宽带情况的信道模型。

另一种双向信道模型由第三代合作伙伴计划(3GPP)和 3GPP2 标准化,二者都是第三代蜂窝系统的标准化组织(见第 26 章)。该模型与 COST 259 类似,在附录及 Calcev et al. [2007]中详细描述。更进一步的双向模型由 COST 273 (Molisch and Hofstetter[2006]), 欧洲 WINNER 计划 (Winner[2007]) 及国际电信联盟 (ITU[2008]) 发布 (参见附录 7.6)。

第8章 信道探测

8.1 引言

8.1.1 信道探测的必要性

无线信道属性(冲激响应)的测量,更多地被称为信道探测(channel sounding),是无线通信工程的一项基本任务,因为任何信道模型都是基于测量数据的。对于统计信道模型,要通过大量的测试活动来获取参数值,而对于确定性模型,则必须通过理论数据与实际测量数据的比较来检验预测质量。

由于系统和所需信道模型日渐复杂,信道探测的任务也变得愈加复杂了。20 世纪 60 年代,测试设备仅用于测量接收场强。随着无线通信系统逐渐过渡到宽带系统,研发可以进行冲激响应(即时延色散)测量的一类新的信道探测器就显得非常必要了。由于人们对多天线系统的兴趣日益增长,20 世纪 90 年代,人们开始重点关注方向性传播属性的测量,这同样对信道探测器的发展带来了影响。目前,一般要求这些信道探测器必须能够测量双向冲激响应。

除了测量内容有所改变外,进行测试的环境也在发生着变化。到 1990 年为止,测试活动通常在宏蜂窝中进行。之后,微蜂窝、尤其是室内传播成为了人们兴趣的焦点。

下面讨论一些最重要的信道探测方法。在对宽带测试的基本要求进行讨论之后,将描述不同类型的信道探测器。并以对空间可分辨信道探测的概括性介绍来结束本章内容。

8.1.2 通用探测器结构

“信道探测器”一词对这一测量设备的功能给予了形象化的描述。发射机发送信号作用于信道,也就是我们所说的“探测”(sound)信道。信道的输出由接收机检测(“监听”)并存储。从发射信号和接收信号的情况可以获得信道的时变冲激响应,或者得到相应的(确定性)系统函数中的某一个。

图 8.1 给出了概念上最简单的信道探测器的框图。

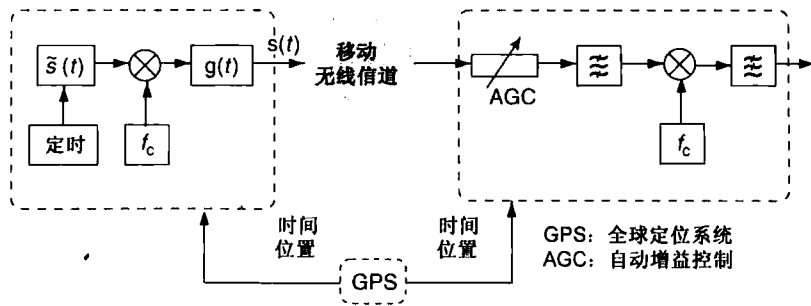


图 8.1 信道探测器原理。发射机和接收机之间的正确同步尤为重要

发射机发送一个信号 $s(t)$, 该信号由周期性重复脉冲 $p(t)$ 组成:

$$s(t) = \sum_{i=0}^{N-1} p(t - iT_{\text{rep}}) \quad (8.1)$$

其中, T_{rep} 代表发送脉冲的重复间隔。一次完整测量包括以固定间隔发送的 N 个脉冲。这些脉冲是基本脉冲 $\tilde{s}(t)$ (由脉冲发生器产生) 与发送滤波器相卷积的结果:

$$p(t) = \tilde{s}(t) * g(t) \quad (8.2)$$

其中, $g(t)$ 是发送滤波器的冲激响应。信道探测时采用的 $p(t)$ 波形因探测器的类型而定, 并且, 与工作于频域的探测器相比, 工作于时域的探测器所用的 $p(t)$ 波形可能有很大不同, 这两种类型的探测器将分别在 8.2 节和 8.3 节讨论到。

上述框图是探测器的通用形式; 探测器的性能则主要取决于探测信号的选择。为实现有效的测量, 探测信号应满足以下要求。

- 较宽的带宽。带宽与探测信号中最短的时间变化量成反比, 因而也就决定了可以达到的时延分辨率。
- 较大的时间带宽积。当探测信号的持续时间大于其带宽的倒数时 (即时间带宽积 TW 大于 1), 往往是会带来好处的。在许多系统中, 发送功率是受限制的。这种情况下, 较大的 TW 就能支持在探测信号中发送更多的能量, 从而在接收端获得较高的信噪比。大 TW 值的探测方案与扩频系统有关 (见第 18 章)。在接收端, 需要进行专门的信号处理 (解扩) 来发挥出大 TW 值所带来的好处。
- 信号持续时间。信号的有效持续时间必须适应于信道特性。一方面, 长探测信号可以给出大的时间带宽积, 这是有好处的 (如上所述)。另一方面, 探测信号的长度不能大于信道的相干时间, 即传输探测信号期间可近似认为信道恒定不变。出于实际的考虑, 脉冲重复间隔 T_{rep} 应大于探测信号的构成脉冲 $p(t)$ 的持续时间, 也应大于信道的最大附加时延。
- 功率谱密度。探测信号的功率谱密度 $|P_{\text{TX}}(j\omega)|^2$ 应在待测带宽上保持不变。这样就允许在所有频率上获得相同质量信道估计结果。出于效率的考虑, 在待测带宽之外, 所传送的信号能量应尽可能地少。
- 较低的峰值因子 (crest factor)。峰值因子的定义如下:

$$C_{\text{crest}} = \frac{\text{峰值幅度}}{\text{均方根幅度}} = \frac{\max\{|s(t)|\}}{\sqrt{s^2(t)}} \quad (8.3)$$

较低的信号峰值因子将有利于发送功率放大器高效地发挥作用。任意信号峰值因子的大致估计可以从 Felhauer et al. [1993] 中得到:

$$1 < C_{\text{crest}} \leq \sqrt{TW} \quad (8.4)$$

- 良好的相关特性。基于相关的信道探测器要求信号的自相关函数具有较高的峰值对次峰值比 (Peak to Off Peak ratio, POP), 并且信号均值为零。后一性质支持无偏估计 (见 8.4.2 节)。相关特性对于那些直接利用相关函数的信道估计方法来说至关重要, 而对基于参数的估计技术 (见 8.5 节) 而言就显得不那么重要了。

因此, 最佳数字合成探测信号的设计可以按照以下步骤进行。

1. 依据信道相干时间和所要求的时间带宽积选定探测信号的持续时间。
2. 为得到恒定的功率谱密度, 所有频率分量必须具有相同的绝对值。

3. 至此所剩下的唯一未定参数是各频率分量的相位。可以通过调整这些参数来获得较低的峰值因子。

8.1.3 无线信道的可辨识性

无线信道的时变性对于能否唯一地辨识(测量)信道具有一定的影响。

我们总可以借助适当的测量手段来识别时不变带限信道, 此时对接收机的唯一要求就是满足时延域的奈奎斯特定理[Proakis 2005], 即以足够快的速率来采样接收信号。

在时变系统中, 探测脉冲 $p(t)$ 的重复周期 T_{rep} 十分重要。任一激励脉冲 $p(t)$ 的信道响应都可以看成信道的一幅“快照”(样本)(见图 8.2)。为了跟踪信道的变化, 就应该足够频繁地获取这些快照。显然, T_{rep} 必须小于信道变化时间。通过建立相应的时域采样定理, 可将这一直观见解形式化。正如想要分辨频谱带限信号, 就必须保证一个最低采样速率, 想要辨别具有带限多普勒频谱的时变过程, 也必须保证一个最低的时间采样速率。因此, 时间采样频率必须是最大多普勒频率 ν_{max} 的两倍:

$$f_{\text{rep}} \geq 2\nu_{\text{max}} \quad (8.5)$$

将式(8.5)两边同时取倒数, 并应用移动台的移动速率和多普勒频率之间的关系式 $\nu_{\text{max}} = f_c v_{\text{max}} / c_0$ [见式(5.8)], 脉冲的重复周期可以表示成

$$T_{\text{rep}} \leq \frac{c_0}{2f_c v_{\text{max}}} \quad (8.6)$$

这样就有

$$\frac{v}{\Delta x_s} \geq 2 \frac{\nu_{\text{max}}}{\lambda_c} \quad (8.7)$$

成立, 因此需要探测的位置之间的距离 Δx_s 就具有如下的上界:

$$\Delta x_s \leq \frac{v}{\nu_{\text{max}}} \frac{\lambda}{2} \leq \frac{\lambda}{2} \quad (8.8)$$

所以, 式(8.8)表明, 要实现无混叠的测量, 每波长距离上至少需要两次快照。

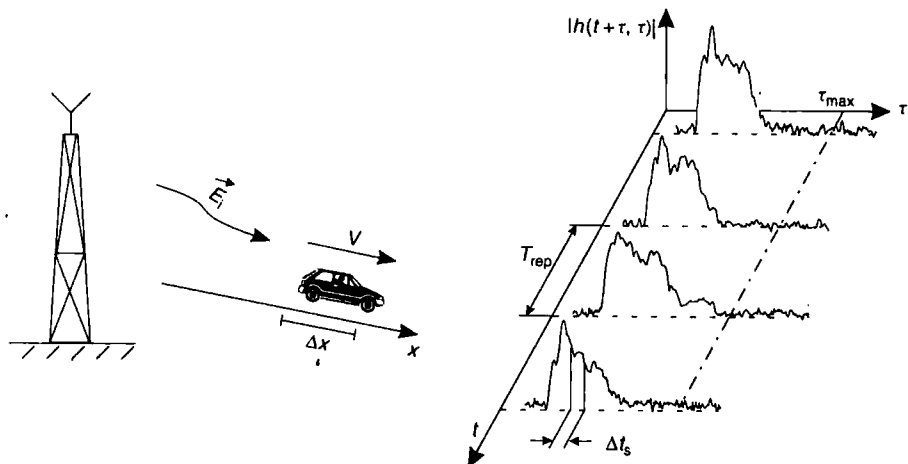


图 8.2 信道的时变冲激响应和信道的可辨识性。前一次激励的冲激响应已经消失之后, 才能获取新的快照

严重时变信道基本上是不可分辨的,因为探测信号的设计要求可能会变得有些自相矛盾。一方面,重复周期 T_{rep} 必须大于信道的最大附加时延 τ_{max} , 否则不同激励脉冲的冲激响应就会开始发生重叠。另一方面,我们刚刚指出重复周期必须满足 $T_{\text{rep}} \leq 1/2\nu_{\text{max}}$ 。因此,只有当信道参数满足:

$$2\tau_{\text{max}}\nu_{\text{max}} \leq 1 \tag{8.9}$$

才可能以某种确定方式识别信道。满足这些要求的信道称为次扩展(underspread)信道。如果不满足式(8.9),则仅在有特定假设的情况下(例如,一个定参模型),信道才是可以识别的。幸运的是,几乎所有无线信道都是次扩展的;许多情况下,甚至可以满足 $2\tau_{\text{max}}\nu_{\text{max}} \ll 1$ 的条件。后面的讨论中都假定满足次扩展条件。这同时意味着信道是慢时变的(见第6章),因此可以将 $h(t, \tau)$ 理解为在给定的(固定不变的)时刻 t 成立的冲激响应 $h(\tau)^{\text{①}}$ 。

例8.1 汽车沿街道以 36 km/h 速率行驶,车中装有信道探测器。探测器在 2 GHz 载频上测量信道的冲激响应。它必须以多大的间隔进行测量?要使信道仍然保持次扩展状态,可以有多大的最大附加时延?

解:

$$\left. \begin{aligned} v &= 36 \text{ km/h} = 10 \text{ m/s} \\ \lambda_c &= \frac{c_0}{f_c} = \frac{3 \times 10^8}{2 \times 10^9} = 0.15 \text{ m} \end{aligned} \right\} \tag{8.10}$$

必须以两倍于最大多普勒频移的最小速率在时域进行信道的采样。利用式(8.5)可得

$$f_{\text{rep}} = 2 \cdot \nu_{\text{max}} = 2 \cdot \frac{v}{\lambda_c} \tag{8.11}$$

采样间隔 T_{rep} 为

$$T_{\text{rep}} = \frac{1}{f_{\text{rep}}} = \frac{\lambda_c}{2 \cdot v} = 7.5 \text{ ms} \tag{8.12}$$

在移动台速率为 36 km/h 时,这就相当于每 75 mm 要获取一次信道快照。至于最大附加时延 τ_{max} 的计算,应用上面得到的结论:可识别信道必须是次扩展的。从而由式(8.9)可得

$$\left. \begin{aligned} 2 \cdot \tau_{\text{max}} \cdot \nu_{\text{max}} &= 1 \\ \tau_{\text{max}} &= \frac{1}{2 \cdot \nu_{\text{max}}} = T_{\text{rep}} = 7.5 \text{ ms} \end{aligned} \right\} \tag{8.13}$$

这一数值比典型无线信道中出现的最大附加时延高出几个数量级(见第7章)。然而,应该注意到,这只是仍可保证可识别性的理论最大时延扩展值。对于这样大的 τ_{max} 值,相关信道探测器的估计质量会有明显的下降。

如果 τ_{max} 更小一些,则探测脉冲的重复周期 T_{rep} 应该有所减小,这样就允许对探测样本进行平均,从而也就改善了信噪比。

8.1.4 对测量数据的影响

在进行测量时,必须意识到所测量的冲激响应中也包含不需要的成分。这些成分主要包括:

- 由同样在使用信道的其他(独立)信号源所形成的干扰;

① 严格地讲,“慢时变”概念还要求探测信号具有远小于信道相干时间的持续时间。在后面的讨论中也假定满足这一条件。

- 加性高斯白噪声。

当进行测量的环境中已经有其他无线系统在同一频率范围运作时, 尤其会引发干扰。例如, 2 GHz 范围的宽带测量就变得十分困难, 因为这些频段被不同的系统高密度地使用着。如果干扰源的数目比较大, 往往可以把总的干扰近似为高斯噪声。这一等效噪声提升了噪声基底 (noise floor), 从而降低了动态范围。

8.2 时域测量

时域测量直接测量信道的(时变)冲激响应。假定信道是慢时变的, 所测得的冲激响应等于实际信道冲激响应和探测器冲激响应的卷积:

$$h_{\text{meas}}(t_i, \tau) = \tilde{p}(\tau) * h(t_i, \tau) \quad (8.14)$$

其中, 探测器的等效冲激响应 $\tilde{p}(\tau)$ 等于发送脉冲波形和接收机滤波器冲激响应的卷积:

$$\tilde{p}(\tau) = p_{\text{TX}}(\tau) * p_{\text{RX}}(\tau) \quad (8.15)$$

上式是在假定信道和收发信机都呈线性的前提下得到的^①。探测器的冲激响应应该尽可能地接近于理想的 δ (狄拉克) 函数^②。这样可以使测量系统对结果的影响最小化。如果探测器的冲激响应不是 δ 函数, 就必须通过解卷积步骤从所测得的冲激响应中消去探测器的冲激响应, 这样做会导致噪声增强和其他额外的误差。

8.2.1 冲激探测器

与冲激雷达类似, 这一类型的信道探测器发出短脉冲序列 $p_{\text{TX}}(\tau)$ 。这些脉冲应尽可能短, 以便获得良好的空间分辨率, 但同时又应该包含尽可能多的能量, 以获得较高的信噪比。图 8.3 给出了一个粗略的草图, 它包括发送脉冲和这一脉冲经过信道传播之后的接收信号。

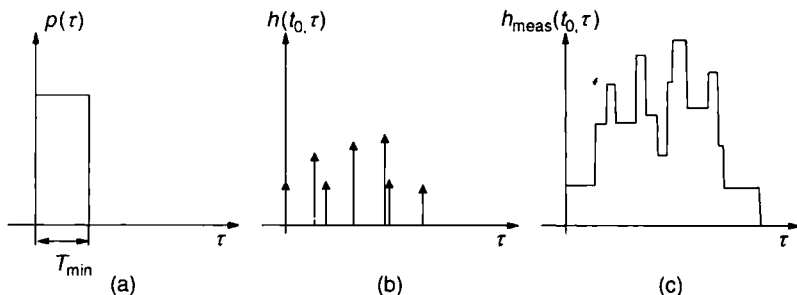


图 8.3 基于脉冲的测量原理。(a) 发送脉冲的一个样例, 整个发送序列由这样的脉冲周期性重复组成; (b) 信道的冲激响应; (c) 接收机测量到的信道输出

接收滤波器为带通滤波器, 即它在待测频率范围具有恒定幅度的频谱。我们可以理想化地认为 $p_{\text{RX}}(\tau)$ 不带来任何影响, 这样就有

$$p(\tau) = p_{\text{TX}}(\tau) \quad (8.16)$$

① 严格地讲, 由于存在像温度漂移这样的次级效应, 探测器冲激响应也是时变的。然而, 在使用过程中对探测器进行重新校准 (recalibrate) 是经常性的事, 所以, 在下次重新校准前都可将探测器的冲激响应看成时不变的。

② 这对应于在所有频率上一样平坦的频谱。在实用意义上, 频谱在待测带宽上是平坦的就足够了。

当将这个探测信号与 8.1.2 节对探测信号的要求进行对照时,就会发现这个信号具有小的时间带宽积、短的信号持续时间及高峰值因子。高脉冲能量和短的发送脉冲持续时间这两项要求,意味着这些脉冲具有非常高的峰值功率。为如此高的峰值功率所设计的放大器和其他射频部件都将是价格不菲的,可能还会伴有其他的严重缺陷,例如非线性。脉冲探测器的一个更严重的缺陷是其较低的抗干扰能力。由于这种探测器直接把接收信号当成信道的冲激响应,所以任何一种干扰信号,比如在待测频段工作的蜂窝电话所引发的干扰,都被认为是信道冲激响应的一个部分。

8.2.2 相关探测器

通过采用相关信道探测器可以增大时间带宽积。式(8.14)和式(8.15)表明,测量系统对所观测的冲激响应的影响并非单独取决于发送脉冲波形。反之,这种影响体现为 $p_{TX}(\tau)$ 和 $p_{RX}(\tau)$ 的卷积。这就为发送信号的设计提供了额外的自由度,而最终可以在低峰值因子下达到高时延分辨率。

第一个步骤就是确定所需的 $p_{TX}(\tau)$ 与 $p_{RX}(\tau)$ 之间的总体关系。正如数字通信理论中众所周知的结论:当接收滤波器是相对于发送波形的匹配滤波器时,接收滤波器的输出信噪比最大[Barry et al. 2003, Proakis 2005]①。此时,发送滤波器和接收滤波器的级联系统就具有了等于发送滤波器自相关函数的冲激响应:

$$\tilde{p}(\tau) = p_{TX}(\tau) * p_{RX}(\tau) = R_{p_{TX}}(\tau) \quad (8.17)$$

因此,探测脉冲应该具有一个能很好地逼近于 δ 函数的自相关函数;换句话说,探测脉冲应该具有高自相关峰值 $R_{p_{TX}}(0)$ 和低自相关函数旁瓣。自相关峰值和最大旁瓣高度的比值称为峰值对次峰值比(POP),这是相关探测信号的一个重要特征量。图 8.4 给出了一个自相关函数的例子,而且借助这样一个信号就可以得到时延分辨率[de Weck 1992]。

实际上,伪噪声(PN)序列或线性调频信号(chirp 信号)已成为最常采用的探测序列。最大长度伪噪声序列(m 序列)可以用带反馈的移位寄存器来产生,应用得尤其广泛。这类序列在码分多址(CDMA)系统里是众所周知的,并且在数学和通信工程文献中都已进行过深入的研究(更多细节见第 18 章)。周期为 M_c 的 m 序列的自相关函数只有唯一的相关峰,峰高为 M_c ,并且 $POP = M_c$ 。沿用 CDMA 文献中使用的术语,我们将这类序列的 M_c 个序列元素中的每一个都称为一个码片(chip)。

码片持续时间恒定时,加大 m 序列的长度可以使 POP 和时间带宽积都有所增加:信号持续时间随着 M_c 的增大而线性增加,而带宽(近似地等于码片持续时间的倒数)则保持不变。增大时间带宽积将有助于提升抗噪声和抗干扰能力。更确切地说,噪声和干扰被抑制为原来的

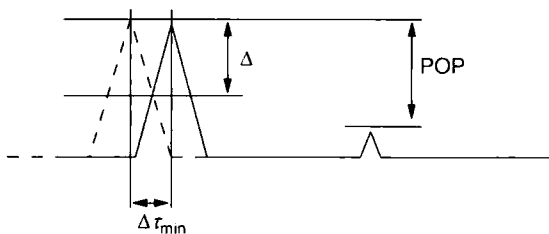


图 8.4 峰值对次峰值比和时延分辨率 $\Delta\tau_{\min}$ 的定义 ($\Delta = 6\text{dB}$)

① 严格地讲,接收滤波器与接收机天线连接器处实际收到的信号相匹配时,接收匹配滤波器的输出信噪比才能达到最大值。然而,这就要求知道信道冲激响应方面的信息,这正是我们需要测量的。因此,将滤波器与发送信号相匹配就是我们可以做到的极致了。

M_c 分之一。由于相关探测的原理与直接序列 CDMA 的原理完全一样, 获得上述性能改善的原因将在第 18 章详细讨论。

在信道时变的情况下, 对相关信道探测器的阐释要多加小心。相关信道探测器的基本原理就建立在将 $p_{TX}(\tau) * h(t, \tau) * p_{RX}(\tau)$ 与 $[p_{TX}(\tau) * p_{RX}(\tau)] * h(t, \tau)$ 视为等价的基础上。也就是说, 我们要求从伪噪声序列开始到其结束的过程中, 信道状况没有变化。这对慢时变信道而言是一个不错的近似。然而, 如果不能满足上述信道状况条件, 就必须采取一些修正措施 [Matz et al. 2002]。

8.3 频域分析

上一节描述的技术都是在时域直接估计信道的冲激响应。除此而外, 还可以尝试直接估计信道的传输函数, 即在频域进行测量。式(8.1)的基本关系仍然成立。然而, 波形 $p(t)$ 的形状在此时会有所不同。这里 $p(t)$ 的主要设计准则就是: 在待测带宽上, 探测脉冲 $p(t)$ 应具有近似不变的功率谱 $|P(j\omega)|^2$, 并能够支持直接在频域对测量结果进行解释。

频域分析的一种方法是基于线性调频(chirping)的。发送波形由下式给定:

$$p(t) = \exp \left[2\pi j \left(f_0 t + \Delta f \frac{t^2}{2T_{\text{chirp}}} \right) \right], \quad 0 \leq t \leq T_{\text{chirp}} \quad (8.18)$$

因此, 瞬时频率为

$$f_0 + \Delta f \frac{t}{T_{\text{chirp}}} \quad (8.19)$$

可以看出, 瞬时频率随时间呈线性变化, 并且其变化覆盖了整个待测频率范围(用 Δf 表示)。接收滤波器仍然是匹配滤波器。形象地说, chirp 滤波器“扫过”各个不同的频率, 在不同的时刻对不同的频率进行测量。

除此以外, 还可以同时在不同频率上探测信道。概念上最简单的方式就是先产生具有不同权值、相位和频率的若干个不同的正弦探测信号, 然后同时从发射天线将它们发送出去:

$$p(t) = \sum_{i=1}^{N_{\text{tones}}} a_i \cdot \exp[2\pi j t (f_0 + i \Delta f / N_{\text{tones}}) + j \varphi_i], \quad 0 \leq t \leq T_{\text{ss}} \quad (8.20)$$

由于成本、调校等方面的问题, $p(t)$ 的模拟产生方式(即采用多部振荡器生成多个频率)并不实用。然而, 以数字方式产生 $p(t)$ 是可以做到的, 与第 19 章描述的正交频分复用原理相似, 基带信号产生以后就可以通过唯一的一部振荡器将信号上变换至所需的通带(并且接收端的处理也是相似的)。

8.4 改进的测量方法

8.4.1 扫描时延互相关器

扫描时延互相关器(STDCC)是相关信道探测器的改进版本, 旨在降低接收端的采样速率。常规相关信道探测器要求在奈奎斯特速率上进行采样。与之不同, 扫描时延互相关器以速率 T_{rep} 进行采样, 即对于每个 m 序列, 扫描时延互相关器只采用唯一的一个采样值, 也就是在自相关函数最大值处采样。通过相对于发射端来改变接收端的时间基准, 对于 m 序列的每一次重复, 这一最大值的位置都会改变。从而, m 序列的 K_{scal} 次传输就给出了单个冲激响应 $h(\tau_i)$

的一系列采样值, $i = 1, \dots, K_{\text{scal}}$ 。所以, 时延分辨率是采样率倒数的 K_{scal} 分之一。这样做可以使采样率大幅度减低, 同时也可以大大降低对后续处理及冲激响应存储的要求。这种方法的缺点在于, 每次测量的持续时间会增大为原来的 K_{scal} 倍。

在扫描时延互相关器中, 就后续的对探测信号的多次重复而言, 自相关函数最大值位置的移动是通过在接收端和发送端采用不同的时间基准来实现的。特别地, 接收端相关器(与发送序列相比)有所延迟的时间基准是通过采用比发送码片速率小 Δf 的码片速率(码片间隔的倒数)来实现的。这样就造成了发送和接收序列之间缓慢的相对移动。在每次序列重复期间, 相关最大值对应于一个不同的时延。经过以下持续时间以后:

$$T_{\text{slip}} = \frac{1}{f_{\text{TX}} - f_{\text{RX}}} = \frac{1}{\Delta f} \quad (8.21)$$

发送和接收信号就再次完全对齐了, 即自相关函数的最大值再次出现在时延 $\tau = 0$ 处。这意味着(对静态信道而言)采样器的输出将依照所谓滑动速率(slip rate):

$$f_{\text{slip}} = f_{\text{TX}} - f_{\text{RX}} \quad (8.22)$$

呈现周期性的变化, 而比值

$$K_{\text{scal}} = \frac{f_{\text{TX}}}{f_{\text{slip}}} \gg 1 \quad (8.23)$$

就是冲激响应的时延尺度换算因子 K_{scal} 。实际冲激响应可以由测量到的采样值获得, 如下式所示:

$$\hat{h}(t_i, k\Delta\tau) = C \cdot h_{\text{STDCC}}(t_i, k\Delta\tau K_{\text{scal}}) \quad (8.24)$$

其中, C 是比例常数。

这一测量方法的缺点在于延长了测量持续时间。之前曾经指出, 只有次扩展信道(即 $2\nu_{\text{max}}\tau_{\text{max}} < 1$, 无线信道通常都能满足这一条件)才是可识别的。就扫描时延互相关器而言, 这个要求改变为 $2K_{\text{scal}}\nu_{\text{max}}\tau_{\text{max}} < 1$, 许多室外信道和 K_{scal} 的一些典型取值都不满足这一关系式。

例 8.2 假定一个扫描时延互相关器在具有 500 Hz 最大多普勒频移和 $1 \mu\text{s}$ 最大附加时延的环境中实施测量。探测器可以支持的最大采样率为 1 Msample/s。这一探测器可以达到的最大时延分辨率(带宽的倒数)是多少?

解:

采用扫描时延互相关器进行测量时, 为保证信道是可识别的(次扩展的), 必须满足如下条件:

$$2 \cdot K_{\text{scal}} \cdot \tau_{\text{max}} \cdot \nu_{\text{max}} = 1$$

$$K_{\text{scal}} = 1/(2 \cdot \tau_{\text{max}} \cdot \nu_{\text{max}}) = 1000$$

探测脉冲每重复一次, 探测器就可以得到一个采样值, 即每微秒一个样值。因此, 这一扫描时延互相关器探测器可以分辨时延为 $1 \mu\text{s}/1000 = 1 \text{ ns}$ 的多径分量。

8.4.2 逆滤波

在某些情况下, 采用虽不能与发送信号理想匹配但却能优化峰值对次峰值比的滤波器, 将是有所助益的。乍看起来, 会觉得在接收端采用不能与发送信号理想匹配的滤波器有些荒谬, 而且因此会得到较差的信噪比。然而, 我们可以为采取这种方式找到合乎实际的恰当理由。通常, 信噪比的微小变动并不比自相关函数的旁瓣重要。虽然可以通过适当的解卷积步骤来

消除旁瓣,但这些处理也会造成额外的误码。因此,针对峰值对次峰值比而不是信噪比来优化接收滤波器也是有意义的。

下面特别考虑逆滤波,并将其与匹配滤波进行比较。就匹配滤波器而言,选择 $P_{\text{TX}}^*(f)$ 作为接收滤波器的传输函数,从而总的滤波器传输函数 $P_{\text{MF}}(f)$ (发送和接收滤波器的级联)由下式给定:

$$P_{\text{MF}}(f) = P_{\text{TX}}(f) \cdot P_{\text{TX}}^*(f) \quad (8.25)$$

就逆滤波而言,在待测带宽上选择 $1/P_{\text{TX}}(f)$ 作为接收滤波器的传输函数,从而使总的传输函数尽可能地接近于1:

$$P_{\text{IF}}(f) = P_{\text{TX}}(f) \cdot \frac{1}{P_{\text{TX}}(f)} \approx 1 \quad (8.26)$$

因此,逆滤波器本质上就是一个迫零均衡器(见第16章),用来补偿由发送滤波器引起的失真。发送频谱 P_{TX} 在待测带宽上不存在任何零点的前提条件是很重要的。逆滤波器会导致噪声增强,因此其信噪比要比匹配滤波器差。好在逆滤波器是无偏的,所以估计误差的均值为零。

8.4.3 平均

通常都要对相继记录的几个传输函数的冲激响应取平均。假设在整个测量时间里信道是不变的,并且对不同测量而言噪声是统计独立的,则对 M 次测量结果取平均能使信噪比增强 $10\log_{10}M$ dB。但是,应当注意到,可测量的最大多普勒频率将减少为原来的 $1/M$ ^①。

以往人们常采用对信道的不同实现取平均的方式来获得诸如小尺度平均(SSA)功率这样的信道参数。

例8.3 一个移动台沿一条直线运动,并且每隔15 cm就能测得一个统计独立的信道冲激响应样本。测量在1.5 m的距离上进行,因此在此距离上可以认为阴影的影响是恒定不变的。目的在于,在小尺度衰落下,通过对测量结果取平均来估计信道增益(或衰减)。先假定测量仅在单一频率上进行。信道增益估计的标准偏差是多少?估计子(estimator)偏离20%以上的概率是多少?

解:

因为测量仅在单一频率上进行,所以可假定信道为平坦衰落信道且平均功率为 $\bar{P}$ 。于是,冲激响应功率的每个样本都是平均功率为 $\bar{P}$ 的指数分布随机变量。平均功率的一个合理估计为

$$\hat{\bar{P}} = \frac{1}{N} \sum_{k=1}^N P_k$$

其中, P_k 是冲激响应功率的第 k 个样本, N 是样本数。我们选择归一化标准偏差 $\sigma_{\hat{\bar{P}}}/\bar{P}$ 作为误差度量。由于各 P_k 是独立同分布的,可得

$$\frac{\sigma_{\hat{\bar{P}}}}{\bar{P}} = \frac{1}{\bar{P}} \sqrt{\text{Var}\left(\frac{1}{N} \sum_{k=1}^N P_k\right)} = \frac{1}{\bar{P}} \sqrt{\frac{1}{N^2} N \bar{P}^2} = \frac{1}{\sqrt{N}}$$

按题设情况共采用11个样本,则相对的标准偏差为0.3。假定估计子的概率密度函数近似地呈高斯型,并具有均值和方差 $\bar{P}^2/N$,则估计子偏离20%以上的概率为

① 由于总测量时间至少为原来的 M 倍。——译者注

$$1 - \Pr(0.8\bar{P} < \hat{\bar{P}} < 1.2\bar{P}) = 2 Q(0.2\sqrt{N}) \quad (8.27)$$

当 $N=11$ 时, 概率为 0.5。

例 8.4 现在考虑宽带测量的情况, 测量在 10 个衰落相互独立的频率上进行。结果将如何变化?

解:

具体测量的目标仍然是窄带信道衰减(小尺度衰落下, 通过对测量结果的平均来获得), 所谓宽带测量仅意味着现在要进行 $N=110$ 次测量。将 $N=110$ 代入式(8.27), 就会发现, 偏离 20% 以上的概率此时会减小至 0.036。

8.4.4 同步

发射和接收的同步是无线信道探测的关键性问题。这就要求在距离可能达几千米的一对发射机和接收机之间建立频率和时间的同步。由于存在多径传播和信道的时变性, 这项工作变得愈加困难。目前采用的同步方式如下所示。

1. 室内环境中, 可以采用线缆实现同步。对于距离达到大约 10 m 的情形, 可以用同轴电缆; 对于更长的距离, 光纤更为合适。两种情况下, 同步信号均是在属性确知的媒体上从发射端传送到接收端。
2. 就许多室外环境而言, 全球定位系统(GPS)为建立共同的时间和频率参考提供了解决方案。信道探测器所需的参考信号是 GPS 卫星发送信号的当然组成部分。这种方式的一个额外好处在于测量位置也是被自动记录的。缺点则在于要求发射和接收两端都具有到 GPS 卫星的视距连接; 这一要求在微蜂窝和室内应用场合很少能够满足。
3. 发射和接收端的铷(Rubidium)钟是 GPS 信号的一种替代方案。它们可以在一次测量活动开始时实现同步; 由于铷钟极其稳定(通常只有 10^{-11} 的相对漂移), 同步保持时间可达几小时。
4. 不存在同步信号的测量。借助无线链路自身也可以实现同步, 也就是在接收端当接收信号超过特定门限时, 自动触发信号的记录过程。这一技术的优点在于简单易行。然而, 它却有以下两方面缺陷: (i) 噪声或干扰可能错误地触发接收机; (ii) 无法确定绝对时延。

8.4.5 矢量网络分析仪测量

8.2 节和 8.3 节所描述的测量技术(包括频域测量技术在内)都需要可能十分昂贵的专用设备, 特别是由于生成短脉冲或短码片持续时间序列都需要高速器件支持。另有一种基于频域慢扫描的测量技术。下面讨论借助矢量网络分析仪进行的测量, 因为许多射频实验室中都有这些设备。

矢量网络分析仪用于测量待测设备(DUT)的 S 参数。待测设备可以是无线信道, 这种情况下参数 S_{21} 就是信道传输函数在用于激励信道的频率处的取值。通过使激励信号在整个待测频段上扫描或者步进变化, 就可以得到传输函数 $H(t, f)$ 的连续形式或者采样版本。

为减小网络分析仪自身对测量结果的影响, 必须进行背对背(back-to-back)校准。对所有类型的信道探测器来说, 都有进行校准的必要性, 而校准也的确是有效测量过程的一项基本原则。通常, 专门用于网络分析仪的校准类型是 SOLT^①(短路、开路、负载、直通)校准。这种校准方式将会设立若干个参考面, 然后去测量网络分析仪的频率响应。在后续测量期间, 网络分

① SOLT: 正确的含义应该是短路(Short)、开路(Open)、负载(Load)、直通(Through), 指用于网络分析仪校准的 4 种形式的校准件。

析仪补偿这一频率响应,因此测量到的就只是待测设备的频率响应。应当注意校准对象并不包括天线。如果把天线作为信道的一部分来考虑,就不成问题。可是,如果想消除天线的影响,就必须进行单独的天线校准,并在进行测量估算期间将其考虑在内^①。

通常,用矢量网络分析仪得到的测量结果是较为精确的,并能以直截了当的方式进行测量。可是,这种方式也存在某些严重缺陷:

- 这类测量进行得比较缓慢,其重复测量的频率一般无法超过若干赫兹。由于我们要求一次测量期间信道无明显变化,网络分析仪测量仅限于静态环境。
- 发射机和接收机常置于同一外壳内。这就限定了发射和接收天线可以分离开的距离。

这些限制决定了网络分析仪基本上只适用于室内测量。

8.5 可分辨方向测量

可分辨方向信道测量及基于这些测量的模型对于多天线系统的设计和仿真都很重要(见第6章和第7章)。在本节的前三小节将讨论如何进行接收机处的可分辨方向测量。然后,在本节结束时就会就多输入多输出测量(在链路两端都应该是可分辨方向的)来推广这些概念。

好在我们不必从头进行方向性信道探测器的设计。更确切地说,只须对现有设备进行某种巧妙组合就足以支持方向性测量。我们可以区分两种基本的测量方式:用方向性天线测量和用天线阵列测量。

- 用方向性天线测量。具高度方向性的天线被设置于接收端。这种天线则与一个“常规”信道探测器的接收机相连接。因而,接收机的输出就是信道与指向某特定方向的接收天线所构成的组合系统的冲激响应;即

$$h(t, \tau, \phi_i) = \int h(t, \tau, \phi) \bar{G}_{RX}(\phi - \phi_i) d\phi \quad (8.28)$$

此处, ϕ_i 表示接收天线方向图的最大值所指向的方向。通过调节天线指向,逐步取遍不同的 ϕ_i 值,就可以得到可分辨方向冲激响应的某种近似。这种测量的一个前提条件是,信道在整个测量持续时间内保持恒定,整个测量持续时间包括了针对所有不同 ϕ_i 的测量。当天线必须通过机械方式旋转来指向新的方向时,整个测量持续时间可达几秒甚至几分钟。方向分辨率越高,测量持续时间就越长。

- 用天线阵列测量。一个天线阵列由许多个天线单元组成,其中每个单元都具有较弱的方向性(或不具方向性),它们之间的间隔距离 d_u 大约等于一个波长。在所有这些天线单元处同时(或基本上同时)进行冲激响应的测量。冲激响应的结果向量可以直接利用(如用于分集性能的预测),或者借助恰当的信号处理技术(阵列处理)从该向量中提取出方向性冲激响应。

不同天线单元处冲激响应的测量可采用如下3种不同方式来完成(见图8.5)。

- 真实阵列。这种情况下每个接收天线单元都有一个解调器电路。因此冲激响应测量真正在所有天线单元处同时发生。其缺点在于高昂的造价,并且必须对多个解调器电路进行校准。

^① 注意,只有在各个多径分量的方向为已知的情况下,才可能校正天线的方向图(见8.5节)。

- 复用阵列。在这项技术中存在多个天线单元,但只有一个解调器电路。不同天线单元通过一个快速射频开关[Thomae et al. 2000]与解调器电路相连(传统信道探测器)。因此,接收机先测量第一个天线单元处的冲激响应,然后将开关连接到第二个天线单元,测量其冲激响应,以此类推。
- 虚拟阵列。在这项技术中,仅存在单个天线单元,该天线单元机械地从一个位置移动到另一个位置,依次测量不同天线单元处的冲激响应。

测量数据能用于信道估算的一个基本假设仍然是整个测量期间环境不发生变化。因此,“虚拟阵列”(由于必须以机械方式移动天线,这种方式的一次完整的测量过程需要几秒甚至几分钟)仅可用于静态环境。这还不包括汽车或处于运动状态的人成为明显的相互作用体的情形。在非静态环境中,复用阵列通常是在测量速度和硬件成本之间达成最好折中的方式。频率漂移和收发之间的同步丢失是一个相关的重要问题,测量持续时间越长,这些负面因素的影响就越大。

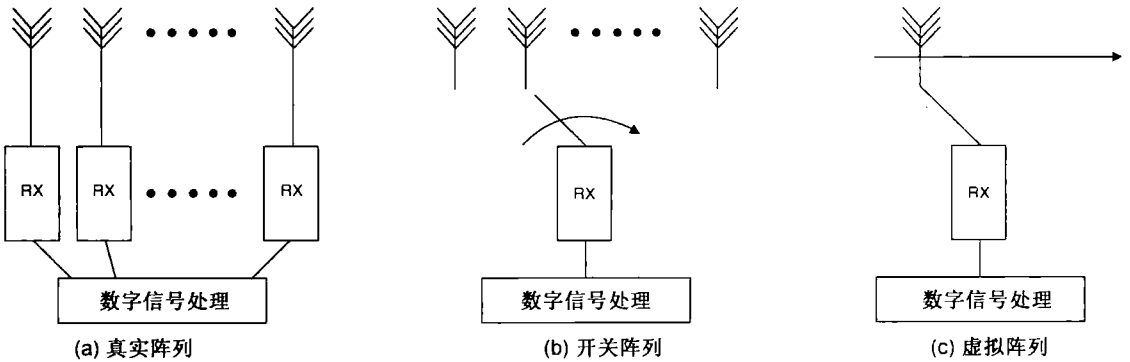


图 8.5 用于信道探测的阵列类型。引自 Molisch and Tufvesson[2005] © Hindawi

现在将问题转到如何从阵列测量中提取方向性信息。8.5.1 节描述基本数据模型,8.5.2 节和 8.5.3 节讨论信号处理方法的不同类型。

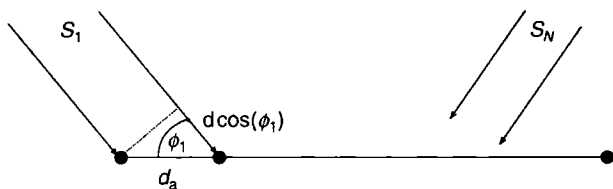
8.5.1 接收阵列的数据模型

下面建立阵列和入射信号的数学模型。对均匀线性阵列(Uniform Linear Array, ULA)的情形进行分析,该阵列包含 N_r 个单元,其中第 n 个单元处所检测的信号为 $r_n(t)$ 。为简化讨论,假定所有的波都在水平面上传播。

现在考虑当平面波从 N 个不同的方向入射到阵列时的情况(见图 8.6),这里用到达方向角 ϕ_i 来描述每个波。入射信号 s_i 和它在第一个天线单元处产生的信号之间的关系可简单表示为

$$r_1(t) = \sum_i a_{i,1} s_i(t - \tau_{i,1}) + n_1(t) \quad (8.29)$$

其中, $\tau_{i,1}$ 表示从第 i 个信号的入射源到第一个天线单元之间的传播时间, $a_{i,1}$ 是信号的(复)幅度,而 $n_1(t)$ 是第一个天线单元处的噪声。

图 8.6 以角度 $\phi_1, \dots, \phi_N$ 入射到均匀线性阵列的 N 个平面波

再来考虑第二个天线单元。该处的接收信号为

$$r_2(t) = \sum_i a_{i,2} s_i(t - \tau_{i,2}) + n_2(t) \quad (8.30)$$

如果第 i 个人射源位于远场，则有 $|a_{i,2}| = |a_{i,1}|$ ，并且

$$s_i(t - \tau_{i,2}) = s_i(t - \tau_{i,1}) \exp(-j(\tau_{i,2} - \tau_{i,1})2\pi f_c) \quad (8.31)$$

上式假定在射频意义上信号是窄带的，即信号带宽远小于载频。这一事实的物理解释就是，天线位置的唯一影响在于因额外的传播时间造成的相移。传播时间差为

$$\tau_{i,2} - \tau_{i,1} = (d_a/c_0) \cos(\phi_i) \quad (8.32)$$

因此， r_1 、 r_2 与 s 之间的关系就可表示为

$$r_1(t) = \sum_i \tilde{s}_i(t) + n_1(t) \quad (8.33)$$

$$r_2(t) = \sum_i \tilde{s}_i(t) \exp(-j2\pi d_a \cos(\phi_i)/\lambda_0) + n_2(t) \quad (8.34)$$

其中， $\tilde{s}_i(t) = a_{i,1} s_i(t - \tau_{i,1})$ 。对于下一个天线单元，可得

$$r_3(t) = \sum_i \tilde{s}_i(t) \exp(-j2\pi 2d_a \cos(\phi_i)/\lambda_0) + n_3(t) \quad (8.35)$$

根据以上这些，可得 r 和 s 之间的一般关系：

$$\mathbf{r}(t) = \mathbf{A} \mathbf{s}(t) + \mathbf{n}(t) \quad (8.36)$$

其中， $\mathbf{r}(t) = [r_1(t), r_2(t), \dots, r_{N_r}(t)]^T$ ， $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ ， $\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_{N_r}(t)]^T$ ，并且

$$\mathbf{A} = \begin{pmatrix} \exp(-jk_0 d_a \cos(\phi_1)) & \exp(-jk_0 d_a \cos(\phi_2)) & \dots & \exp(-jk_0 d_a \cos(\phi_N)) \\ \exp(-j2k_0 d_a \cos(\phi_1)) & \exp(-j2k_0 d_a \cos(\phi_2)) & \dots & \exp(-j2k_0 d_a \cos(\phi_N)) \\ \vdots & \vdots & \vdots & \vdots \\ \exp(-j(N_r-1)k_0 d_a \cos(\phi_1)) & \exp(-j(N_r-1)k_0 d_a \cos(\phi_2)) & \dots & \exp(-j(N_r-1)k_0 d_a \cos(\phi_N)) \end{pmatrix} \quad (8.37)$$

是控制矩阵 (steering matrix)。

此外，假定不同天线单元处的噪声相互独立 (空间上呈白特性)，这样噪声的相关矩阵就是主对角线上的项为 δ_n^2 的对角阵。

8.5.2 波束成形

确定入射角度的最简单方法是对信号向量 $\mathbf{r}$ 进行傅里叶变换。这种方法以阵列尺寸决定的角度分辨率来给出到达方向角 ϕ_i ，分辨率约为 $2\pi/N_r$ 。此方法的优点在于简单易行 (只须进行快速傅里叶变换)，缺点则在于较低的分辨率。

更确切地, 角度谱 $P_{\text{BF}}(\phi)$ 由下式给出:

$$P_{\text{BF}}(\phi) = \frac{\alpha^\dagger(\phi) \mathbf{R}_{\text{rr}} \alpha(\phi)}{\alpha^\dagger(\phi) \alpha(\phi)} \quad (8.38)$$

其中, $\mathbf{R}_{\text{rr}}$ 是入射信号的相关矩阵, 并且

$$\alpha_{\text{RX}}(\phi) = \begin{pmatrix} 1 \\ \exp(-jk_0 d_a \cos(\phi)) \\ \exp(-j2k_0 d_a \cos(\phi)) \\ \vdots \\ \exp(-j(N_r - 1)k_0 d_a \cos(\phi)) \end{pmatrix} \quad (8.39)$$

是方向 ϕ 上的控制向量, 可以与式(8.37)进行对照。

8.5.3 高分辨率算法

低分辨率问题可以借助所谓高分辨率方法予以消除。这些方法的分辨率不受天线阵列尺寸的限制, 而仅受限于建模误差和噪声。这些方法带来的好处是以高的计算复杂度为代价的。而且, 这些方法往往会对方向待估的多径分量的数目施加一定的限制。

高分辨率方法包括如下几种。

- **ESPRIT**。这种方法确定信号子空间, 进而以闭合形式提取出到达方向。这种算法主要适用于均匀线性阵列, 对该算法的描述将在附录 8. A 中给出 (www.wiley.com/go/molisch)。
- **多信号分类 (Multiple Signal Classification, MUSIC)**。这种算法同样需要确定信号和噪声子空间, 但随后却采用了谱搜索的方法来找到到达方向。
- **最小方差方法 (MVM: Capon 波束成形器)**。这种方法是一种纯谱搜索方法, 用来确定一个角度谱, 对于所考虑的每一个方向而言, 这个谱都能使来自其他方向的噪声和干扰的总和最小化。改进后的角度谱是易于计算的, 其表达式如下:

$$P_{\text{MVM}}(\phi) = \frac{1}{\alpha^\dagger(\phi) \mathbf{R}_{\text{rr}}^{-1} \alpha(\phi)} \quad (8.40)$$

- **入射波参数的最大似然估计**。最大似然参数提取的问题在于高计算复杂度: 空间交替广义期望最大化 (SAGE) 算法是最大似然估计的一种高效迭代实现方法。自从 2000 年起, 这种方法成为最流行的信道探测计算方法。其缺点在于迭代可能收敛于一个局部最优值, 而未必能收敛于全局最优值。

阵列校准是所有算法共同的问题。绝大多数算法都采用了关于阵列的某些假设: 所有单元的天线方向图相同, 天线单元之间不存在互耦合, 并且所有相邻天线单元之间的距离相等。如果实际阵列不满足这些假设, 就必须进行校准, 以便对测量进行适当的修正。这种校准必须经常性地反复进行, 因为温度漂移、器件老化等因素总是会对校准构成破坏。

许多高分辨率算法 (包括基于子空间的算法) 都要求相关矩阵是非奇异的。如果不同方向到达波的波源相关, 一般就会出现这种奇异的 $\mathbf{R}_{\text{rr}}$ 。对于信道探测而言, 通常所有的信号都来自于同一个源, 所以它们是完全相关的。那么, 必须采用子阵列平均 (“空间平滑” 或 “前向-后向” 平均) 来得到合乎要求的相关矩阵 [Harardt and Nosssek 1995]。子阵列平均的缺点在于减小了阵列的有效尺寸。

例8.5 3个相互独立的信号,幅度为1、0.8和0.2,入射方向分别为 10° 、 45° 和 72° 。由噪声电平引起的第一个信号的信噪比为15 dB。对于五单元线性阵列(天线单元相隔 $\lambda/2$)。先计算相关矩阵和控制向量,再画出 $P_{BF}(\phi)$ 和 $P_{MVM}(\phi)$ 。

解:

首先假定3个信号到达线性阵列的初始相位均为0。因此,对于这3个多径分量,可以得到如下参数:

$$\begin{aligned} a_1 &= 1 \cdot e^{j \cdot 0}, \quad \phi_1 = 10\pi/180 \text{ rad} \\ a_2 &= 0.8 \cdot e^{j \cdot 0}, \quad \phi_2 = 45\pi/180 \text{ rad} \\ a_3 &= 0.2 \cdot e^{j \cdot 0}, \quad \phi_3 = 72\pi/180 \text{ rad} \end{aligned} \quad (8.41)$$

假定测量噪声服从复高斯分布,且在空间上呈白特性。噪声样本的公共方差 σ_n^2 如下:

$$\sigma_n^2 = \frac{1}{10^{15/10}} = 0.032 \quad (8.42)$$

按照式(8.37),单元间隔为 $\lambda/2$ 的五单元线性阵列的控制矩阵如下:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ \exp(-j \cdot \pi \cdot \cos(\phi_1)) & \exp(-j \cdot \pi \cdot \cos(\phi_2)) & \exp(-j \cdot \pi \cdot \cos(\phi_3)) \\ \exp(-j \cdot 2 \cdot \pi \cdot \cos(\phi_1)) & \exp(-j \cdot 2 \cdot \pi \cdot \cos(\phi_2)) & \exp(-j \cdot 2 \cdot \pi \cdot \cos(\phi_3)) \\ \exp(-j \cdot 3 \cdot \pi \cdot \cos(\phi_1)) & \exp(-j \cdot 3 \cdot \pi \cdot \cos(\phi_2)) & \exp(-j \cdot 3 \cdot \pi \cdot \cos(\phi_3)) \\ \exp(-j \cdot 4 \cdot \pi \cdot \cos(\phi_1)) & \exp(-j \cdot 4 \cdot \pi \cdot \cos(\phi_2)) & \exp(-j \cdot 4 \cdot \pi \cdot \cos(\phi_3)) \end{bmatrix} \quad (8.43)$$

注意,此矩阵中的各列就是各个多径分量对应的控制向量。将到达方向角数值代入后得到

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ -0.9989 - 0.0477j & -0.6057 - 0.7957j & 0.5646 - 0.8253j \\ 0.9954 + 0.0953j & -0.2663 + 0.9639j & -0.3624 - 0.9320j \\ -0.9898 - 0.1427j & 0.9282 - 0.3720j & -0.9739 - 0.2272j \\ 0.9818 + 0.1898j & -0.8582 - 0.5133j & -0.7374 + 0.6755j \end{bmatrix} \quad (8.44)$$

由这3个入射信号确定的被观测阵列响应如下:

$$\mathbf{r}(t) = \mathbf{A} \cdot \mathbf{s}(t) + \mathbf{n}(t) \quad (8.45)$$

此处, $\mathbf{s}(t) = [1 \ 0.8 \ 0.2]^T$, $\mathbf{n}(t)$ 为加性噪声样本向量。就10 000次实现的结果(对应于彼此独立的实现^①,在它们的控制矩阵中,同一控制向量列可能具有不同的相位 ϕ_i)来计算相关矩阵 $\mathbf{R}_{rr} = E[\mathbf{r}(t)\mathbf{r}^\dagger(t)]$,从而有

$$\mathbf{R}_{rr} = \begin{bmatrix} 1.7295 & -1.3776 + 0.5941j & 0.8098 - 0.6763j & -0.4378 + 0.3755j & 0.4118 + 0.1286j \\ -1.3776 - 0.5941j & 1.7244 & -1.3621 + 0.5957j & 0.7993 - 0.6642j & -0.4275 + 0.3666j \\ 0.8098 + 0.6763j & -1.3621 - 0.5957j & 1.7080 & -1.3493 + 0.5850j & 0.7932 - 0.6599j \\ -0.4378 - 0.3755j & 0.7993 + 0.6642j & -1.3493 - 0.5850j & 1.6903 & -1.3439 + 0.5780j \\ 0.4118 - 0.1286j & -0.4275 - 0.3666j & 0.7932 + 0.6599j & -1.3439 - 0.5780j & 1.6875 \end{bmatrix} \quad (8.46)$$

普通波束成形器的角度谱由式(8.38)给出,并以虚线形式绘制于图8.7中。注意普通波束成形器无法分辨3个入射信号。MVM方法(Capon波束成形器)的角度谱由式(8.40)给定;结果以实线形式绘制于图8.7中。在实际到达角 10° 、 45° 和 72° 的附近可以分辨出3个尖峰。

① 此处假定可以达成不同的实现。正如上面讨论过的,许多信道探测应用需要采取某些辅助手段(如子阵列平均)来得到那些实现的最后结果。

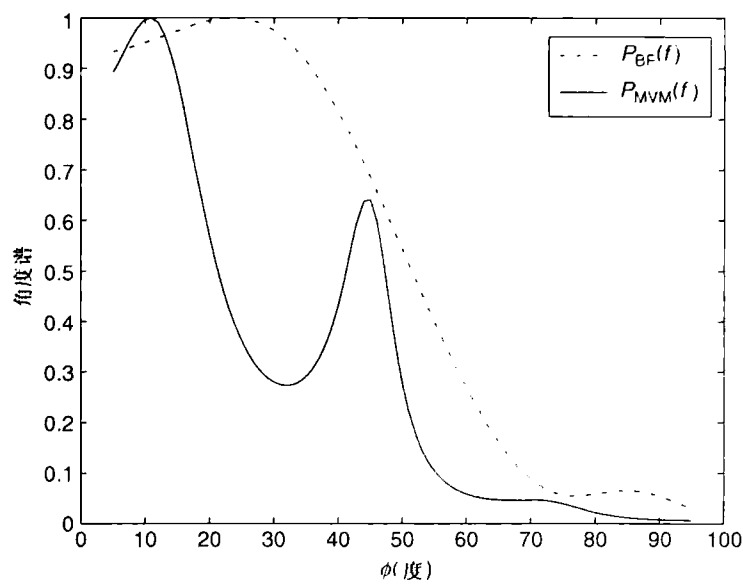


图 8.7 对于五单元线性阵列,传统波束成形器和 MVM 所得角度谱的比较 实际到达角度为 10°, 45°和 72°

8.5.4 多输入多输出测量

前面描述的方法都是为了在链路的一端获得到达方向。很容易将它们推广应用到双向测量或多输入多输出(MIMO)测量中。链路两端都可以采用天线阵列。在这种情况下,所采用的发送信号必须满足以下条件:接收机可以判断出某发送信号是由哪个天线发送的 这一点是可以做到的。例如,在不同时刻发送或以不同载频发送,或用不同码字调制后再发送信号(见图 8.8)。这些方法中,采用不同时刻发送所要求的硬件成本最低,因而在商用 MIMO 信道探测器中得到了广泛应用。判定链路两端方向性的信号处理技术也和单端情形所采用的技术相当地相似。

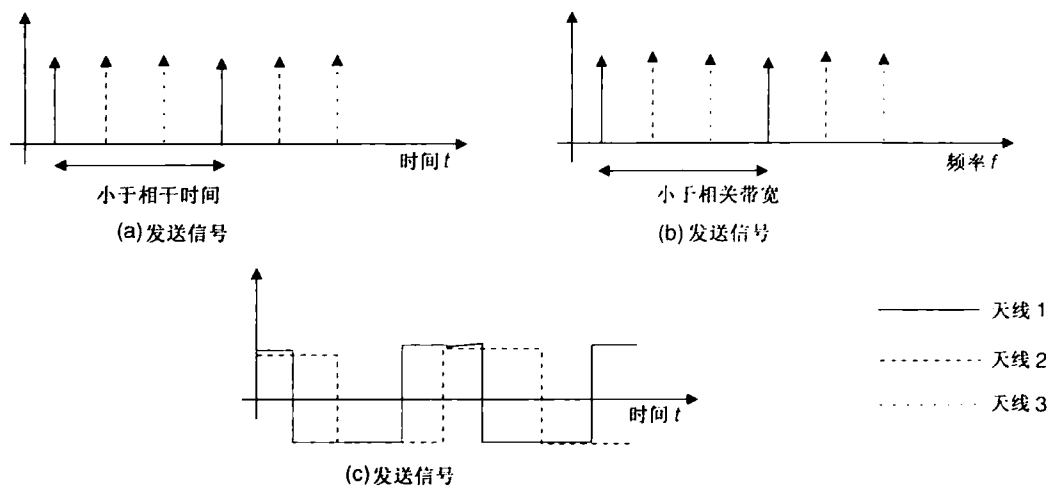


图 8.8 不同天线处探测信号的发送:信号在时间上正交(a)、在频率上正交(b)或用于调制的码字相互正交(c)。引自 Molisch and Tufvesson [2005] © Hindawi

8.6 附录

请访问 www.wiley.com/go/molisch。

深入阅读

Parsons[1992]和 Parsons et al.[1991]中对不同的信道探测技术进行了综述。Cullen et al.[1993]重点讨论相关信道探测器；在 Cox[1972]中第一次引入了扫描时延互相关器，并对这一方法进行了详尽的描述。Matz et al.[2002]讨论了信道时变性对测量结果的影响。大多数发布测量结果的论文(特别是一对一校准和解卷积)也会涉及测量步骤。就方向性属性的测量而言，人们对采用旋转方向性天线的替代方案有所讨论[Pajusco 1998]。

在 Haardt and Nossek[1995]与 Roy et al.[1986]中描述了 ESPRIT 算法；Schmidt[1986]中描述了 MUSIC 算法；要进一步了解 MVM (Capon 波束成形器)，可参阅 Krim and Viberg[1996]；Fleury et al.[1999]中描述了 SAGE。SAGE 算法与梯度方法的巧妙结合在 Thomae et al.[2005]中有所描述。Richter[2006]中讨论了将扩散辐射(diffuse radiation)结合到信息提取中。其他高分辨率算法包括：非常适合于分析超宽带信号的 CLEAN 算法[Cramer et al. 2002]、JADE 算法[van der Veen et al. 1997]和其他许多算法。Salmi et al.[2009]讨论了通过一系列的测量来达成对多径分量的跟踪。

第9章 天 线

9.1 引言

9.1.1 天线的系统集成

天线是无线传播信道与发射机和接收机之间的接口，从而对无线系统的性能有着重要的影响。因此，本章将讨论无线系统天线(包括基站天线和移动台天线)的某些重要特性。讨论的重点主要是专门用于实际无线系统的天线。

无线通信的天线设计受到两个因素的影响：(i)性能因素；(ii)尺寸和造价因素。后者对于移动台天线显得尤为重要。移动台天线不仅必须体现出优良的电磁性能，而且必须尺寸小、坚固耐用且易于制造。此外，移动台天线的性能受到起固定作用的机壳及操作手机的人的影响。而另一方面，基站天线更类似于“传统的”天线。在尺寸和造价上都没有过于严格的约束，并且(至少就室外应用而言)天线附近的环境无障碍物(这种规则的某些例外情况也将在本章讨论到)。

因此，本章其余部分会区分移动台天线和基站天线。我们将讨论不同类型的天线，同时也将讨论环境对天线性能的影响。

9.1.2 天线参量特征

方向因子

天线的方向因子 D 是发射天线在特定方向的集中辐射程度(或者是接收天线从特定方向集中接受的辐射程度)的度量值。更确切地，将其定义为[Vaughan and Andersen 2003]：

$$D(\Omega) = \frac{\text{在方向 } \Omega \text{ 上每单位立体角度的总辐射功率}}{\text{每单位立体角度的平均辐射功率}} \quad (9.1)$$

根据互易原理，发送和接收情况下的方向因子是相同的。它与远场天线功率方向图 $G(\Omega)$ 有关^①。值得一提的是，天线的功率方向图是归一化的，因此有

$$\frac{1}{4\pi} \int G(\Omega) d\Omega = 1 \quad (9.2)$$

在许多情形下，针对不同的极化情况，天线具有不同的方向图。就这些情形而言，

$$D(\phi_0, \theta_0) = \frac{G_\phi(\phi_0, \theta_0) + G_\theta(\phi_0, \theta_0)}{\frac{1}{4\pi} \int \int (G_\phi(\phi, \theta) + G_\theta(\phi, \theta)) \sin(\theta) d\theta d\phi} \quad (9.3)$$

其中， θ 和 ϕ 分别是垂直方位角(elevation，也称为极角)和水平方位角(azimuth，也称为方位角)，而 G_θ 和 G_ϕ 分别是辐射被分解到 θ 和 ϕ 所在平面之后各自的功率增益。

^① 注意， $G(\Omega)$ 指的是天线的功率值，当我们将 $\tilde{G}(\Omega)$ 定义为复幅度增益时，就有 $G(\Omega) = |\tilde{G}(\Omega)|^2$ 。这两个量都是就天线远场定义的。

特定方向上的天线增益与方向因子有关。然而,增益数值的确定还必须考虑到损耗,如电阻损耗。这些损耗用天线效率来描述,接下来讨论天线效率。

效率

几种不同的现象都可能导致天线损耗。首先,由于天线材料的有限传导特性会导致电阻损耗。其次,接收天线和入射的辐射场之间的极化态可能不匹配(参见以下关于“极化”的内容)。最后,电路的不完全匹配也会造成损耗。因此,效率可以表达为下式:

$$\eta = \frac{R_{\text{rad}}}{R_{\text{rad}} + R_{\text{ohmic}} + R_{\text{match}}} \quad (9.4)$$

其中, R_{rad} 是辐射电阻,它的定义是:与天线等效的网络元件中的电阻值。通过这种等效可以将辐射功率 P_{rad} 写成 $0.5 |I_0|^2 R_{\text{rad}}$ 的形式,其中 I_0 是激励电流的幅度值。例如,半波偶极子的输入阻抗为 $Z_0 = 73 + j42 \Omega$, 所以其辐射电阻为 $R_{\text{rad}} = 73 \Omega$ 。

就仅仅工作在单一频率上的天线而言,完全匹配是可以做到的。然而,在更宽的频段上,要实现很好的天线匹配则存在着诸多限制。对于天线阻抗可以建模为一个电阻和一个电容串联的情形,所谓的 Fano 界表示为

$$\int \frac{1}{(2\pi f_c)^2} \ln \left(\frac{1}{|\rho(f)|} \right) df < \pi RC \quad (9.5)$$

其中, ρ 是反射系数,而 R 和 C 分别表示电阻值和电容值。

高效率对于任何无线天线而言都是关键性指标。从发射天线的角度,高效率意味着达到特定场强所需的放大器输出功率可以降低。从接收天线的角度,可达到的信噪比与天线效率完全成正比。因此,天线效率决定了移动台的电池使用时间及链路的传输质量。然而,很难就天线效率定义一个绝对的指标。最理想地,在整个感兴趣的带宽上应该都能取得 $\eta = 1$ 的效率值。然而,必须指出的是,近年来移动台天线的天线效率有所下降。这主要是出于装饰性的外在原因(也就是要减小天线尺寸)而牺牲了效率。

影响辐射效率的另一个因素是在移动台附近有绝缘及导电材质(即用户)存在。在计算可用于链路预算的天线增益时,必须将这一材质的影响考虑在内,因为它会造成天线方向图的严重变形,并会吸收部分能量。因此,当考虑到这些效应时,辐射效率会有所下降。为消除任一用户在个性特征上的差异(手持手机时,相对于用户头部呈现什么角度?用户惯用右手还是左手操作?持机的手部的绝缘特性怎么样?等等),必须针对大量用户分析有效增益。图 9.1 显示了一个贴片天线和一个螺旋天线的典型累积分布函数。注意,有效天线增益和理论值之间的差值可能超过 10 dB。这一点对网络规划具有很大的影响。

Q 因数

天线的基本数量参数是其 Q 因数,它的定义为[Vaughan and Andersen 2003]:

$$Q = 2\pi \frac{\text{储备的能量}}{\text{每周周期所消耗的能量}} \quad (9.6)$$

Q 因数可以与输入阻抗联系起来:

$$Q = \frac{f_c}{2R} \frac{\partial X}{\partial f} \quad (9.7)$$

其中,输入阻抗为 $Z = R + jX$ 。对于包含在直径为 L_a 的球面内的天线而言, Q 因数由下式给出:

$$Q \geq \frac{1}{(k_0 L_a/2)^3} + \frac{1}{k_0 L_a/2}$$

(9.8)

此处 k_0 仍然等于 $2\pi/\lambda_0$ 。

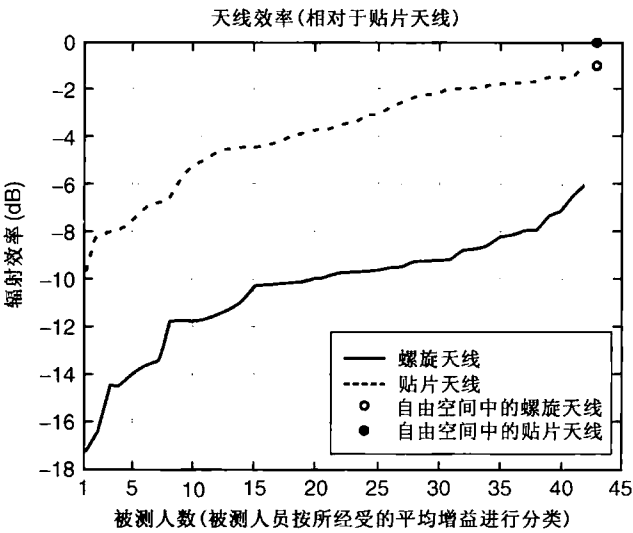


图 9.1 靠近人体的移动台天线的效率(相对于自由空间中天线的效率)。在测试样本集中, 采用了不同用户和移动台位置的样本。引自Pedersen et al. [1998]© IEEE

平均有效增益

天线的方向图及其所表示的天线的方向性都是针对远场区(即假定在这样的区域内存在均匀平面波)定义的, 同时它们还是所讨论的方向的函数。方向图是在吸波室(anechoic chamber)中进行测量的, 这种测试环境中去除了天线周边的所有障碍物及可能造成方向图变形的物体。这种方式下测得的方向图和增益与任何一本天线理论书给出的数字相当接近, 例如一个赫兹偶极子的增益为 1.5。当天线工作于一个随机散射环境时, 考察其平均有效增益(MEG)就显得很有意义了。平均有效增益是在天线入射波的众多方向由一个随机环境所确定的情况下, 不同方向增益的平均值[Andersen and Hansen 1977, Taga 1990]。平均有效增益定义为: 移动台天线接收到的平均功率与各向同性天线接收到的垂直和水平极化波平均功率的总和之比。

极化

在某些无线系统(如卫星电视)中, 到达波的极化和接收天线的极化必须非常小心地匹配。这种匹配程度可以通过庞加莱(Poincaré)球来度量。每一个极化态都与球面上的一个点有关: 右旋圆极化和左旋圆极化分别对应于球的北极和南极, 而不同的线性极化态则对应于球的赤道; 其他所有的点都对应于椭圆极化。球面上两点之间的角度是它们的不匹配程度的度量。

非视距应用场合一般对天线的特定极化没有非常严格的要求。即使在发射天线以单一极化波发射为主的情况下, 无线信道中的传播也会引起去极化现象(见第 7 章), 以至于接收机处的交叉极化很少高于 10 dB。在许多实际应用环境下, 这一事实的存在可以带来好处: 由于移动台天线的方位是不可预测的(不同用户的持机方式不同), 天线对入射波极化的低敏感度及入射波的均匀极化往往是有益的。

要求天线具有良好的交叉极化分离能力的一种情况是采用极化分集(见第13章)的场合,此时天线的绝对极化形式并非关键所在,而两个天线单元的交叉极化分离能力却显得至关重要。

带宽

天线带宽定义为天线特性(反射系数、天线增益等)满足其规格要求情况下的带宽。绝大多数无线系统都具有大约10%的相对带宽。这个数字业已成为绝大多数系统采用频分双工(FDD,见第17章),即收发采用不同频率的原因。例如,在GSM 1800^①系统(见第21章)中,带宽约为200 MHz,而载频位于1800 MHz附近。由于人们希望针对收发两种情形只使用一个天线,天线带宽就必须宽到足以把收发频段都包括在内。如同在前面所指出的,大的带宽意味着匹配电路不再能做到完全匹配。天线的物理尺寸越小,这种作用就越明显。

另一引人注意的具体案例与双模或多模设备有关。例如,大多数GSM手机必须能够工作在900 MHz和1800 MHz的载频上。这种设备的天线就必须按在整个0.9~2 GHz频段都能达到良好性能的要求进行设计。然而,采用这种具有50%相对带宽的天线显得有些“矫枉过正”了,因为1~1.7 GHz的频率范围根本没有必要覆盖。因此,仅针对所需频段设计具有良好性能的天线要更好些;如果不同载频彼此呈整数倍关系,这一目标将更容易达到。尽管如此,设计可以覆盖两个或更多个频段的高效率天线仍然是相当具有挑战性的工作。就基站天线而言,从技术角度出发,通常更愿意对不同的频段采用不同的天线。然而,出于美观的原因和降低视觉影响的目的,有时也愿意采用多波段天线。

9.2 移动台天线

9.2.1 单极子和偶极子天线

线天线是用于移动台的“经典”天线,长久以来决定着这些移动设备的典型“样貌”。最常见的一种是安置于导电表面(机壳)上的电单极子,以及偶极子。沿 z 轴布置的短(赫兹)偶极子的天线方向图对水平方位角具有均匀性,而对极角 θ (该角度从 z 轴量起)则呈正弦曲线状:

$$\tilde{G}(\varphi, \theta) \propto \sin(\theta) \quad (9.9)$$

并具有如下最大增益:

$$G_{\max} = 1.5 \quad (9.10)$$

$\lambda/2$ 偶极子具有以下性质(见图9.2):

$$\tilde{G}(\varphi, \theta) \propto \frac{\cos\left(\frac{\pi}{2} \cos(\theta)\right)}{\sin(\theta)} \quad (9.11)$$

和如下的最大增益:

$$G_{\max} = 1.64 \quad (9.12)$$

于是,可以看出 $\lambda/2$ 偶极子和赫兹偶极子的方向图差别并不显著。然而,辐射电阻却可能差得很远。对于电流均匀分布的偶极子,辐射电阻等于

^① 工作于1800 MHz频段的全球移动通信系统(GSM)。

$$R_{\text{rad}}^{\text{uniform}} = 80\pi^2 (L_a/\lambda)^2$$

(9.13)

对于呈锥形电流分布(馈电端电流最大,朝向天线末端电流呈线性递减)的偶极子,辐射电阻为 $0.25 R_{\text{rad}}^{\text{uniform}}$ 。

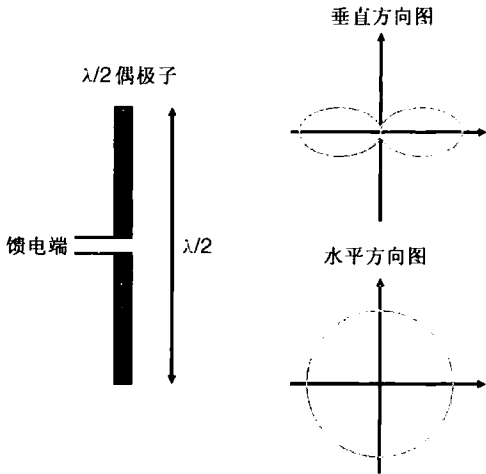


图 9.2 λ/2 偶极子天线的外形和辐射方向图

由镜像原理可以得出以下结论:位于一个导电平面上方的单极子与位于上半平面的偶极子天线有完全一样的辐射方向图。由于单极子发射到上半空间($0 \leq \theta \leq \pi/2$)的能量二倍于偶极子,所以最大增益是偶极子的两倍,而辐射电阻是偶极子的一半。可是应当注意到,镜像原理仅在导电平面延伸至无群大时才有效,这并非移动台机壳的情形。因此,可以预料到即使是单极子天线也会有相当一部分辐射进入了下半空间。

通常我们不希望减小辐射电阻,因为这样会使匹配变得更困难,并导致由电阻损耗造成的效率下降。不增加天线的物理尺寸的前提下,提升效率的一种方式是采用折合偶极子。折合偶极子由一对在非馈电端连接起来的半波长导线构成[Vaughan and Andersen 2003],这样就加大了输入阻抗。

单极子和偶极子天线的最大优点就是便于制作和价格低廉。相对带宽足以支持单天线系统的绝大多数应用。缺点在于移动台机壳上必须连接一根相对较长的金属杆。在 900 MHz 频段,一个 $\lambda/4$ 单极子的长度为 8 cm,常常比移动台机身还要长。即使以可伸缩单元的形式来实现,也很容易受到损坏。基于上述原因,更短的或内置的低效率天线正在获得日益广泛的应用。在通常覆盖程度较好的欧洲和日本,这一点并不是个重要问题。但是,在美国及其他那些在国内许多地区呈现出不规则覆盖状况的国家,这一点会在性能方面带来相当大的影响。

9.2.2 螺旋天线

螺旋天线的外形略图如图 9.3 所示。螺旋天线可以看成环天线和线天线的混合体,因此也就具有两种工作模式。天线的尺寸决定了它工作于何种模式。如果螺旋天线的尺寸远小于一个波长,则天线工作在常规模式。这时其表现类似于线天线,并具有集中成形于辐射方向的波瓣。这就是应用于移动台天线的工作状况。一般而言,极化形式是椭圆极化,而当比值 $2\lambda d/(\pi D)^2 = 1$ 时,会变为圆极化[Balanis 2005]。如果将螺旋天线放置于导电平面上,由于实际天线的水平分量与其镜像互相抵消,极化形式将变成垂直极化。当螺旋天线的周长接近

于一个波长时,天线方向图将会沿螺旋轴向呈现最大辐射,而极化形式也接近于圆极化。

由于处于常规模式的螺旋天线类似于线天线,天线的螺旋数目不会对天线方向图产生影响。而带宽、效率和辐射电阻将随着 h 的增加而增加。一般而言,螺旋天线相对于单极子天线具有较窄的带宽和较小的输入阻抗;然而,通过采用适当的匹配电路,可以达到 10% 的相对带宽。螺旋天线的最大好处就是其更小的尺寸;这一点使得它(与线天线一起)成为应用最广泛的移动台外置天线。

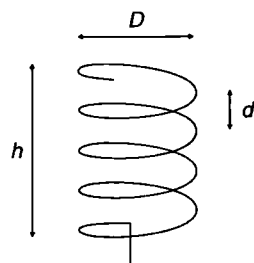


图 9.3 螺旋天线的外形

9.2.3 微带天线

微带天线(贴片天线)由一面覆盖着薄导电材料层(接地面),另一面附着着导电材料贴片的薄电介质底层构成。其结构略图如图 9.4 所示[Fujimoto 2008]。

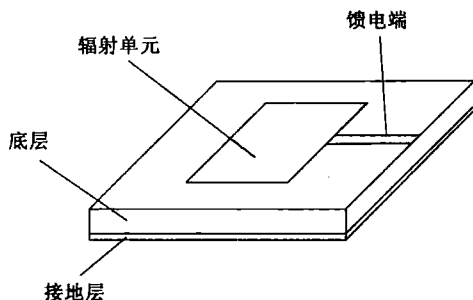


图 9.4 微带天线外形

微带天线的属性由金属贴片的形状和尺寸决定,也取决于所采用底层的介电特性。实质上,贴片就是一个尺寸必须为有效介质波长倍数的谐振器。因此,高介电常数的底层可以支持构造小天线。最常采用的贴片形状是矩形、圆形和三角形。

通常采用同轴电缆或者微带线向贴片馈电。也可能通过电磁耦合向贴片馈电。后一种情况所采用的是接地层夹在两层电介质材料之间的底层。一层材料之上有贴片,馈线则位于另一层电介质之下。耦合是通过接地层上的一个槽(缺口)来起作用的。因此,这些天线又称为缺口耦合贴片天线。这种设计的好处在于根据贴片和馈线的需要,可以选择两层介电特性不同的电介质材料作为底层。此外,这种设计具有比传统贴片天线更宽的带宽。

如同上面指出的,微带天线的尺寸和效率取决于电介质底层的参数。大的 ϵ_r 可以减小天线尺寸。由谐振贴片的单边长度必须服从下式可以立即得出这一结论:

$$L = 0.5\lambda_{\text{substrate}} \quad (9.14)$$

其中,

$$\lambda_{\text{substrate}} = \lambda_0 / \sqrt{\epsilon_r} \quad (9.15)$$

遗憾的是,物理尺寸的减小也导致了更窄的带宽,这常常又是我们所不希望的。基于这一原因,实践中用到的底层一般都具有很小的 ϵ_r ,很多底层甚至是用空气做的。进一步减小贴片天线尺寸的可能性在于使用短路谐振器,它将贴片谐振器的所需尺寸从 $\lambda/2$ 减小到了 $\lambda/4$ (见图 9.5)。

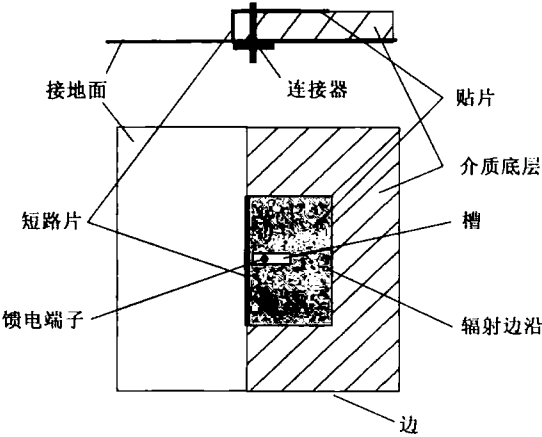


图 9.5 短路 $\lambda/4$ 贴片天线

采用各种不同措施都可以增加微带天线的带宽。最直接的一种就是增大天线的体积，即采用具有较小 ϵ_r 的厚底层。另外还可以采用匹配电路和应用寄生单元。

微带天线在无线应用上有如下几个重要好处：

- 尺寸小并能以低廉的成本制造；
- 馈线可以与天线一样制作在同一底层上；
- 可以将它们集成到移动台内部，而不需要伸出机壳。

可是，微带天线也有如下一些严重缺陷：

- 具有较低的带宽(常常只是载波频率的百分之几)；
- 效率低。

9.2.4 平面倒 F 形天线

微带天线的某些问题可以通过采用平面倒 F 形天线 (PIFA) 来得到缓解。平面倒 F 形天线的形状与 $\lambda/4$ 短路微带天线相似(见图 9.6)。一个平面状的辐射单元平行于接地面固定。此单元在距离为 W 的一段上短路到地。如果所选择的 W 等于边沿长度 L ，则可以得到 $\lambda/4$ 短路微带天线。如果所选定的 W 更小一些，则谐振长度会有所增加，并且辐射单元上的电流分布也会有所变化。

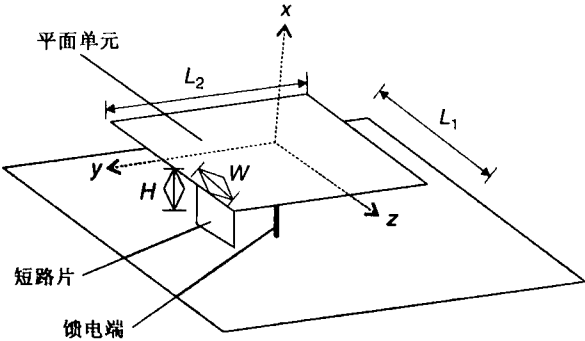


图 9.6 平面倒 F 形天线

9.2.5 辐射耦合双 L 形天线

进一步的改进通过所谓辐射耦合双 L 形天线 (RCDLA) 来实现 (见图 9.7 及 Rasinger et al. [1990])。它由两个 L 型直角构件组成, 其中只有一个被直接馈电 (传导式馈电)。另一个 L 形构件则由第一个 L 形构件以辐射耦合的形式馈电。这就增加了整个装置的带宽。通过将这种天线最佳地置于机壳上以后, 可以达到至多 10% 的相对带宽, 约为平面倒 F 形天线带宽的两倍。

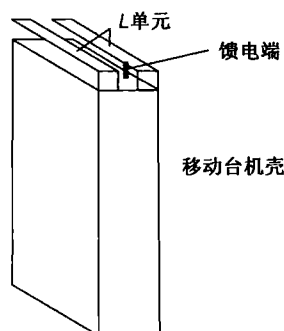


图 9.7 辐射耦合双 L 形天线

9.2.6 多波段天线

人们期待时髦的蜂窝手机可以支持不同的通信频率。如同 9.2.2 节讨论过的, 例如, GSM 手机在绝大多数国家必须至少可以支持标准所预先规定的 900 MHz 和 1800 MHz 通信。许多手机还必须支持美国使用的 1900 MHz 模式, 以及在需要蓝牙连接 (如手机到无线耳机的连接) 的情况下支持 2.4 GHz 频段, 这就进一步加大了困难。支持 GSM 和宽带码分多址 (WCDMA, 见第 26 章) 的双模手机使情况变得更加复杂。设计内置的多波段天线是十分复杂的, 几乎不存在闭式设计准则。图 9.8 显示了一个多波段微带天线的例子。

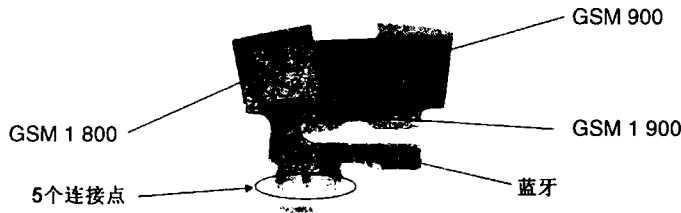


图 9.8 集成多波段天线。引自 Ying and Anderson [2003] © Z. Ying

9.2.7 移动台天线的安置

天线并非在空荡荡的空间中工作, 而是要被安装在可看成辐射器一部分的机壳上。此外, 天线的特性参数也会受到用户手部和头部的影响; 天线在移动台上的装配情况也会对天线参数产生影响。因此, 勘察可供选择的安置天线的不同位置, 并确定相应的安置方式可能对性能产生的影响, 就显得很重要了。

线天线和螺旋天线通常置于机壳顶部的窄面之上, 即它们从机壳的这一部分伸出去。这主要是出于人机工程学的考虑, 如果它们从底面伸出去, 用户将会感到不适, 而且用户手部常常会盖住天线, 导致额外的衰减。

就微带天线、平面倒 F 形天线 (PIFA) 及辐射耦合双 L 形天线 (RCDLA) 而言, 可以有更多可供选择的安置位置。这些天线常常用来作为内置天线, 即将其与机壳合并, 或者封闭于机壳内。这样就大大降低了机械损伤的危险性。然而, 用户手部置于天线上的可能性却有所增长, 从而也就增加了对电磁能量的吸收, 并因此使链路性能变差了 [Erätuuli and Bonek 1997]。辐

射耦合双 L 形天线安装于机壳不同位置的示例在图 9.9 中可以见到;其他类型的微带天线也可以采用类似的安置方式。

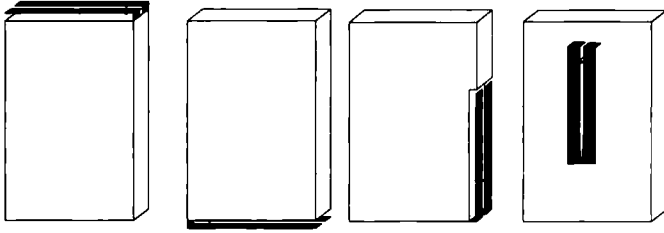


图 9.9 辐射耦合双 L 形天线在移动台机壳上的不同安装方式

9.3 基站天线

9.3.1 天线类型

基站天线和移动台天线的设计要求是不同的。由于基站的造价已经相当昂贵,天线的成本就成为了相对影响较小的因素。同时,对天线尺寸的约束也没有那么严苛,就宏蜂窝而言,仅仅要求:(i)天线支架的机械强度必须足够,尤其在受到风力作用时;(ii)对周围环境“美观”的影响必须足够小。就微蜂窝和微微蜂窝而言,由于天线被安装在建筑物表面、街灯或办公室墙上,天线应当尽量地更小一些。基站天线与移动台天线相比,所要求的天线方向图十分不同。当基站天线的方位和布局确定以后,就应该按照不损失任何能量的方式来对方向图进行成形(见 9.3.3 节)。这种方向图成形可借助多单元天线最便利地加以实现。

基于上述原因,宏蜂窝天线一般为天线阵列或由线天线单元构成的八木天线。而对于微蜂窝和微微蜂窝,贴片天线组成的天线阵列相当普遍。除了对尺寸的要求(与移动台天线相比,这方面的要求已放宽了许多)以外,9.2 节关于带宽、效率等的全部论断仍然有效。

9.3.2 阵列天线

阵列天线通常用于基站天线。采用阵列天线可以更容易地实现天线方向图的成形。这种成形能按预先确定的方式实现(见 9.3.3 节),也可以以自适应方式实现(见第 13 章和第 20 章)。天线阵列的方向图可以表示成单个单元方向图和阵列因子 $M(\phi, \theta)$ 的乘积。在阵列天线平面上,均匀线阵列的阵列因子为(可与 8.5.1 节比较)[Stutzman and Thiele 1997]

$$M(\phi) = \sum_{n=0}^{N_t-1} w_n \exp[-j \cos(\phi) 2\pi d_n n / \lambda] \quad (9.16)$$

其中, d_n 是天线单元之间的距离^①,而 w_n 是单元激励的复权值。对于 $|w_n| = 1$ 对所有 n 值均成立,而 $\arg(w_n) = n\Delta$ 的情形,阵列因子变成了

① 通常选定 $d_n = \lambda/2$ ($d_n \leq \lambda/2$ 是必须的,以避免天线方向图的周期性带来的空间混叠) 后面假定线阵列的各个单元都位于 x 轴。显然,当天线单元竖直地叠放起来时,同一原理仍然有效,并应当对阵列的垂直方向图进行成形。

$$|M(\phi)| = \left| \frac{\sin \left[\frac{N_r}{2} \left(\frac{2\pi}{\lambda} d_a \cos \phi - \Delta \right) \right]}{\sin \left[\frac{1}{2} \left(\frac{2\pi}{\lambda} d_a \cos \phi - \Delta \right) \right]} \right| \quad (9.17)$$

馈电电流的相移 Δ 决定了天线主瓣的方向。在相控阵天线理论中,这一原理是众所周知的([Hansen 1998], 见第8章)。通过迫使 $\varphi_n = n\Delta$, 自由度已减小为1; 与此同时主瓣就可以指向任意方向, 而旁瓣和零点的分布将唯一地取决于已指定的主瓣方向。

例9.1 假定基站天线安装在51 m的高度, 向半径为1 km的小区提供覆盖。天线由8个垂直叠放的短偶极子组成, 相邻偶极子之间的距离为 $\lambda/2$ 。要使最大点(干涉相长点)位于小区边缘1 m高的移动台处, 相移 Δ 应该等于多少?

解:

为了在方向 θ_0 处获得由来自不同天线的众多分量形成的相长干涉, 角度 Δ 必须满足:

$$\Delta = \frac{2\pi}{\lambda} d_a \cos \theta_0 \quad (9.18)$$

很明显, 当 $\theta_0 = \pi/2$ 时 $\Delta = 0$, 即不必有相移。由此例题设, 现在可以得到一个倾角:

$$\theta_0 = \frac{\pi}{2} + \arctan \left(\frac{51 - 1}{1000} \right) = \frac{\pi}{2} + 0.05 \quad (9.19)$$

因此, 相移必须满足:

$$\Delta = \frac{2\pi}{\lambda} d_a \cos \left(\frac{\pi}{2} + 0.05 \right) = -0.05\pi \quad (9.20)$$

单元方向图的影响可以忽略。

9.3.3 改进天线方向图

基站天线呈各向同性的辐射是我们所不希望的。辐射被投射到一个具有大仰角的方向上(即投射到天空中)不仅带来能量的浪费, 而且实际上增加了对其他系统的干扰。若能做到在整个小区内接收功率处处为常数, 而小区外完全没有功率, 就得到了最佳的天线方向图。下面的问题就是如何合成这样的垂直方向图。通过阵列天线可以近似做到这一点, 其复权值将按照使方向图的均方误差最小化的方式来选取。一种更简单的措施是采用 $w_n = 1$ 的天线阵列, 而将主瓣下倾 5° 左右。这种下倾既可以通过机械方式(通过下倾整个天线阵列)完成, 也可以通过电子方式实现。

所需的水平天线方向图或者是全向的, 即在 $[0, 2\pi]$ 上处处一致, 或者在一个扇区上保持一致, 而在扇区外则为零。扇区的角度通常是 60° 或 120° 。农村地区的宏小区多采用全向天线, 由线天线来实现全向覆盖。线性阵列中天线的排布仅沿竖直轴展开。扇区天线可以采用带有适当的成形反射器的线天线, 或者可以采用微带天线(具备固有的非一致天线方向图)。

9.3.4 环境对天线方向图的影响

基站天线方向图的定义往往针对天线位于自由空间的情形, 或者天线位于理想导电平面上的情形。这正是在吸波室中测到的东西, 也是最容易计算的。然而, 在其实际工作位置上,

基站天线被有限大小和有限导电性的不同物体包围着。如此环境下的天线方向图与理论方向图之间存在明显的偏差。天线支架通常是金属的,并因此具有较强的导电能力,会造成方向图变形。类似地,用某些特定建筑材料(例如钢筋混凝土)建造的屋顶也会造成方向图的变形(见图 9.10)。

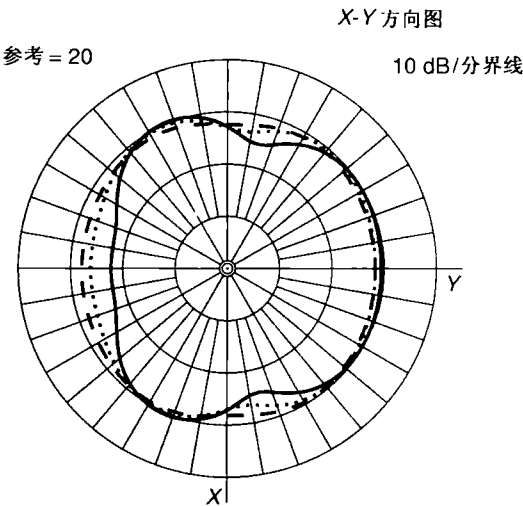


图 9.10 靠近天线支架的全向天线方向图。与天线支架之间的距离为 30 cm。支架的直径分别为:非常小(实线)、5 cm(虚线)和 10 cm(点状线)。引自 Molisch et al. [1995] © EMA (欧洲微波协会)

垂直方向图的变形绘制于图 9.11 中。这种变形源于两方面的作用:(有限大小)的屋顶比地面更接近于天线,以及屋顶的介电属性与地面的不同。

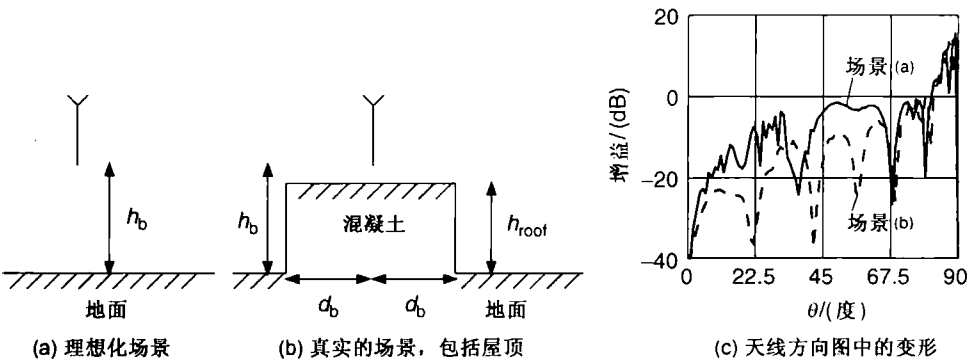


图 9.11 垂直方向图的变形。所有材料的属性:相对介电常数 $\epsilon_r = 2$; 传导率 0.01 S/m ; $d_b = 20 \text{ m}$ 。引自 Molisch et al. [1995] © EMA

人体的存在也会造成天线方向图的失真。对这一状况加以考虑的一种方式就是将天线和人体视为一个“超级天线”(superantenna),可以用与“常规”天线完全相同的方式来对它的特征量(效率、辐射方向图)进行测量和表征。图 9.12 给出了存在人的头部和身体时天线方向图测量结果的示例。

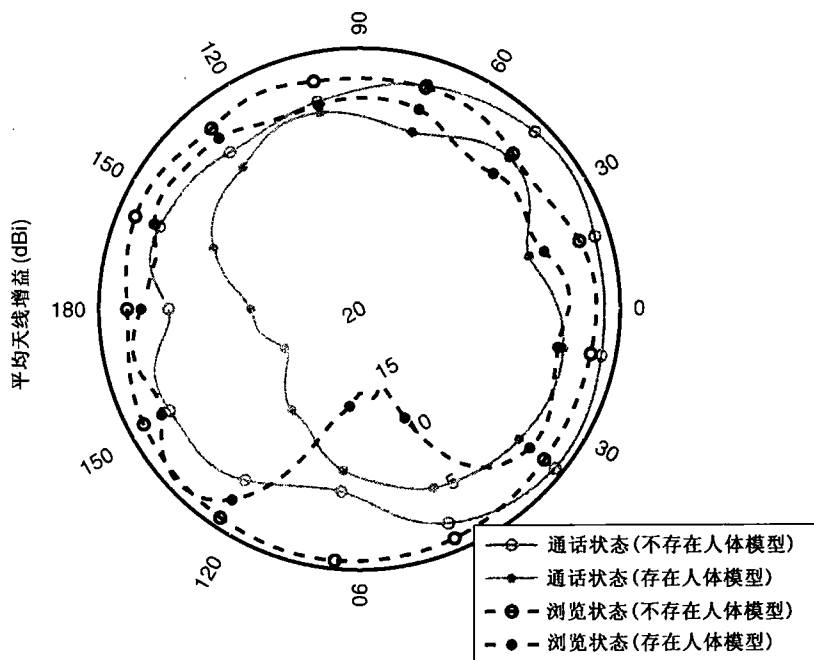


图 9.12 在通话和数据浏览状态下, 人体造成的天线方向图失真。引自 Harryson et al. [2010] © IEEE

深入阅读

就天线理论的一般性介绍而言, 上文已提到了许多优秀的天线理论著作: Balanis[2005], Kraus and Marhefka[2002], Ramo et al. [1967] 及 Stutzman and Thiele[1997], 也包括 Collin [1985] 的略显更高级的教科书; 后者还专门辟出单独一章讨论接收天线的性质。无线通信专用天线的更多细节可以参考 Godara[2001] 与 Vaughan and Andersen[2003]。移动台天线在 Fujimoto[2008] 及 Hirasawa and Haneishi[1991] 中有所讨论。基站天线的设计在专著 Chen and Luk[2009] 和会议论文 Beckman and Lindmark[2007] 中进行了综述。最后, 在 Hansen[1998] 与 Mailloux[1994] 中详尽讨论了相控阵天线。有关波束下倾的讨论在 Manholm et al. [2003] 中可以找到。人的头部和身体对天线特征量的影响也被讨论过, 如 Ogawa and Matsuyoshi[2001], Kivekaes et al. [2004] 和 Harryson et al. [2010] 所述。

第三部分 收发信机和信号处理

第 10 章	无线通信链路结构
第 11 章	调制
第 12 章	解调
第 13 章	分集
第 14 章	信道编码
第 15 章	语音编码
第 16 章	均衡器

无线通信系统最终的性能极限取决于前面所讨论的无线传输信道。实际收发信机设计的主要任务包括寻找合适的调制方案、编码和信号处理算法，使其性能尽可能趋于极限值，同时要在性能和软硬件的实现之间寻求平衡。随着技术的进步，我们能够实现的方案越来越复杂。例如，第三代蜂窝电话的计算能力与(2000 年)个人计算机相当，其所完成的信号处理算法在 20 世纪 90 年代中期是不可想象的。因此，这部分将不过多关注算法的复杂度，因为目前看起来非常复杂的算法，有可能是若干年后的标准解决方案。

本部分首先在第 10 章介绍收发信机的一般结构，以及发射机和接收机框图中的各部分，并给出了可用于系统仿真或设计的简化模型。第 11 章讨论了不同的调制方式及其应用于无线系统时的优缺点。例如，恒包络调制对于电池供电的发射机而言非常有用，因为此时发射机可以使用高效率的放大器。第 12 章通过对这些调制方式进行数学分析，阐述了它们在不同类型的衰落信道中的误比特率性能。可以发现，某种调制方式的性能主要受到两个因素的影响，即衰落和时延色散。通过分集可明显减弱衰落的影响，所谓的分集就是通过不同的路径传输同一信号。第 13 章介绍了获得不同路径的多种方法，如采用多个天线，在不同的频率或时间重复发送信号等。该章同时讨论分集对不同调制方案性能的影响。时延色散的不利影响通过分集也可得到一定的改善，当然更有效的方法是采用均衡。均衡器不但可以克服由时延信号产生的码间干扰，而且可以利用这些时延信号中所包含的能量，因而均衡技术可使系统的性能得到相当大的改善，尤其是高数据速率的系统和/或信道具有较大时延扩展的系统。第 16 章介绍了各种均衡器的结构，从简单的线性均衡器到最佳的(但高复杂度)最大似然序列检测器。

分集和均衡并不总能充分或有效地降低错误概率。在很多情况下，编码能显著地增强系统性能，从而使之满足某一具体应用所需的传输质量要求。第 14 章概述了无线通信中最常用的各种编码方案，包括从 20 世纪 90 年代初期就开始受到广泛关注的准最佳 Turbo 码和低密度

奇偶校验(LDPC)码。这些编码方案用来纠正由传输信道引入的误码。同时本章也介绍信息论基本知识,这些是建立理想码所能达到最终性能极限的基础。另一种不同类型的编码是信源编码,它把来自信源的消息转换成能在无线信道中进行最有效传输的比特流。第15章概述了语音编码,它是无线应用中最重要信源编码类型。最后,第16章介绍了均衡,即信道时延色散的补偿方法。

一般来讲,无线通信中的调制、编码和均衡当然与数字通信有很大的关系,但是因为本书的这部分并不是侧重数字通信的教科书,所以假设读者或者通过以前的课程,或者通过某一本经典教科书,例如 Barry et. al[2003], Proakis[1995], Sklar[2001]和 Anderson[2005],已经对上述知识比较熟悉。本书将简明扼要地给出一些最关键的内容,仅为帮助读者回顾和复习。

第 10 章 无线通信链路结构

10.1 收发信机结构

这一节将给出无线通信链路的框图，并对各个模块进行简要介绍。更具体的内容将在后面第三部分和第四部分的章节中予以介绍。下面首先概述各个模块的功能，随后介绍侧重于不同硬件元件的框图。

图 10.1 给出了一个通信链路功能框图。在许多情况下，无线链路的作用就是将来自模拟信源(麦克风、录像机)的信息经过一个模拟无线传输信道传到模拟信宿(扩音器、电视屏幕)，信息的数字化仅仅是为了增加链路的可靠度。第 15 章将介绍语音编码，它代表了数字化模拟信息的一种最常见形式。第 23 章介绍了视频编码。对于其他传输(如文件传输)，它的信息已经是数字化的。

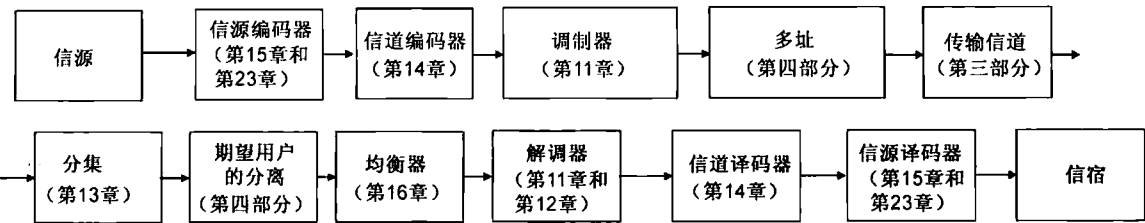


图 10.1 发射机和接收机框图，标明各个模块对应于哪一章

为了更易于抵抗由信道引发的错误，发射机以前向纠错码的方式来增加冗余(注意，虽然绝大部分无线系统都进行这种编码，但并不是所有的)。随后将这种已编码的数据作为调制器的输入，调制器将其映射成适于传输的输出波形。通过在特定频率或特定时间发送这些符号，可以区分不同的用户^①。然后，信号被发送到传输信道，会产生衰减和失真，并引入噪声，正如第二部分所介绍的。

在接收机中，可采用一副或多副天线(关于如何合并来自多个天线的信号的讨论见第 13 章)来接收信号，不同的用户被分离(例如，仅在某个特定频率接收信号)。如果信道是时延色散信道，就可以采用均衡器来抵消这种色散，消除码间干扰。接下来，对信号进行解调，并用信道译码器消除比特流中的(大部分)错误。最后，信源译码器将该比特流映射成模拟信息流，并送入信宿(扩音器、电视屏幕，等等)；当然，如果信息一开始就是数字的，最后一步可以省略。

以上关于各模块的描述当然是过于简化了，并且模块的划分不必太确切，尤其是在接收机中。最佳接收机将使用采样后的接收信号作为其输入，并且从中计算出以最大似然发送的信号。然而对于很多应用来说，最大似然的计算量仍然太大，所以在文献中已经研究了联合译码和解调及其他类似方案。

^① 不同的多址方法在第 18 章至第 22 章中介绍。

图 10.2 和图 10.3 给出了一个更具体的数字发射机和接收机框图,它着重强调了硬件部分和模拟与数字器件的接口。

- 信源提供模拟源信号,并且送入信源模数转换器(ADC)。模数转换器首先对来自模拟信源的信号进行带限处理(如果有必要),然后以某一采样速率和分辨率(每个样值的比特数)把信号转变为数字数据流。例如,语音信号典型的采样速率为 8 k 个样值/秒,分辨率为 8 比特,可以得到 64 kbps 的数据流。对于数字数据的传输,这些步骤可以省略,数字信源直接提供图 10.2 中接口“G”的输入。

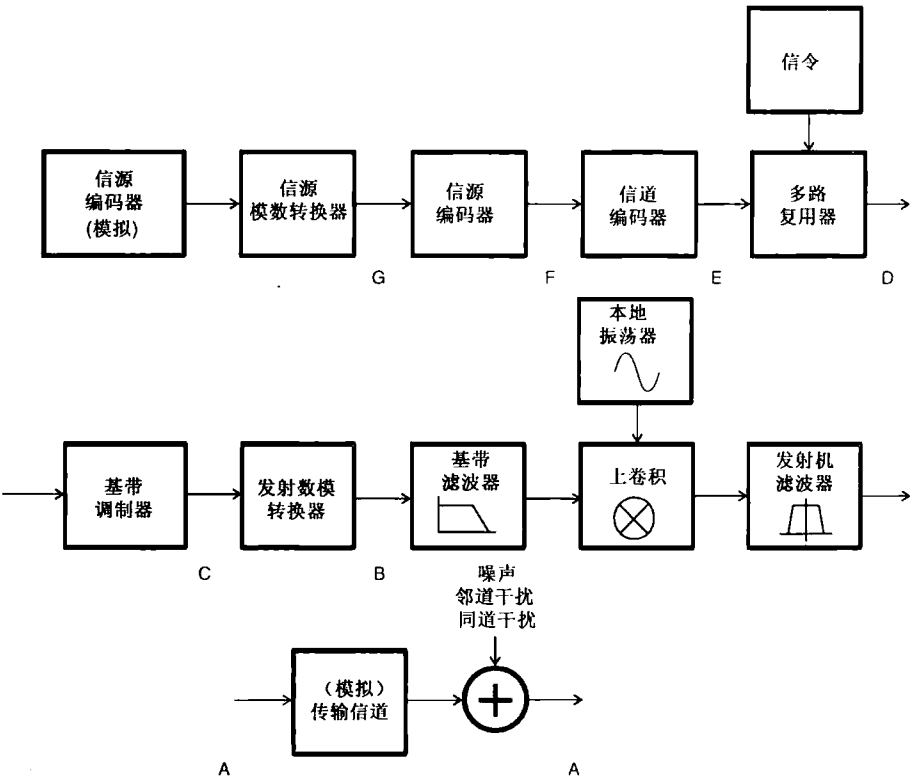


图 10.2 采用数字发射机、数字接收机和模拟传输信道的无线链路的框图

- 信源编码器利用信源数据性质的先验信息,以减小信源信号的冗余度,从而减少了需要传输的信源数据量,也减小了所需的传输时间和/或带宽。例如,全球移动通信系统的语音编码把信源数据率从上面提到的 64 kbps 减小到 13 kbps。对于音乐和视频(MPEG 标准)来说,类似的减小也是可能的。传真信息同样能够被显著地压缩。1000 个连续的字符“00”(代表“白”颜色),必须要用 2000 比特来表示,它可以用以下描述来取代:“现在紧接着的是 1000 个 00 字符”,而它仅需要 12 比特。对于一个典型的传真,可以获得 10 倍的压缩效果。信源编码器增加了接口 F 处数据的信息熵(每比特的信息量);因此比特错误会产生较大的影响。对于某些应用,信源数据需要进行加密,以防止未授权的偷听。
- 信道编码器通过增加冗余度来保护数据,减小传输错误的影响。这增大了接口 E 处发送数据的速率。例如,全球移动通信系统信道编码把数据率从 13 kbps 增加到 22.8 kbps。信道编码器常常利用信道中的错误源(噪声功率,干扰的统计特性)的统计信息来设

计信道码, 该信道码特别适用于特定的信道类型(如, Reed-Solomon 码主要用来抵抗突发错误)。数据可以根据其重要性进行排序, 比较重要的比特可获得更强的保护, 进而可以用交织拆散突发错误。值得注意的是, 如果交织和信道编码组合使用, 将非常有效。

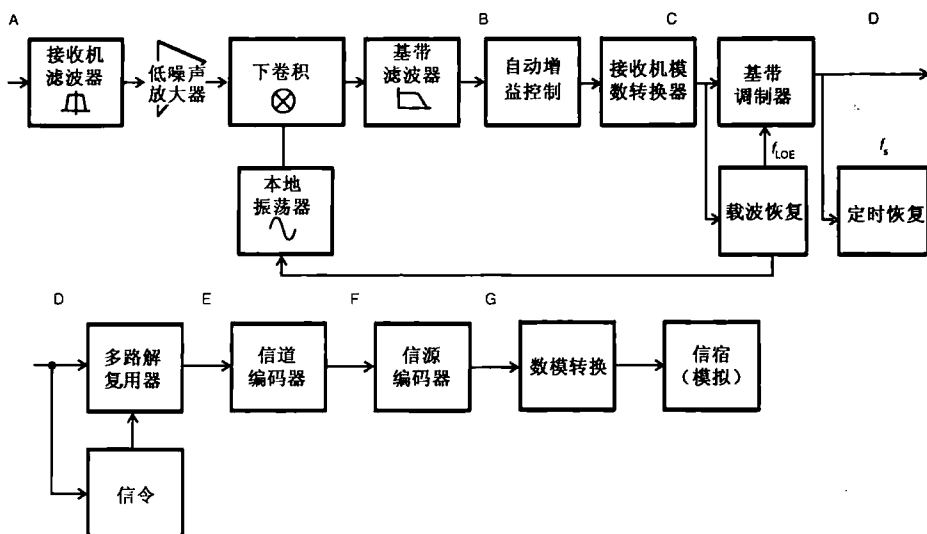


图 10.3 用于移动通信的数字接收机框图

- 信令是为建立和终止连接、与正确用户联系信息、同步等施加的控制信息。信令信息通常采用纠错码进行严格保护。
- 多路复用器合并用户数据和信令信息, 并合并来自多个用户的数据^①。如果采用时分复用, 则需要进行一些时间压缩。在全球移动通信系统中, 对于 8 个参与者的标准情况, 多址复用把数据率从 22.8 kbps 增加到了 182.4 kbps (8×22.8)。信令信息的加入使数据率增加到 271 kbps。
- 基带调制器将总数据比特(接口 D 处的用户数据和信令)分配到基带的复发送符号。这一步确定了频谱性质、符号间干扰、峰平比和发送信号的其他性质。基带调制器的输出(接口 C)以过采样形式提供了时间和振幅都离散的发送符号。
过采样和量化决定混叠和量化噪声。因此, 希望系统具有高的分辨率, 并且基带调制器输出端的数据率应远高于输入端的数据率。对于一个全球移动通信系统, 过采样参数为 16, 8 比特振幅分辨率的数据率约为 70 Mbps。
- 发射机的数模转换器产生一对模拟的离散振幅的电压, 分别对应于发射字符的实部和虚部。
- 发射机中的模拟低通滤波器滤除(不可避免的)所需传输带宽之外的频率分量。这些成分由(理想)基带调制器的带外辐射产生, 它由所选调制形式的性质决定。此外, 基带调制器和数模转换器的非理想性产生了附加的杂散辐射, 它必须通过发射机滤波器来抑制。

^① 实际上, 只有基站的多路复用器真正将不同用户的数据进行合并, 以进行发射。在移动台, 多路复用器仅仅确保基站的接收机能区分来自不同用户的数据流。

- 发射机本地振荡器提供一个未调制的正弦信号,对应于所考虑系统可接受的一个中心频率。对于频率稳定性、相位噪声和各频率之间切换速度的要求,取决于调制和多址方法。
- 上变频器通过与本地振荡器的信号混频,把一个模拟的滤波后的基带信号转换成一个带通信号。上变频既可以一步完成,也可以分几步完成。最后,需要在射频进行放大。
- 射频发送机滤波器滤除射频的带外辐射。即使低通滤波器成功地滤除了所有的带外发射,上变频仍然会导致附加的带外成分的产生。特别是混频器和放大器的非线性引起了互调产物和频谱再生,即产生了额外的带外辐射。
- (模拟)传输信道使信号受到衰减,导致时延和频率色散。另外,环境还会引入噪声(加性高斯白噪声)和同道干扰。
- 接收机滤波器对接收带宽进行粗略地选择。滤波器的带宽对应于某一业务所分配的总带宽,因此它可以覆盖同一业务的多个通信信道。
- 低噪声放大器对信号进行放大,因此降低了由接收链路之后的器件产生的噪声对信噪比的影响。在后续下变频的步骤中,信号还将被进一步放大。自动增益控制减小了信号的动态波动范围。
- 接收机本地振荡器产生的正弦信号与发射机本地振荡器的可能输出信号相对应。本地振荡器的振荡频率可通过载波恢复算法(见下文)进行精确调谐,以保证发射机和接收机的本地振荡器具有相同的频率和相位。
- 接收机下变频器把接收信号(用一步或几步)变到基带。在基带,信号是以模拟复信号的形式存在的。
- 接收机低通滤波器为某一特定用户提供所需频带的选择(它区别于接收机通带滤波器,后者选择设备工作的频率范围)。该滤波器滤除了邻道干扰和噪声,对所需信号的影响应该尽可能小。
- 自动增益控制放大信号,使其电平达到使下面的模数转换器能更好地进行量化。
- 接收机模数转换器把模拟信号转换成时间和幅度都离散的数值。模数转换器所需的分辨率本质上是由后续信号处理的动态范围决定的。采样速率只要满足了采样定理就可以了。过采样提高了对模数转换器的要求,但是简化了后续的信号处理。
- 载波恢复决定接收信号中载波的频率和相位,并用来调整接收机的本地振荡器。
- 基带解调器从数字基带数据处获得软判决数据,并且送入译码器。基带解调器可以是最佳的相干解调器,也可以是一个简化的差分或非相干解调器。这一步也可以包含进一步的信号处理,比如均衡。
- 如果存在多个天线,那么接收机或者选择其中的一路信号进行进一步的处理,或者将所有天线上的信号都进行处理(滤波、放大、下变频)。在后一种情况下,基带信号或者在送入传统的基带解调器之前进行合并,或者直接送入联合解调器,这样可以利用来自不同天线信号的信息。
- 符号定时恢复利用解调后的数据来估计符号的持续时间,并且利用它来精确调整采样时间间隔。
- 译码器利用来自解调器的软估计来确定原始(数字)数据源。在最简单的未编码系统中,译码器只是一个硬判决(门限)器件。对于卷积码,采用最大似然序列估计器(MLSE,

例如维特比译码器)。最近,提出了一种联合解调和译码的迭代接收机。剩余误码或者进行数据包重传(自动重传请求,即 ARQ),或者被忽略。后一种方案通常用在语音通信中,因为重传带来的时延是不可接受的。

- 信令恢复是指在数据中识别出表示信令信息的那一部分,并由它来控制后面的解复用器。
- 解复用器将用户数据和信令信息分开,并且恢复由发射机复用器可能引入的时间压缩。值得注意的是,在传输过程中解复用器的位置可以提前。它的最佳位置取决于具体的复用和多址方式。
- 信源译码器根据信源编码规则来重构源信号。如果源数据是数字的,输出信号就送到信宿。否则,数据就送入数模转换器,它把发送的信息转换成模拟信号,再交给信宿。

10.2 简化模型

通常人们更喜欢使用链路的简化模型。图 10.4 给出了适用于调制方法分析的链路模型。发射机中,从信源到复用器输出之间的各部分都被归入数字数据源的“黑盒子”中。模拟无线信道、上变频器、下变频器、射频元件(滤波器,放大器)及所有的噪声和干扰信号都被归入等效时间离散低通信道中,可采用时变的冲激响应和加性干扰的统计特性来描述。评价调制方式质量的标准是从发端接口 D 到收端接口 D 之间的误比特率。

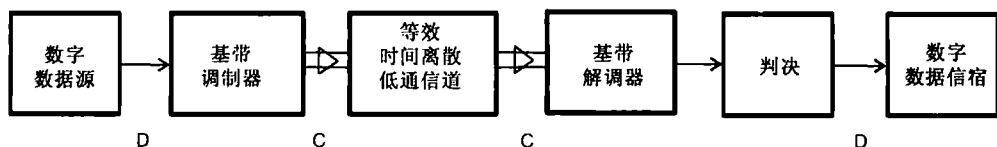


图 10.4 用于调制方法分析的数学链路模型

采用数字化信道(例如二进制对称信道)的其他简化模型主要适用于编码方案的分析。

深入阅读

本章简要地介绍了发射机和接收机的总体结构。在后续章节中将讨论数字信号处理的多个方面,关于这些内容,下面给读者推荐一些详细的参考文献。

数模和模数转换器是高速无线设备的功耗和制造成本的一个重要方面, van de Plassche [2003] 详细介绍了用 CMOS 技术实现的数模和模数转换器,该技术被大部分厂商所采用。射频硬件,包括放大器、混频器和合成器在 Pozar[2000], Razavi[1997] 和 Sayre[2001] 中有详细的介绍。在 Gonzalez[1984] 中讨论了放大器的设计。另一个重要的主题,即同步,在 Mengali and D'Andrea[1997], Meyr and Ascheid[1990] 及 Meyr et al. [1997] 中进行了讨论。

第11章 调 制

11.1 概述

数字调制是将数据比特映射为能够在(模拟)信道中传输的信号波形。如前一章所述,在无线信道中,传输的数据是数字的,这是因为需要传输的可能是数据文件,也可能是信源消息经过信源编码后的数字形式。另一方面,无线信道是模拟介质,在其中只能传输模拟信号波形。基于此,发射机中的数字调制器需要将数字信源数据转换为模拟信号波形。在接收端,解调器尽可能从接收到的波形中恢复出数据比特。本章将简要介绍数字调制方式,并重点介绍调制的结果。若读者想了解更详细的内容,可参考本章末的“深入阅读”一节。第12章将介绍最佳和次最佳解调器及其在无线信道中的性能。

根据调制类型,一个波形可以表示一个或一组比特。二进制调制是最简单的调制方式,其中+1用一个特定波形表示,而-1用另一个不同的特定波形表示。更一般的情况是,一组 K 个比特代表一个符号,每个符号分别与 $M=2^K$ 个波形之一相对应,这种情况称为 M 进制调制、高阶调制、多电平调制或者大小为 M 的符号集调制。无论怎样,不同调制方案的区别在于不同的传输波形,以及数据比特与信号波形的映射方式。通常与一个符号相关的波形限定在时间段 T_s 内,与不同符号对应的信号波形一个接一个地传输。显然数据速率(即比特速率)是符号速率(即信号速率)的 K 倍。

在无线通信系统中,选择某一种调制方案的最终目的是为了在特定的传输质量(即误比特率)和信道带宽限定下,以特定的能量传输尽可能多的信息。根据这一基本需求,在选择调制方案时应遵循以下准则。

- 调制方案的频谱效率应尽可能高。这可以通过采用高阶的调制方式来实现。这使得在传输中,每个符号承载多个数据比特。
- 相邻信道的干扰应尽可能小。这就要求信号频谱在带外滚降要快,而且信号在传输前必须进行滤波处理。
- 对噪声的敏感度要低。这可以通过使用低阶的调制方式获得,这样可以使各个信号波形之间的差异(假设平均功率相同)达到最大。
- 对时延和多普勒扩展的鲁棒性应尽可能大。因为滤波引起的时延扩展会使系统对信道引入的时延扩展更敏感,所以对传输信号应尽可能少地滤波。
- 波形应便于硬件实现,易于产生,并具有较高的功率效率。这些需求源于无线发送设备的实际要求。为了能够使用高效的C类(或E类和F类)放大器,通常采用具有恒包络特性的调制方式。另一方面,如果调制方式对信号包络畸变非常敏感,那么发送设备就必须使用线性放大器(A类或B类)。前者的功率效率可达80%,而后者却不到40%。这对电池的使用寿命影响很大。

上面的各点中有一些是相互冲突的,这就是无线通信中没有“理想”调制方案的原因。而且调制方案应该根据特定系统和应用的需求来选择。

11.2 基本概念

本章的其余部分采用一种等效基带表示方法来描述调制方式的等效基带形式。带通信号, 即实际物理信号, 可写成复基带表示形式^①:

$$s_{\text{BP}}(t) = \text{Re}\{s_{\text{LP}}(t) \exp[j2\pi f_c t]\} \quad (11.1)$$

11.2.1 脉冲幅度调制

许多调制方式都可以归为脉冲幅度调制(PAM), 这种调制方式以一定的调制系数 c_i 对基带脉冲 $q(t)$ 加权求和得到

$$s_{\text{LP}}(t) = \sum_{i=-\infty}^{\infty} c_i q(t - iT_s) \quad (11.2)$$

其中, T_s 为码元间隔。这就是说, 发射的模拟信号波形 s 包含一系列由(复)标量线性加权的时移基本脉冲, 这个加权标量与发射的符号相关。各个基本脉冲的调制相互独立。

不同脉冲幅度调制方式的区别在于其数据比特 b 和调制系数 c_i 的映射方式, 这将在后续章节中详细分析。现在转而研究各种可能的基带脉冲 $q(t)$ 及其对信号频谱的影响。假设基带脉冲均归一化为具有单位平均功率, 即

$$\int_{-\infty}^{\infty} |q(t)|^2 dt = \int_{-\infty}^{\infty} |G(f)|^2 df = T \quad (11.3)$$

式中第二个等式服从帕斯瓦尔关系。

矩形基带脉冲

最简单的基带脉冲即为矩形脉冲, 图 11.1 和图 11.2 分别给出了其时域波形和相应频谱:

$$\left. \begin{aligned} g_{\text{R}}(t, T) &= \begin{cases} 1, & 0 \leq t \leq T \\ 0, & \text{其他} \end{cases} \\ G_{\text{R}}(f, T) &= \mathcal{F}\{g_{\text{R}}(t, T)\} = T \text{sinc}(\pi f T) \exp(-j\pi f T) \end{aligned} \right\} \quad (11.4)$$

其中 $\text{sinc}(x) = \sin(x)/x$ 。

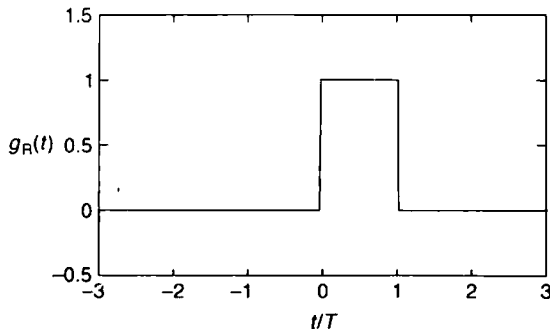


图 11.1 矩形基带脉冲

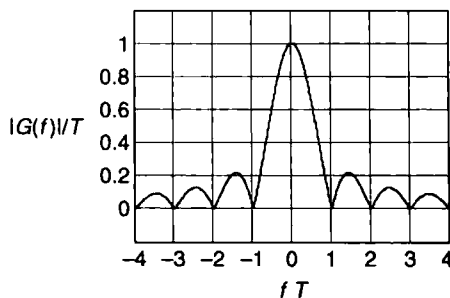


图 11.2 矩形基带脉冲的频谱

^① 注意, 定义式(11.1)导致带通信号的能量为基带信号能量的一半。为了获得同等的能量, 就必须定义 $E = \|s_{\text{BP}}\|^2 = 0.5 \|s_{\text{LP}}\|^2$ 。而且要求通带内滤波器的冲激响应 $h_{\text{filter, BP}}(t)$ 的等效基带表示形式为 $h_{\text{filter, BP}}(t) = 2\text{Re}\{h_{\text{filter, LP}}(t) \exp[j2\pi f_c t]\}$, 从而确保了滤波器的输出具有和输入信号同样的能量标准。尽管这样使用有些不符合逻辑, 但这种表示方法在文献中使用较为广泛。也有一些文献, 如 Barry et al. [2003] 使用 $s_{\text{BP}}(t) = \sqrt{2}\text{Re}\{s_{\text{LP}}(t) \exp[j2\pi f_c t]\}$ 。

奈奎斯特脉冲

矩形脉冲的频谱分布在很宽的带宽上,其第一个旁瓣峰值仅比主瓣峰值衰减 13 dB,这导致邻道干扰很大,从而使蜂窝系统的频谱效率大大降低,因此通常需要一个在频域具有较大滚降的脉冲波形。最常见的具有较大频谱滚降的一类基带脉冲是奈奎斯特脉冲,即满足奈奎斯特准则的一类脉冲,这样就不会引起码间干扰^①。

下面以升余弦脉冲为例加以介绍(见图 11.3)。该脉冲的频谱具有余弦滚降特性,决定频谱衰减陡度的参数称为滚降系数 α 。定义函数为

$$G_{N0}(f, \alpha, T) = \begin{cases} 1, & 0 \leq |2\pi f| \leq (1 - \alpha) \frac{\pi}{T} \\ \frac{1}{2} \cdot \left(1 - \sin \left(\frac{T}{2\alpha} \left(|2\pi f| - \frac{\pi}{T} \right) \right) \right), & (1 - \alpha) \frac{\pi}{T} \leq |2\pi f| \leq (1 + \alpha) \frac{\pi}{T} \\ 0, & (1 + \alpha) \frac{\pi}{T} \leq |2\pi f| \end{cases} \quad (11.5)$$

升余弦脉冲的频谱可以表示为

$$G_N(f, \alpha, T) = \frac{T}{\sqrt{1 - \alpha/4}} \cdot G_{N0}(f, \alpha, T) \exp(-j\pi f T_S) \quad (11.6)$$

其中,归一化因子 $\frac{T}{\sqrt{1 - \alpha/4}}$ 来源于式(11.3)。

在实际应用中,发射端滤波器和接收端滤波器级联即可产生升余弦脉冲,因此接收端滤波器之后观察到的脉冲形状满足奈奎斯特准则。由于接收端滤波器必须和发送波形匹配,所以发射端和接收端滤波器的频谱必须都是升余弦滤波器频谱的平方根,这种滤波器就称为平方根升余弦滤波器,用下标 NR 表示。

升余弦和平方根升余弦的脉冲可从傅里叶逆变换中得出(见图 11.4),其时域完整表达式可在 Chennakeshu and Saulnier[1993]中找到。

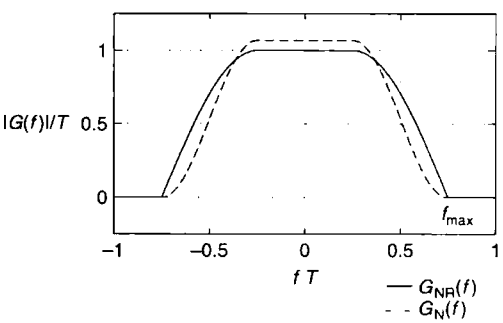


图 11.3 升余弦脉冲 G_N 和平方根升余弦脉冲 G_{NR} 频谱

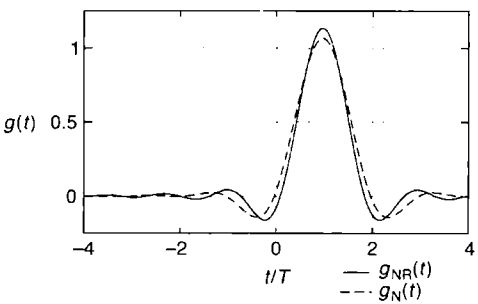


图 11.4 升余弦脉冲和平方根升余弦脉冲

例 11.1 计算升余弦脉冲和平方根升余弦脉冲在 $\alpha = 0.5$ 时的信号功率与邻信道干扰功率比,其中心频率分别为 0 和 $1.25/T$ 。

解:

两个信号在滚降域具有微小的重叠,且由 $\alpha = 0.5$ 可知,信号在 $-0.75/T \leq f \leq 0.75/T$ 范

① 更精确地说,奈奎斯特脉冲是在假设信道不产生码间干扰和采样时钟理想的前提下,不产生码间干扰。

国内具有频谱分量, 在区间 $0.5/T \leq f \leq 0.75/T$ 范围内存在干扰, 不失一般性可假设 $T=1$ 。对于升余弦脉冲, 假设在接收端采用匹配滤波器, 则信号能量正比于(信号和干扰中产生的下降归一化因子):

$$\begin{aligned}
 S &= 2 \int_{f=0}^{0.75/T} |G_N(f, \alpha, T)|^2 df \\
 &= 2 \int_{f=0}^{0.25} |1|^4 df + 2 \int_{f=0.25}^{0.75} \left| \frac{1}{2} \cdot \left(1 - \sin \left(\frac{2\pi}{2\alpha} \left(|f| - \frac{1}{2} \right) \right) \right) \right|^4 df \\
 &= 0.77
 \end{aligned} \tag{11.7}$$

干扰信号能量(假设上下相邻信道均有干扰)正比于:

$$\begin{aligned}
 I &= 2 \int_{f=0.5}^{f=0.75} |G_N(f, \alpha, T) G_N(f - 1.25, \alpha, T)|^2 df \\
 &= 2 \int_{f=0.5}^{f=0.75} \left| \frac{1}{2} \cdot \left(1 - \sin \left[2\pi \left(|f| - \frac{1}{2} \right) \right] \right) \right| \frac{1}{2} \cdot \left(1 - \sin \left[2\pi \left(|f - 1.25| - \frac{1}{2} \right) \right] \right) \right|^2 df \\
 &= 0.9 \times 10^{-4}
 \end{aligned} \tag{11.8}$$

则信干比为 $10 \log_{10}(S/I) \approx 39$ dB。对于平方根升余弦脉冲, 信号能量为

$$S = 2 \int_{f=0}^{f=0.75/T} |G_N(f, \alpha, T)|^2 df = 0.875 \tag{11.9}$$

干扰能量为

$$\begin{aligned}
 I &= 2 \int_{f=0.5}^{f=0.75} |G_N(f, \alpha, T) G_N(f - 1.25, \alpha, T)| df \\
 &= 2 \int_{f=0.5}^{f=0.75} \left| \frac{1}{2} \cdot \left(1 - \sin \left[2\pi \left(|f| - \frac{1}{2} \right) \right] \right) \right| \frac{1}{2} \cdot \left(1 - \sin \left[2\pi \left(|f - 1.25| - \frac{1}{2} \right) \right] \right) \right| df \\
 &= 5.6 \times 10^{-3}
 \end{aligned} \tag{11.10}$$

则信干比为 $10 \log_{10}(S/I) \approx 22$ dB。显然, 在频域衰减较为平坦的平方根升余弦滤波器导致了更差的信干比。

11.2.2 多脉冲调制和连续相位调制

脉冲幅度调制可以进一步推广为多脉冲调制, 其信号由一组基带脉冲组成, 待传输的脉冲取决于调制系数 c_i :

$$s_{LP}(t) = \sum_{i=-\infty}^{\infty} g_{c_i}(t - iT) \tag{11.11}$$

M 阶频移键控(MFSK)就是一个例子: 其基带脉冲与载波频率之间的频率偏移量为 $f_{\text{mod}} =$

$i\Delta f/2, i = \pm 1, \pm 3, \dots, \pm (M-1)$ 。通常选择一组正交的或双正交的脉冲,以简化检测器。

在连续相位频移键控(CPSK)中,在某一时间发射的波形不仅与当前符号有关,而且与发射信号的历史有关。尤其是调制符号之间的相互作用应按照某种方式变化,以保证整个信号的相位是连续的。整个信号具有恒定包络,其相位 $\Phi(t)$ 可以表示为

$$\Phi_{\text{CPFSK}}(t) = 2\pi h_{\text{mod}} \sum_{i=-\infty}^{\infty} c_i \int_{-\infty}^t \tilde{g}(u - iT) du \tag{11.12}$$

其中, u 为积分变量, h_{mod} 为调制指数, $\tilde{g}(t)$ 为基带脉冲波形,归一化为

$$\int_{-\infty}^{\infty} \tilde{g}(t) dt = 1/2 \tag{11.13}$$

需要指出的是,归一化后的基本相位脉冲与脉冲幅度调制中的基本脉冲本质上不同,与信号的能量无关。

高斯基带脉冲

高斯基带脉冲为矩形脉冲和高斯函数的卷积。换句话说,高斯基带脉冲就是一个具有高斯冲激响应的滤波器在输入为矩形波形时的输出(见图 11.5)。矩形脉冲波形由式(11.4)给出,高斯滤波器的冲激响应在数学上可表示为

$$\frac{1}{\sqrt{2\pi}\sigma_G T} \exp\left(-\frac{t^2}{2\sigma_G^2 T^2}\right) \tag{11.14}$$

其中,

$$\sigma_G = \frac{\sqrt{\ln(2)}}{2\pi B_G T} \tag{11.15}$$

B_G 为高斯滤波器的 3 dB 带宽。高斯滤波器的传输函数为

$$\exp\left(-\frac{(2\pi f)^2 \sigma_G^2 T^2}{2}\right) \tag{11.16}$$

应用式(11.13)中的归一化条件得

$$\tilde{g}_G(t) = \frac{1}{4T} \left[\text{erfc}\left(\frac{2\pi}{\sqrt{2\ln(2)}} B_G T \left(-\frac{t}{T}\right)\right) - \text{erfc}\left(\frac{2\pi}{\sqrt{2\ln(2)}} B_G T \left(1 - \frac{t}{T}\right)\right) \right] \tag{11.17}$$

其中, $\text{erfc}(x)$ 为互补误差函数:

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt \tag{11.18}$$

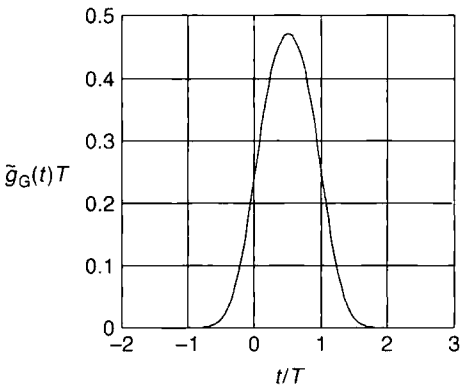


图 11.5 具有 $B_G T = 0.5$ 的高斯基带脉冲波形

11.2.3 功率谱

占用带宽

传输信号的带宽是调制方式的一个重要参数,对于频谱资源非常珍贵的无线信号更是如此。在进行深入讨论前,首先需要讲清楚“带宽”的概念,这是因为对于带宽有如下多种不同的定义。

- 噪声带宽定义为一个具有矩形传输函数 $|H_{\text{rect}}(f)|$ (且恒定的峰值振幅为 $\max(|H(f)|)$) 的系统的带宽, 且其噪声功率与实际系统接收噪声功率相同。
 - 3 dB 带宽定义为 $|H(f)|^2$ 下降到比峰值低 3 dB 时的带宽。
 - 90% 能量带宽定义为包含整个发射能量 90% 的带宽; 类似地, 可以定义 99% 能量带宽或者包含其他百分比能量的带宽。
- 频带利用率定义为数据(比特)速率和所占带宽之比。

循环平稳随机过程的功率谱密度

循环平稳随机过程 $x(t)$ 定义为一类均值函数和自相关函数都以 T_{per} 为周期的随机过程, 即

$$\left. \begin{aligned} E\{x(t + T_{\text{per}})\} &= E\{x(t)\} \\ R_{x,x}(t + T_{\text{per}} + \tau, t + T_{\text{per}}) &= R_{x,x}(t + \tau, t) \end{aligned} \right\} \quad (11.19)$$

大多数调制方式都具有这些特性, 周期 T_{per} 即为符号周期 T_s 。这样一个脉冲幅度调制已调信号的功率谱密度就可以由基带脉冲的功率谱 $|S_G(f)|^2$ 和数据的谱密度 σ_s^2 乘积确定:

$$S_{\text{LP}}(f) = \frac{1}{T_{\text{per}}} \cdot |S_G(f)|^2 \sigma_s^2(f) \quad (11.20)$$

基带脉冲的功率谱密度就是基带脉冲 $g(t)$ 的傅里叶变换 $G(f)$ 的模平方:

$$|S_G(f)|^2 = |G(f)|^2 \quad (11.21)$$

带通信号的功率谱密度为

$$S_{\text{BP}}(f) = \frac{1}{2} [S_{\text{LP}}(f - f_c) + S_{\text{LP}}(-f - f_c)] \quad (11.22)$$

进一步假设数据符号的均值为零且不相干, 则此时的数据符号的功率谱密度为均匀的, $\sigma_s(f) = \sigma_s^{\text{①}}$ 。注意, 连续相位频移键控信号具有记忆性, 且因此使其码元之间具有相关性, 所以其功率谱密度的计算将更为复杂。详细介绍见 Proakis[2005]。

11.2.4 信号空间图

信号空间图用模拟发送符号表示有限维空间中的向量(点), 它是分析调制方式的非常重要的工具之一, 它对各种不同的调制方式得到的信号进行了直观和统一的图形表示。第 12 章将详细讨论如何通过信号空间图计算误比特率。现在主要将其作为一种简便的方法来表示各种不同的调制方式。

为了便于说明, 假设基带脉冲为矩形基带脉冲 $q(t) = q_n(t, T)$ 。在第 i 个码元期间, $iT_s < t < (i+1)T_s$ 的信号 $s(t)$ 组成了一个有限函数集(不失一般性, 下面均假设 $i=0$)。这个有限集的大小为 M 。这种表示方法同样适用于脉冲幅度调制和多脉冲调制。

现在选取一个定义在区间 $[0, T_s]$ 上的正交扩展函数^② $\varphi_n(t)$ 集(大小为 N)。扩展函数集必须是完备的, 这样才能保证传输信号都能够用扩展函数的线性组合表示。这种完备的扩展函数集可以通过一种 Gram-Schmidt 正交化过程得到(Wozencraft and Jacobs [1965]给出了示例)。

① 未编码数据通常假定是不相关的。数据编码增加了数据之间的相关性(见第 14 章)。许多系统采用扰码(采用伪随机数据与编码数据相乘)来消除发送符号之间的相关性。

② 扩展函数通常又称为“基函数”。我们没有使用“基函数”这个名字是为了避免与“基脉冲”混淆。很显然, 扩展函数都是定义为正交形式的, 而基脉冲却不一定必须是正交的。

给定一组扩展函数集 $\{\varphi_n(t)\}$,任一复(基带)传输信号 $s_m(t)$ 都可以表示为一个向量 $\mathbf{s}_m = (s_{m,1}, s_{m,2}, \dots, s_{m,N})$,其中

$$s_{m,n} = \int_0^{T_S} s_m(t) \varphi_n^*(t) dt \quad (11.23)$$

其中,上标 $*$ 表示复共轭。而带通信号的向量分量为

$$s_{BP,m,n} = \int_0^{T_S} s_{BP,m}(t) \varphi_{BP,n}(t) dt \quad (11.24)$$

那么实际传输的信号可以从向量分量(无论是通带信号还是基带信号)中获得:

$$s_m(t) = \sum_{n=1}^N s_{m,n} \varphi_n(t) \quad (11.25)$$

由于每个信号均可以通过向量 $\mathbf{s}_m$ 表示,所以可在所谓的信号空间图^①上画出这些向量。当 $N=2$ 时,用图形表示这些点非常方便,而 $N=2$ 包括了多数重要的调制方式。

对于带通表示,具有矩形基本脉冲波形的脉冲幅度调制的扩展函数为

$$\left. \begin{aligned} \varphi_{BP,1}(t) &= \sqrt{\frac{2}{T_S}} \cos(2\pi f_c t) \\ \varphi_{BP,2}(t) &= \sqrt{\frac{2}{T_S}} \sin(2\pi f_c t) \end{aligned} \right\} \quad i T_S \leq t < (i+1) T_S; 0 \quad \text{其他} \quad (11.26)$$

其中假设 $f_c \gg 1/T_S$ 。这就意味着所有包含 $\cos(2 \cdot 2\pi f_c t)$ 的乘积项都被忽略或被滤波器滤除。

例如,对于二进制双极信号(二进制相移键控,见11.3.1节),可以表示为

$$s_{BP,1} = \pm \sqrt{\frac{2E_S}{T_S}} \cos(2\pi f_c t) \quad (11.27)$$

其在信号空间图上的点位于 $\pm \sqrt{E_S}$, E_S 为码元能量。

一般地,一个码元内的能量可以通过带通信号计算出来:

$$E_{S,m} = \int_0^{T_S} s_{BP,m}^2(t) dt = \|\mathbf{s}_{BP,m}\|^2 \quad (11.28)$$

其中, $\|\mathbf{s}\|$ 表示 $\mathbf{s}$ 的 L_2 范数(欧氏范数)。值得注意的是,对于很多调制方式(例如二进制相移键控)来说, E_S 与 m 无关。

$s_m(t)$ 和 $s_k(t)$ 之间的相关系数在带通形式时可以计算如下:

$$\text{Re}\{\rho_{k,m}\} = \frac{\mathbf{s}_{BP,m} \mathbf{s}_{BP,k}}{\|\mathbf{s}_{BP,m}\| \|\mathbf{s}_{BP,k}\|} \quad (11.29)$$

两信号之间的欧氏距离的平方为

$$d_{k,m}^2 = [E_{S,m} + E_{S,k} - 2\sqrt{E_{S,m} E_{S,k}} \text{Re}\{\rho_{k,m}\}] \quad (11.30)$$

后面将会看到这个距离对于计算误比特率非常重要。

如果使用复基带表示,那么信号的空间图可以通过扩展向量得到:

$$\left. \begin{aligned} \varphi_1(t) &= \sqrt{\frac{1}{T_S}} \cdot 1 \\ \varphi_2(t) &= \sqrt{\frac{1}{T_S}} \cdot j \end{aligned} \right\} \quad i T_S \leq t < (i+1) T_S; 0 \quad \text{其他} \quad (11.31)$$

① 更准确地讲,向量的端点从系统的坐标原点开始。

很重要的一点是,带通和低通情况下得到的信号空间图存在一个 $\sqrt{2}$ 的因子差异。因而,信号的能量可以表示为

$$E_{S,m} = \frac{1}{2} \int_0^{T_s} \|s_{LP,m}(t)\|^2 dt \approx \frac{1}{2} \|s_{LP,m}\|^2 \quad (11.32)$$

$s_{LP,m}$ 就是从等效基带表示中得到的信号向量。相关系数为

$$\rho_{k,m} = \frac{s_{LP,m}(s_{LP,k})^*}{\|s_{LP,m}\| \|s_{LP,k}\|} \quad (11.33)$$

11.3 几种重要的调制

本节将简要总结不同数字调制方案的重要性质。对于每一种调制方案将给出如下几项信息:

- 带通及基带信号的时域函数;
- 脉冲幅度调制或多脉冲调制的表示方式;
- 信号空间图;
- 频谱效率。

11.3.1 二进制相移键控

二进制相移键控(BPSK)是一种最简单的调制方法:根据所发送的码元 a 为 $+1$ 还是 -1 ,载波相位改变 $\pm\pi/2$ ^①。尽管二进制相移键控非常简单,但却有两种不同的表示方法。一种是将其看成相位调制,此时数据流影响传输信号的相位,根据数据比特 b_i ,发射信号的相位为 $\pi/2$ 或 $-\pi/2$ 。

另一种更常用的表示方法是将 BPSK 看成脉冲幅度调制,其基带脉冲为幅度为 1 的矩形脉冲,因此有

$$s_{BP}(t) = \sqrt{2E_B/T_B} p_D(t) \cos\left(2\pi f_c t + \frac{\pi}{2}\right)$$

其中,

$$p_D(t) = \sum_{i=-\infty}^{\infty} b_i g(t - iT) = b_i * g(t) \quad (11.34)$$

且

$$g(t) = g_R(t, T_B) \quad (11.35)$$

图 11.6 给出了信号时域波形,图 11.7 给出了信号空间图。在等效基带表示中,复调制符号为 $\pm j$:

$$c_i = j \cdot b_i \quad (11.36)$$

则信号的实部为

$$\text{Re}\{s_{LP}(t)\} = 0 \quad (11.37)$$

虚部为

$$\text{Im}\{s_{LP}(t)\} = \sqrt{\frac{2E_B}{T_B}} p_D(t) \quad (11.38)$$

除了在 $t = iT_B$ 时刻外,信号包络具有恒定幅度。由于使用的是未经滤波的矩形脉冲作为基带脉冲,所以信号频谱的滚降非常缓慢。考虑 90% 能量带宽时,带宽效率可达 0.59 bps/Hz;但当考虑 99% 能量带宽时,带宽效率仅为 0.05 bps/Hz(见图 11.8)。

^① 严格地讲,参考相位 φ 也必须给出,这决定了 $t=0$ 时刻的相位。一般没有限定时均假设 $\varphi_{s,0}=0$ 。

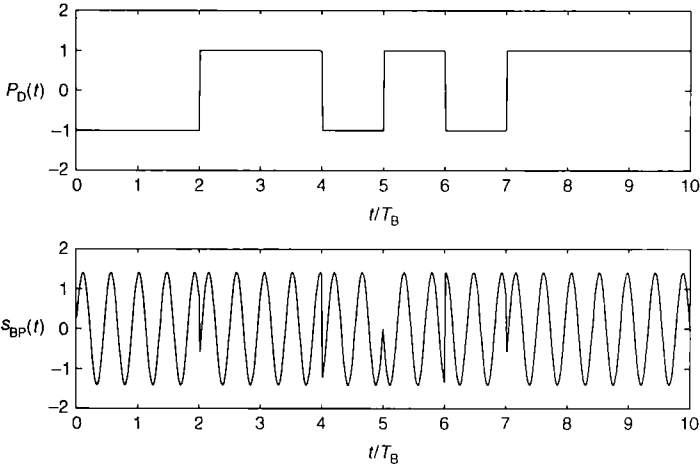


图 11.6 二进制相移键控信号的时域波形

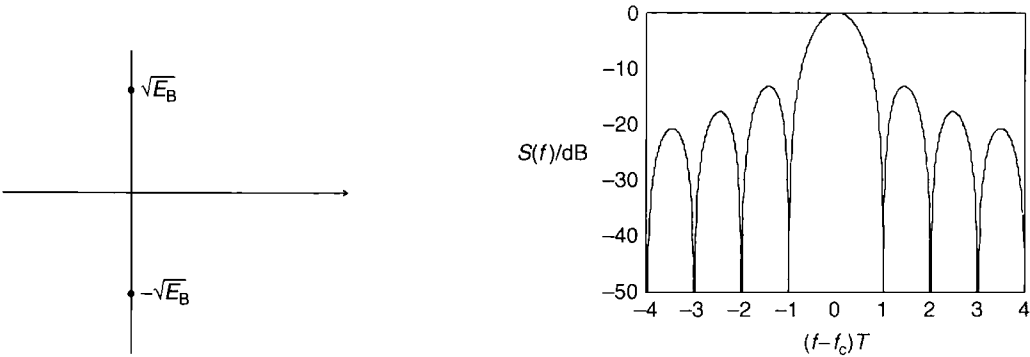


图 11.7 二进制频移键控信号空间图

图 11.8 二进制频移键控的归一化功率谱密度

由于矩形脉冲的带宽效率太低,实际的发送端通常使用奈奎斯特脉冲作为基带脉冲^①。即使采用 $\alpha=0.5$ 的相对缓慢的滤波,都会使频谱效率得到很大提高:对于 90% 和 99% 的能量带宽,频谱利用率分别达到 1.02 bps/Hz 和 0.79 bps/Hz,但另一方面却使信号不再具有恒定包络(见图 11.9 和图 11.10)。

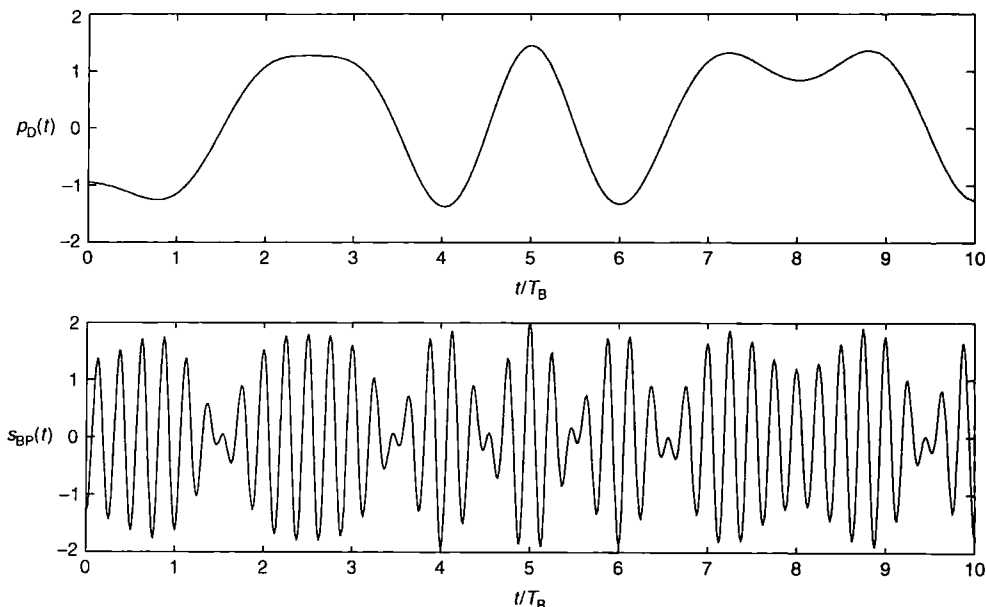
相移键控的一个重要的改进就是差分相移键控(DPSK)。其基本思想就是传输的相位不是仅由当前码元决定,而是等于前一码元的相位加上当前码元引起的相位增量。对于二进制相移键控,这种变换就是一种特别简单的形式,首先将数据比特进行如下编码:

$$\tilde{b}_i = b_i + b_{i-1} \tag{11.39}$$

再用 $\tilde{b}_i$ 替换式(11.34)中的 b_i 。

差分编码的优点是能够应用差分解码器,而差分解码器只需要比较前后两个码元的相位就能解调接收到的信号。这样就避免了要恢复接收信号绝对相位的难题,从而简化了接收机,降低了制造成本。

① 值得注意的是,以奈奎斯特脉冲为基带脉冲的脉冲幅度调制和以奈奎斯特脉冲为相位脉冲的相位调制是不同的以后只考虑以奈奎斯特脉冲为基带脉冲的脉冲幅度调制,并简记为 BAM(Binary Amplitude Modulation, 二进制幅度调制)。

图 11.9 二进制幅度调制信号时域波形(滚降系数 $\alpha = 0.5$)

11.3.2 正交相移键控

正交相移键控(QPSK)已调信号是同相分量和正交分量在每个符号期间各承载 1 比特信息的一种脉冲幅度调制。原始的数据流被分成 b_{1i} 和 b_{2i} 两个数据流:

$$\left. \begin{aligned} b_{1i} &= b_{2i} \\ b_{2i} &= b_{2i+1} \end{aligned} \right\} \quad (11.40)$$

每个数据流的数据速率都是原来数据流速率的一半:

$$R_S = 1/T_S = R_B/2 = 1/(2T_B) \quad (11.41)$$

首先考虑基带脉冲为矩形脉冲的情况, 即 $g(t) = g_R(t, T_S)$ 。这时可以将正交相移键控信号看成一种相位调制或脉冲幅度调制。首先定义两个序列的矩形脉冲(见图 11.11):

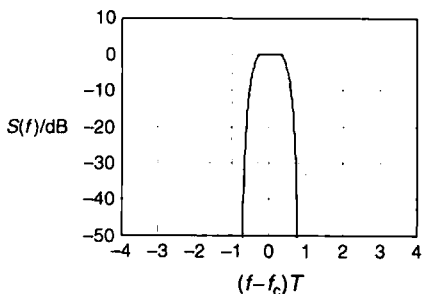
$$\left. \begin{aligned} p_{1D}(t) &= \sum_{i=-\infty}^{\infty} b_{1i} g(t - iT_S) = b_{1i} * g(t) \\ p_{2D}(t) &= \sum_{i=-\infty}^{\infty} b_{2i} g(t - iT_S) = b_{2i} * g(t) \end{aligned} \right\} \quad (11.42)$$

将正交相移键控看成脉冲幅度调制时, 带通信号可表示为

$$s_{BP}(t) = \sqrt{E_B/T_B} [p_{1D}(t) \cos(2\pi f_c t) - p_{2D}(t) \sin(2\pi f_c t)] \quad (11.43)$$

一个码元间隔内的能量归一化为 $\int_0^{T_S} s_{BP}(t)^2 dt = 2E_B$, 其中 E_B 指传输 1 比特所需的能量。

图 11.12 给出了信号空间图, 基带信号为

图 11.10 二进制幅度调制信号归一化功率谱密度(滚降系数 $\alpha = 0.5$)

$$s_{LP}(t) = [p1_D(t) + jp2_D(t)]\sqrt{E_B/T_B}$$
 (11.44)

若将正交相移键控解释为相位调制, 则低通信号可以表示为 $\sqrt{2E_B/T_B}\exp(j\Phi_s(t))$, 且

$$\Phi_S(t) = \pi \cdot \left[\frac{1}{2} \cdot p2_D(t) - \frac{1}{4} \cdot p1_D(t) \cdot p2_D(t) \right]$$
 (11.45)

从上式显然可以看出, 传输信号除了在转换时刻 $t = iT_s$ 外都具有恒包络(见图 11.13)。

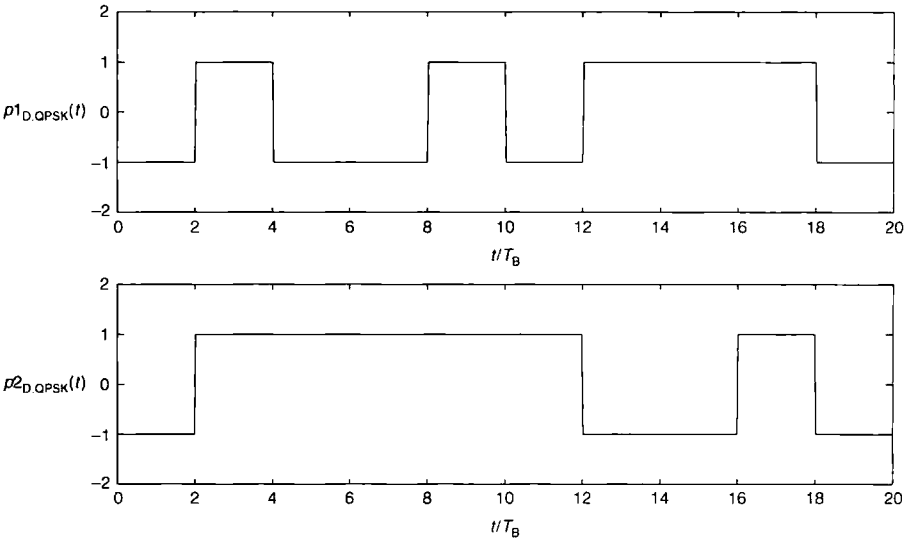


图 11.11 正交相移键控的同相数据分量和正交数据分量

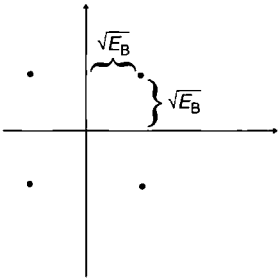


图 11.12 正交相移键控的信号空间图

正交相移键控的频谱效率是二进制相移键控的两倍, 这是因为正交相移键控的同相分量和正交分量都被用于信息传输。这意味着考虑 90% 能量带宽时, 正交相移键控的带宽效率为 1.1 bps/Hz; 考虑 99% 能量带宽时, 带宽效率为 0.1 bps/Hz(见图 11.14)。与二进制相移键控相似, 缓慢的频谱滚降问题使系统需要采用升余弦基带脉冲(见图 11.15)。下面将讲到的采用升余弦基带脉冲的正交幅度调制方式, 其频谱效率分别增加到 2.04 bps/Hz 和 1.58 bps/Hz(见图 11.16)。但是, 信号包络却起伏较大(见图 11.17)。

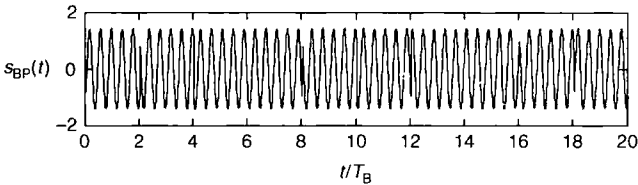


图 11.13 正交相移键控的时域波形

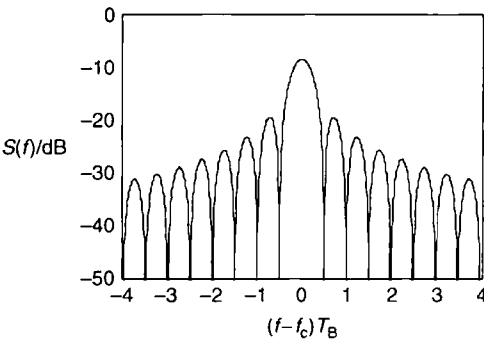


图 11.14 正交相移键控的归一化功率谱密度

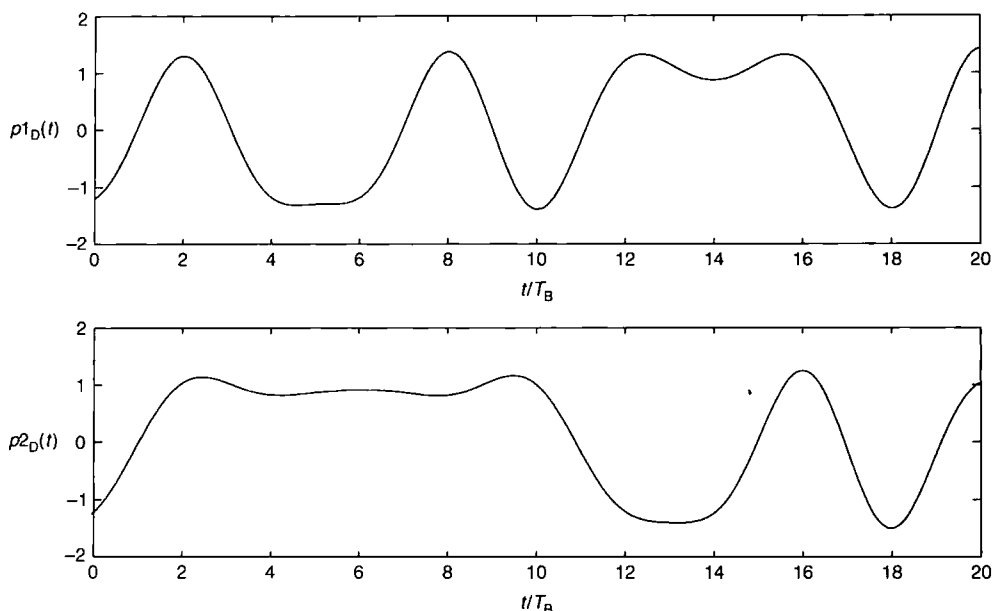
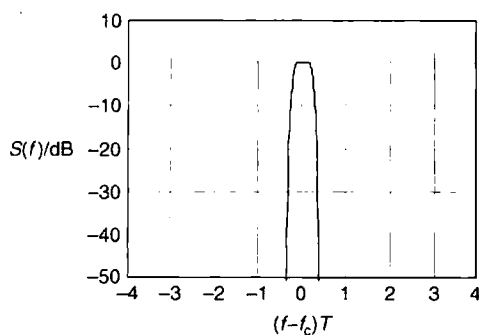
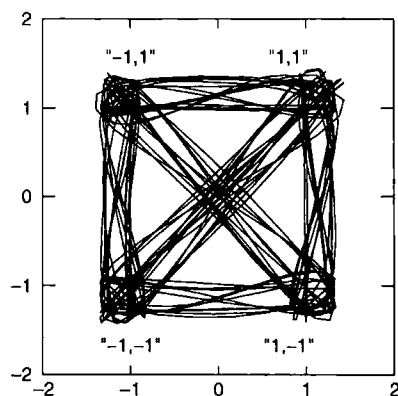


图 11.15 正交幅度调制的脉冲序列

图 11.16 采用 $\alpha = 0.5$ 的升余弦脉冲的正交幅度调制的归一化功率谱密度图 11.17 升余弦脉冲正交幅度调制的 I-Q 分量眼图。同时给出了归一化信号空间图中的 4 个归一化点 $(1, 1)$, $(1, -1)$, $(-1, -1)$ 和 $(-1, 1)$

11.3.3 $\pi/4$ -DQPSK

尽管正交相移键控看上去是一种恒包络调制,但其幅度在比特转换点仍然存在凹陷点,这也可以从 I-Q 眼图中有一些比特转换时通过原点的轨迹看出。当使用非矩形基带脉冲时,衰减持续的时间将更长。信号包络的这种变化是不希望出现的,因为这会使相配的放大器设计非常困难。解决这一问题的一种可行方法是使用 $\pi/4$ -DQPSK。这种调制对于第二代蜂窝通信是非常重要的,已经在北美的一些标准(如 IS-54, IS-136, PWT)、日本的蜂窝电话(JDC)和无绳电话(PHS)标准,以及欧洲的陆地集群无线标准(TETRA)中得到了应用。

$\pi/4$ -DQPSK 的原理可以从差分正交相移键控的信号空间图来理解(见图 11.18)。这里存

在两组信号星座：(0°, 90°, 180°和 270°)和(45°, 135°, 225°和 315°)。具有偶时序号的所有码元选择第一组，奇时序号的所有码元则选择第二组。也就是说，无论 t 是不是码元间隔的整数倍，传输信号的相位都以 $\pi/4$ 增加，而且相位的变化只取决于传输码元。所以前后信号星座图的转移从不过原点(见图 11.19)，从物理意义上讲，这就意味着信号包络具有较小的起伏。

信号的相位为

$$\Phi_s(t) = \pi \left[\frac{1}{2} p_{2D}(t) - \frac{1}{4} p_{1D}(t) p_{2D}(t) + \frac{1}{4} \left\lfloor \frac{t}{T_s} \right\rfloor \right]$$

(11.46)

其中， $\lfloor x \rfloor$ 表示小于或等于 x 的最大整数。与式(11.45)比较，可以明显看出 T_s 的各个整数倍时刻的相位变化。图 11.20 给出了基本的数据序列，图 11.21 画出了使用矩形和升余弦基带脉冲得到的最终带通信号波形。

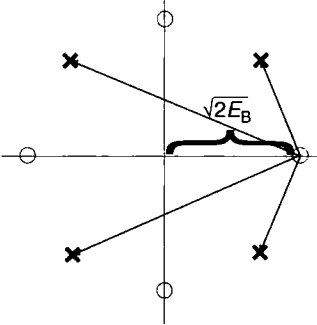


图 11.18 $\pi/4$ -DQPSK 信号空间
图中允许的所有转移

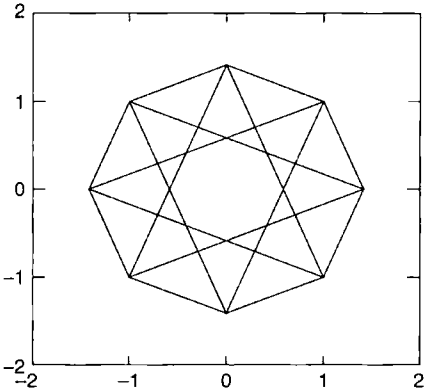


图 11.19 采用矩形基带脉冲的
 $\pi/4$ -DQPSK的I-Q眼图

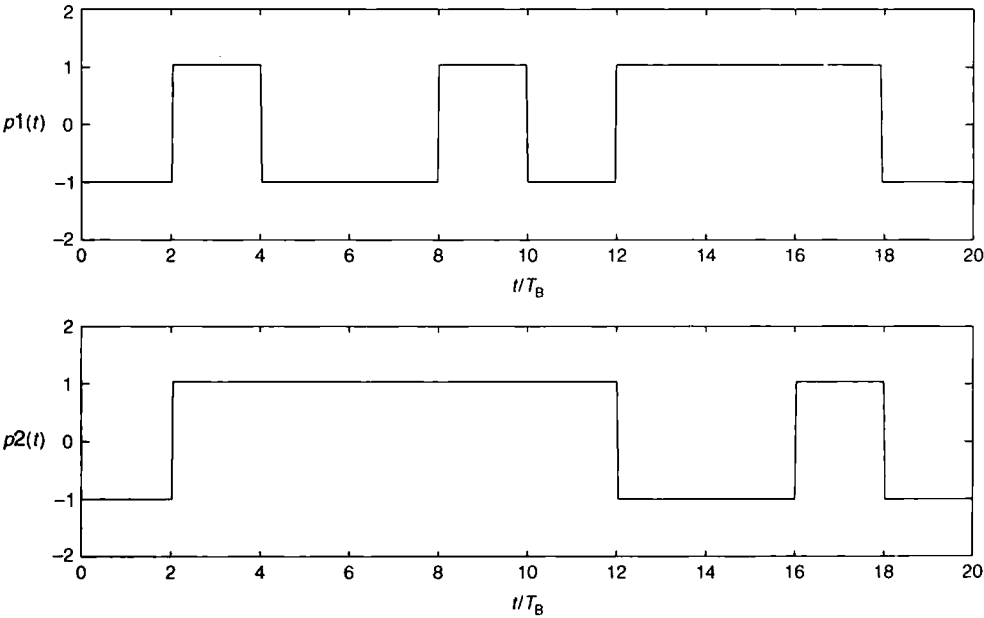
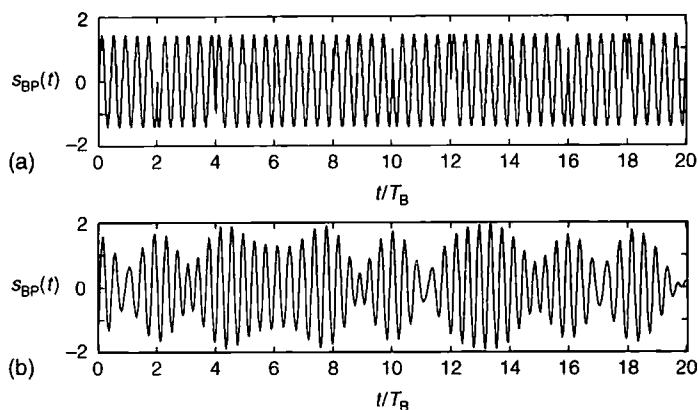


图 11.20 $\pi/4$ -DQPSK 的基脉冲序列

图 11.21 采用(a)矩形基带脉冲和(b)升余弦基带脉冲的 $\pi/4$ -DQPSK 时域波形

11.3.4 偏移正交相移键控

另一种提高正交相移键控峰平比的方法就是确保同相分量和正交分量的比特传输在不同的时刻, 这种方法就是偏移正交相移键控(OQPSK)。通过调制, 使同相和正交分量的数据流之间偏移半个码元间隔(见图 11.22), 这样同相分量的传输发生在码元间隔的整数倍时刻(比特间隔的偶数倍), 而正交分量的传输延迟半个码元间隔(即 1 比特间隔)。所以传输的脉冲流就是

$$\left. \begin{aligned} p_{1D}(t) &= \sum_{i=-\infty}^{\infty} b_{1i} g(t - iT_S) = b_{1i} * g(t) \\ p_{2D}(t) &= \sum_{i=-\infty}^{\infty} b_{2i} g\left(t - \left(i + \frac{1}{2}\right) T_S\right) = b_{2i} * g\left(t - \frac{T_S}{2}\right) \end{aligned} \right\} \quad (11.47)$$

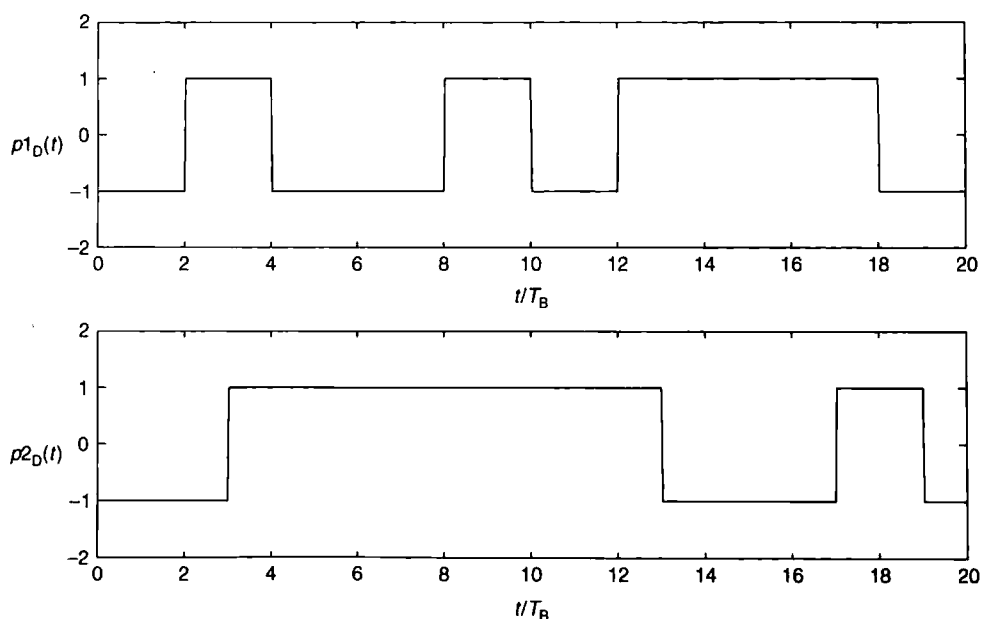


图 11.22 偏移正交相移键控的基脉冲序列

这些数据流也可以表示为脉冲幅度调制[式(11.44)]或相位调制[式(11.45)]。得到的最终带通信号如图 11.23 所示。

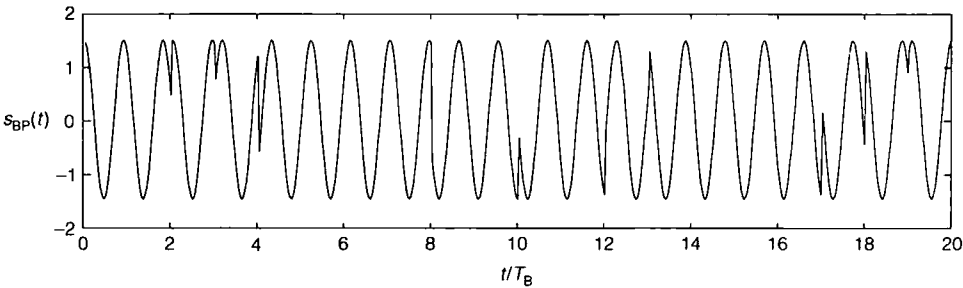


图 11.23 偏移正交相移键控的时域波形

在信号包络图(见图 11.24)中,可以清楚地看到没有传输信号通过坐标原点,因此这种调制方式能够较好地抵抗包络波动。

和一般的正交相移键控一样,也可以使用升余弦基带脉冲等平滑的基带脉冲来改善频谱效率。图 11.25 给出了采用升余弦基带脉冲波形的 I-Q 眼图。

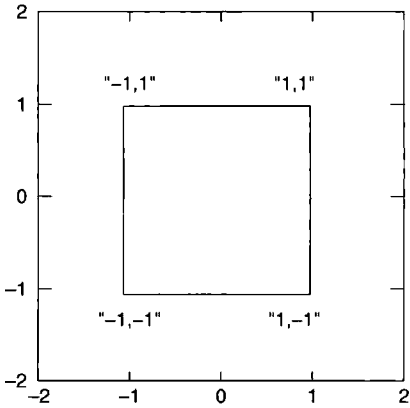


图 11.24 采用矩形基带脉冲的偏移正交相移键控的 I-Q 图。同时给出了归一化信号空间图中的四个归一化点(1,1),(1,-1),(-1,-1)和(-1,1)

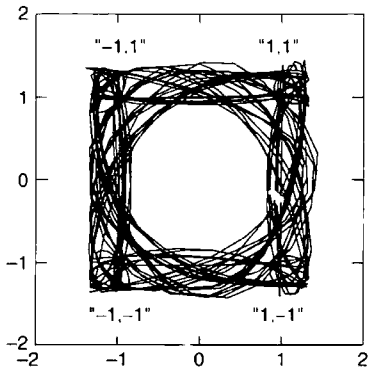


图 11.25 采用升余弦脉冲的偏移正交相移键控的 I-Q 眼图

11.3.5 高阶调制

前面研究了每个码元间隔最多传输 2 比特的调制方式。本节将讨论如何将这些调制方式推广,使其在每个码元间隔中传输更多的信息,这样可以获得更高的频谱效率,但也对噪声和干扰更加敏感,所以它们在以前的无线通信中并不常用。然而,多电平正交幅度调制只是在最近才应用于无线局域网标准(见第 29 章)和第四代蜂窝系统(见第 27 章和第 28 章)中。

高阶正交幅度调制

高阶正交幅度调制(QAM)在同相和正交分量上发送多个比特。它通过在每个分量上传输具有正负极性的多个电平信号来实现。高阶正交幅度调制的数学表示类似于 4-QAM,只是在式(11.42)中的脉冲序列的电平不仅仅是 ± 1 ,而是 $2m-1-\sqrt{M}$,其中 $m=1,\cdots,\sqrt{M}$ 。

16-QAM 的信号空间图如图 11.26 所示。64-QAM 和 256-QAM 之类的更大的星座图也可以

按相同的原理构造出来。一般星座图越大, 输出信号的峰平比就越大。

例 11.2 求 16-QAM 信号的平均能量与其空间图中相邻两点之间距离的平方的关系。

解:

这里只考虑信号空间图的第一象限。假设信号点位于 $d + jd$, $d + j3d$, $3d + j$ 和 $3d + j3d$, 则信号平均能量为 $E_s = \frac{d^2}{4}(2 + 10 + 10 + 18) = 10d^2$, 两点之间距离的平方为 $4d^2$, 其比值为 2.5。

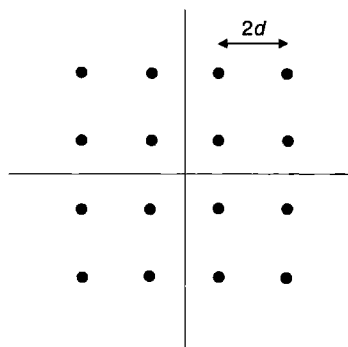


图 11.26 16-QAM 的信号空间图

高阶相位调制

高阶正交幅度调制的缺点是输出信号幅度具有较大的波动, 这就要求系统必须使用线性放大器。另一种选择就是使用高阶相移键控, 其发射信号可以表示为

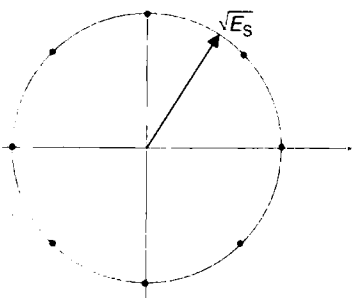


图 11.27 8-PSK 包络图和信号空间图中的 8 个点

$$s_{BP}(t) = \sqrt{2E_s/T_s} \cos\left(2\pi f_c t + \frac{2\pi}{M}(m-1)\right), \quad m = 1, 2, \dots, M \quad (11.48)$$

这是指发送端选择 M 个可能传输相位中的一个(见图 11.27), 注意, 这里按照码元能量和码元间隔对信号能量进行了归一化。等效的低通信号为

$$s_{LP}(t) = \sqrt{2E_s/T_s} \exp\left(j \frac{2\pi}{M}(m-1)\right) \quad (11.49)$$

两个信号的相关系数为

$$\rho_{km} = \exp\left(j \frac{2\pi}{M}(m-k)\right) \quad (11.50)$$

11.3.6 二进制频移键控

在频移键控(FSK)中, 每个码元通过传输一个正弦信号(持续时间为 T_s)来表示, 而其频率取决于所传输的码元。频移键控不能用脉冲幅度调制信号形式表示, 却可以采用多脉冲调制形式表示: 由待传输比特确定中心频率的不同基带脉冲信号(见图 11.28)为

$$g_m(t) = \cos[(2\pi f_c + b_m 2\pi f_{\text{mod}})t/T + \psi] \quad (11.51)$$

注意, 发射信号的相位在比特转换时刻可能存在突跳, 从而导致非理想谱特性。

频移键控的功率谱包含离散部分(谱线)和连续部分:

$$S(f) = S_{\text{cont}}(f) + S_{\text{disc}}(f) \quad (11.52)$$

式中(见 Benedetto and Biglieri[1999]):

$$S_{\text{cont}}(f) = \frac{1}{2T} \left\{ \sum_{m=1}^2 |G_m(f)|^2 - \frac{1}{2} \left| \sum_{m=1}^2 G_m(f) \right|^2 \right\} \quad (11.53)$$

且

$$S_{\text{disc}}(f) = \frac{1}{(2T)^2} \left| \sum_{m=1}^2 G_m(f) \right|^2 \sum_n \delta\left(f - \frac{n}{T}\right) \quad (11.54)$$

其中, $G_m(f)$ 为 $g_m(t)$ 的傅里叶变换, 图 11.29 给出了一个例子。

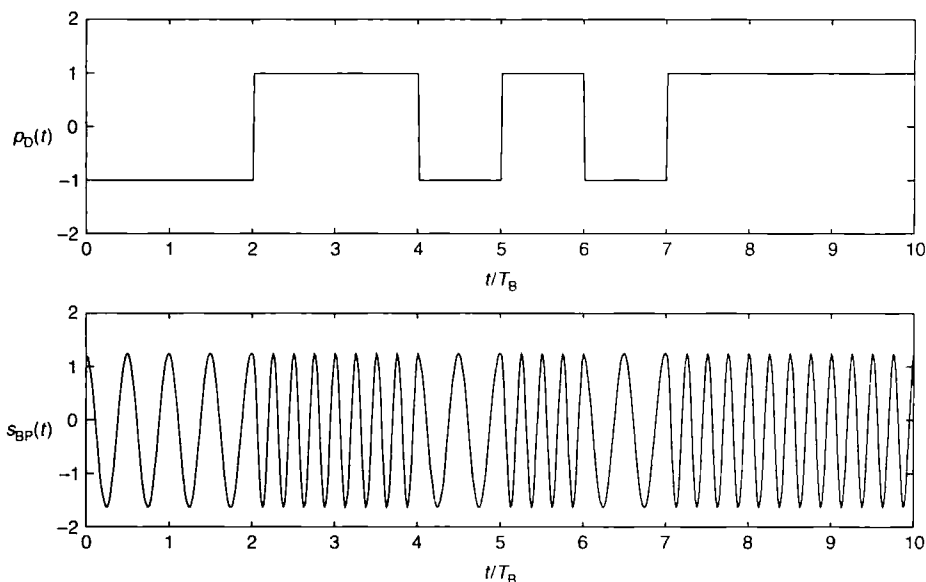


图 11.28 频移键控信号时域波形

连续相位频移键控 (CPFSK) 信号在比特转换时刻使相位变化平滑, 其相位为

$$\Phi_S(t) = 2\pi h_{\text{mod}} \int_{-\infty}^t \tilde{p}_{\text{D.FSK}}(\tau) d\tau \quad (11.55)$$

连续相位频移键控信号具有恒定包络, 其中 h_{mod} 为调制指数。使用相位脉冲的归一化形式 [见式(11.13)], 并假设使用的是矩形相位基带脉冲:

$$\tilde{g}_{\text{FSK}}(t) = \frac{1}{2T_B} g_R(t, T_B) \quad (11.56)$$

相位脉冲序列为

$$\tilde{p}_{\text{D.FSK}}(t) = \sum_{i=-\infty}^{\infty} b_i \tilde{g}_{\text{FSK}}(t - iT_B) = b_i * \tilde{g}_{\text{FSK}}(t) \quad (11.57)$$

瞬时频率就是

$$f(t) = f_c + b_i f_{\text{mod}}(t) = f_c + f_D(t) = f_c + \frac{1}{2\pi} \frac{d\Phi_S(t)}{dt} \quad (11.58)$$

等效的复基带信号的实部和虚部如下:

$$\text{Re}(s_{\text{LP}}(t)) = \sqrt{2E_B/T_B} \cos \left[2\pi h_{\text{mod}} \int_{-\infty}^t \tilde{p}_{\text{D.FSK}}(\tau) d\tau \right] \quad (11.59)$$

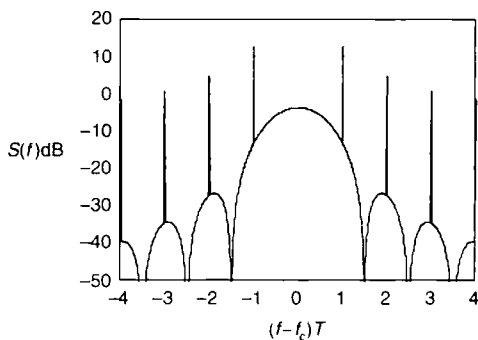


图 11.29 (非连续相位) $h_{\text{mod}} = 1$ 的频移键控调制信号功率谱

$$\text{Im}(s_{\text{LP}}(t)) = \sqrt{2E_B/T_B} \sin \left[2\pi h_{\text{mod}} \int_{-\infty}^t \tilde{p}_{\text{D,FSK}}(\tau) d\tau \right] \quad (11.60)$$

连续相位频移键控信号具有一定的记忆，在时间 t 时刻的信号与前面发送的比特相关。

连续相位频移键控信号的空间图，有如下两种不同的表示形式。

1. 使用在两个可能信号频率 $f_c \pm f_{\text{mod}}$ 上振荡的正弦振荡器，采用相位脉冲的形式来表示，（通带的）展开函数基为

$$\left. \begin{aligned} \varphi_{\text{BP},1}(t) &= \sqrt{2/T_B} \cos(2\pi f_c t + 2\pi f_{\text{mod}} t) \\ \varphi_{\text{BP},2}(t) &= \sqrt{2/T_B} \cos(2\pi f_c t - 2\pi f_{\text{mod}} t) \end{aligned} \right\} \quad (11.61)$$

在这种情况下，信号空间图中垂直的两个坐标轴上存在两个点（见图 11.30）。注意，这里隐含着 $\varphi_{\text{BP},1}(t)$ 和 $\varphi_{\text{BP},2}(t)$ 两个信号相互正交的假设。

2. 使用中心频率为 f_c 的信号同相分量和正交分量。在这种情况下，信号的包络图不再表现为离散的点，而是表现为连续的轨迹，即一个圆（见图 11.31）。在任何时刻，传输的信号与图中不同的点相对应。

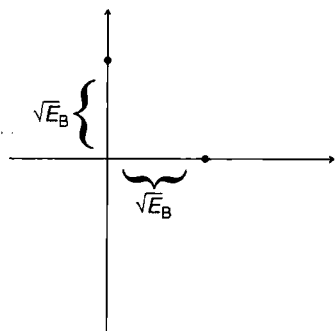


图 11.30 使用基函数 $\cos[2\pi(f_c \pm f_D)t]$ 的频移键控的信号空间图

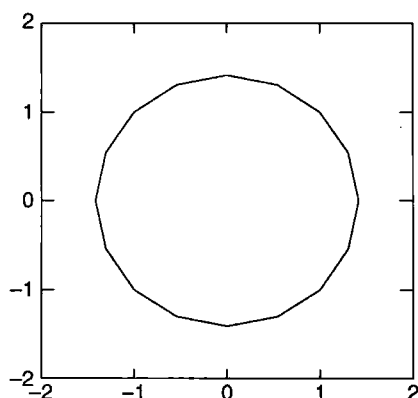


图 11.31 频移键控包络眼图，这等同于使用 $\cos(2\pi f_c t)$ 和 $\sin(2\pi f_c t)$ 基函数的频移键控的信号空间图

11.3.7 最小频移键控

最小频移键控 (MSK) 是无线通信中最重要的调制方式之一。然而最小频移键控存在多种不同的表示方式，容易引起混淆。

1. 第一种表示方式是具有以下调制指数的连续相位频移键控：

$$h_{\text{mod}} = 0.5, \quad f_{\text{mod}} = 1/4T \quad (11.62)$$

这表示相位在 1 比特间隔的时间内变化 $\pm \pi/2$ （见图 11.32）。最小频移键控的带通信号如图 11.33 所示。

2. 另一种方式是将最小频移键控表示为偏移正交幅度调制。如附录 11. A (www.wiley.com/go/molisch) 所述，其基带脉冲为扩展到 $2T_B$ 时间间隔上的半个正弦波（见图 11.34）：

$$g(t) = \sin(2\pi f_{\text{mod}}(t + T_B))q_R(t, 2T_B) \quad (11.63)$$

证明见附录 11. A (www.wiley.com/go/molisch)。

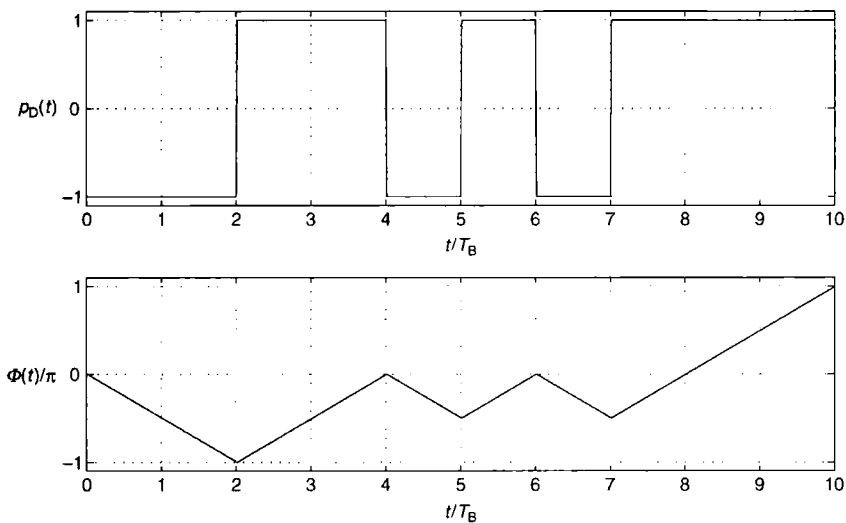


图 11.32 最小频移键控信号的相位脉冲及其相位函数波形

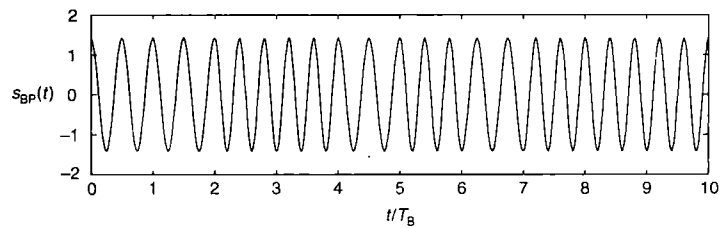


图 11.33 最小频移键控已调信号

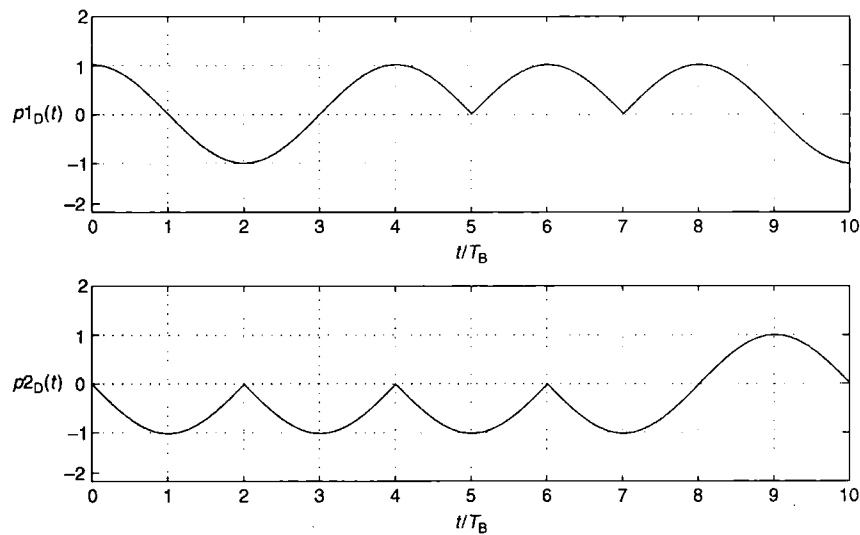


图 11.34 由半波正弦合成的最小频移键控

由于使用了平滑的基带脉冲，最小频移键控的频谱下降速度比采用矩形脉冲的偏移正交相移键控更快：

$$S(f) = \frac{16T_B}{\pi^2} \left(\frac{\cos(2\pi f T_B)}{1 - 16f^2 T_B^2} \right)^2 \tag{11.64}$$

在图 11.35 中也可看到这一点。另一方面, 最小频移键控仅为一种二进制调制, 但偏移正交相移键控的一个码元间隔传输 2 比特。因而, 当考虑 90% 能量带宽时, 最小频移键控的谱效率低 (1.29 bps/Hz), 但是当考虑 99% 能量的带宽时, 最小频移键控的谱效率 (0.85 bps/Hz) 仍然较好。

例 11.3 比较最小频移键控和采用矩形脉冲的正交相移键控的功率谱密度。考虑具有相同比特间隔的系统, 并求出在 $1/T_B$, $2/T_B$ 和 $3/T_B$ 点的带外能量。

解:

最小频移键控的功率谱密度表达式为

$$S_{\text{MSK}}(f) = \frac{16T_B}{\pi^2} \left(\frac{\cos(2\pi f T_B)}{1 - 16f^2 T_B^2} \right)^2 \quad (11.65)$$

而采用矩形脉冲的正交相移键控的功率谱密度和普通的正交幅度调制相同 (注意, 通过归一化使功率谱密度的积分变为 1):

$$S_{\text{OQPSK}}(f) = (1/T_S)(T_S \text{sinc}(\pi f T_S))^2 \quad (11.66)$$

必须注意, 对于正交相移键控, $T_S = 2T_B$ 。对于最小频移键控, 当 $T_B = 1$ 时, 带外功率为

$$\begin{aligned} P_{\text{out}}(f_0) &= 2 \int_{f_0}^{\infty} S(f) df = 2 \int_1^{\infty} \frac{16}{\pi^2} \left(\frac{\cos(2\pi f)}{1 - 16f^2} \right)^2 df \\ &= \frac{32}{\pi^2} \int_1^{\infty} \frac{\cos^2 2\pi f}{256f^4 - 32f^2 + 1} df = \frac{32}{\pi^2} \int_1^{\infty} \frac{\frac{1}{2} \cos 4\pi f + \frac{1}{2}}{256f^4 - 32f^2 + 1} df \end{aligned} \quad (11.67)$$

对于正交相移键控, 当 $T_B = 1$ 时, 带外功率为

$$P_{\text{out}}(f_0) = \int_1^{\infty} \left(\frac{2 \sin(2\pi f)}{(2\pi f)} \right)^2 df \quad (11.68)$$

在 T_B 整数倍时刻的各个带外能量如下:

	$1/T_B$	$2/T_B$	$3/T_B$
QPSK	0.050	0.025	0.017
MSK	0.0024	2.8×10^{-4}	7.7×10^{-5}

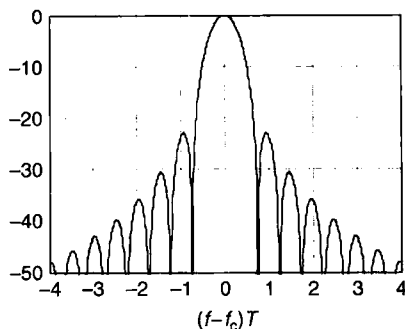


图 11.35 MSK 的功率谱密度曲线

11.3.8 最小频移键控的解调

对于最小频移键控的不同表示方式不仅有利于深入了解这种调制方案, 而且也有利于构造解调器。不同的解调器和不同的表示方式相对应。

- 鉴频检测。由于最小频移键控是一种频移键控, 这样就能直接通过检测瞬时频率是否大于或者小于载波频率 (在等效的基带则是和 0 比较) 来解调数据。瞬时频率可在比特的中间采样, 或者在 (部分) 比特周期内进行积分, 以减少噪声的影响。这种接收机结构都很简单, 但不是最佳的, 这是因为它没有利用到传输比特之间的连续相位特性。
- 差分检测。在一个比特周期内信号 $+\pi/2$ 或 $-\pi/2$ 的相位改变取决于所传输的比特。为

了判决,接收端只需判断信号在 iT 和 $(i+1)T$ 时刻的相位。显然不需要对传输的信号进行差分编码,且在一个采样时刻对相位的错判会造成两个(但不会更多)比特出错。

- 匹配滤波器接收。众所周知,匹配滤波器接收是最佳的(见第12章)。这对于最小频移键控和偏移正交相移键控也一样,此时将其看成多脉冲调制。然而,必须注意最小频移键控是一种具有记忆特性的调制,所以逐比特检测也是次最佳的:信号空间图中存在4个可能的星座点(在 0° , 90° , 180° 和 270°),对于逐比特检测,判决界就是第一个和第二个主对角线。信号星座位置和判决界之间的距离为 $\sqrt{E}/\sqrt{2}$, 这 and 二进制相移键控相比差了 3 dB。但是,这种判决方法没有利用从系统记忆得到的信息:如果前一个星座点已经在 0° , 后续的信号星座点只能在 90° 或者 270° 。判决界也就是 x 轴,星座到判决界的距离就是 $\sqrt{E}$, 这 and 二进制相移键控相同。可以利用记忆信息,即可以利用最大似然序列估计进行检测(见 14.3 节)①。

11.3.9 高斯最小频移键控

高斯最小频移键控(GMSK)是调制指数 $h_{\text{mod}}=0.5$ 的连续相位频移键控,其高斯相位基带脉冲为

$$\tilde{q}(t) = g_G(t, T_B, B_G T) \quad (11.69)$$

因此传输的相位脉冲序列(见图 11.36)为

$$p_D(t) = \sum_{i=-\infty}^{\infty} b_i \tilde{q}(t - iT_B) = b_i * \tilde{q}(t) \quad (11.70)$$

图 11.37 给出了其频谱图。可以看到,高斯最小频移键控可获得比最小频移键控更好的谱特性,因其采用了比最小频移键控更平滑的高斯相位基本脉冲。

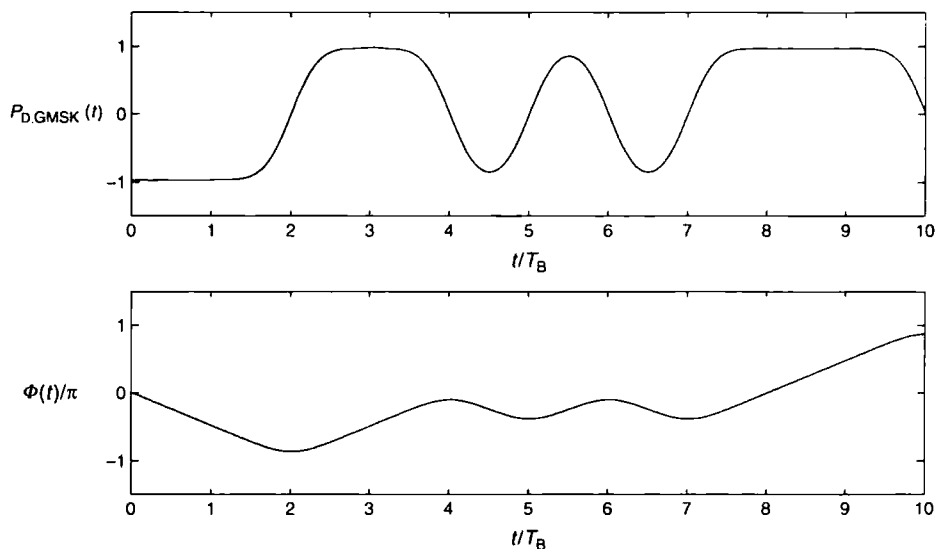


图 11.36 高斯最小频移键控信号的脉冲序列和相位

① 对于最大似然序列检测,只有 3 个采样值能影响到对特定比特的判决[Benedetto and Biglieri 1999]。

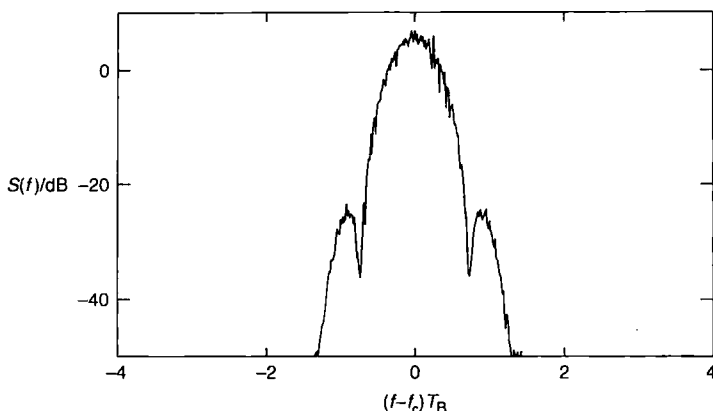


图 11.37 高斯最小频移键控功率谱密度(基于仿真)

高斯最小频移键控是欧洲应用最广泛的调制方式,并被作为全球蜂窝移动通信系统($B_c T = 0.3$)和无绳电话 DECT($B_c T = 0.5$)的标准(分别见第 24 章和关于 DECT 的附录)。高斯最小频移键控也用于无线个人局域网的蓝牙(IEEE 802.15.1)标准^①。

值得注意的是,高斯最小频移键控不能表示为脉冲幅度调制形式,但是 Laurent[1986]推导出了将高斯最小频移键控表示为具有有限记忆的脉冲幅度调制形式的公式。

11.3.10 脉冲位置调制

脉冲位置调制(PPM)是另一种多脉冲调制方式。前面的频移键控使用具有不同中心频率的基带脉冲,而脉冲位置调制则使用具有不同时间延的基带脉冲。下面将考虑 M 进制脉冲位置调制:

$$s_{LP}(t) = \sqrt{\frac{2E_B}{T_B}} \sum_{i=-\infty}^{\infty} g_{c_i}(t - iT) \quad (11.71)$$

$$= \sqrt{\frac{2E_B}{T_B}} \sum_i g_{PPM}(t - iT - c_i T_d) \quad (11.72)$$

其中, T_d 为调制时延(见图 11.38)。由于待调制的码元都是直接映射到脉冲时延的,所以其已调信号是实信号。

脉冲位置调制的包络具有很大的波动,但可以用非线性放大器,这是因为其输出只允许有两个幅度电平。

由于脉冲位置调制是一种非线性调制,所以其频谱不能由式(11.20)计算出,而是由复杂的多脉冲调制频谱表达式给出[见式(11.52)]:

$$S(f) = \frac{1}{M^2 T_S^2} \sum_{i=-\infty}^{+\infty} \left(\left| \sum_{m=0}^{M-1} G_m(f) \right|^2 \delta\left(f - \frac{i}{T_S}\right) \right) + \frac{1}{T_S} \left(\sum_{m=0}^{M-1} \frac{1}{M} |G_m(f)|^2 - \left| \sum_{m=0}^{M-1} \frac{1}{M} G_m(f) \right|^2 \right) \quad (11.73)$$

注意其频谱存在一些离散的谱线。

① 严格来讲, DECT 和蓝牙使用调制指数为 0.5 的高斯频移键控,这可以等效为高斯最小频移键控,由调制指数引起的偏差是可容忍的。

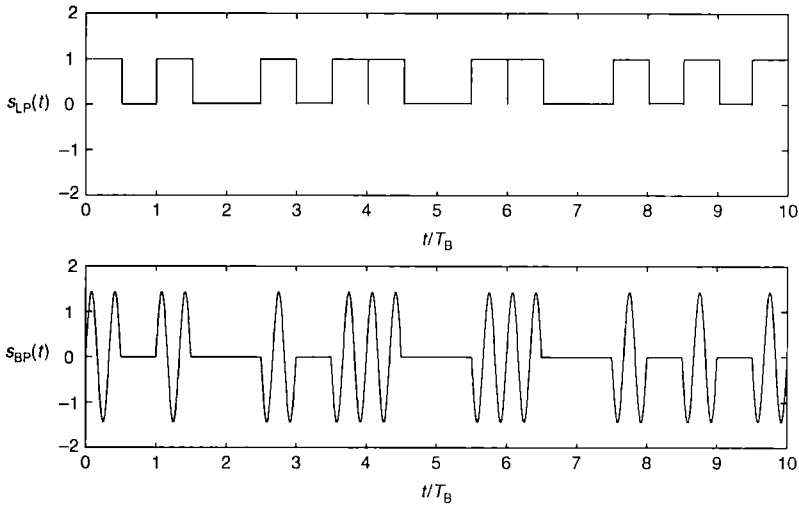


图 11.38 脉冲位置调制在低通和带通域的瞬时信号波形

到目前为止，我们一直假设传输脉冲为矩形脉冲，至于相互之间的偏移则仅限于一个脉冲宽度。但因为矩形脉冲频谱比较宽，所以这样并不能获得较高的频谱效率。也可以使用其他波形的脉冲，而这些不同脉冲的性能取决于检测的方法。当使用相干检测时，性能好坏的关键因素取决于代表 +1 和 -1 的脉冲波形之间的相关性(假设为二进制脉冲位置调制)：

$$\rho = \frac{\int g(t)g^*(t - T_d) dt}{\int |g(t)|^2 dt} \tag{11.74}$$

若 $\rho = 0$ ，则类似于使用矩形脉冲，当 $T_d > T_b$ 时调制是正交的。实际上应尽可能选择使 $\rho < 0$ 的脉冲波形，这不仅能获得较好的频谱效率，而且能获得较好的性能。然而当接收端使用非相干(能量)检测，脉冲之间没有重叠时将获得最好的性能。此时的相关系数为

$$\rho_{env} = \frac{\int |g(t)||g(t - T_d)| dt}{\int |g(t)|^2 dt} \tag{11.75}$$

脉冲位置调制可以和其他调制混合使用，例如二进制脉冲位置调制可以和二进制相移键控组合，这样每个码元表示 2 比特(1 比特由传输脉冲的相位确定，另 1 比特则由位置确定)。

例 11.4 对于一个 $g(t) = \sin(t/T)/t$ 的脉冲位置调制系统，求当 T_d 为 (i) T ，(ii) $5T$ 时的相关系数和包络的相关系数。

解：

相关系数的表达式为

$$\rho = \frac{\int g(t)g(t - T_d) dt}{\int |g(t)|^2 dt} \tag{11.76}$$

假设 $T = 1$ ，则上式的分母就是

$$\int |g(t)|^2 dt = \int_{-\infty}^{\infty} \frac{\sin^2(t)}{t^2} dt = \pi \tag{11.77}$$

对于 $T_d = 1$ ，计算数值积分，分子为

$$\int g(t)g(t - T_d) dt = \int_{-\infty}^{\infty} \frac{\sin(t)}{t} \frac{\sin(t - 1)}{t - 1} dt \approx 2.63 \tag{11.78}$$

则相关系数就是 $\rho \approx 0.84$ 。对于 $T_d = 5$, 计算积分, 分子为

$$\int_{-\infty}^{\infty} g(t)g(t - T_d) dt = \int_{-\infty}^{\infty} \frac{\sin(t)}{t} \frac{\sin(t - 5)}{t - 5} dt \approx -0.61 \quad (11.79)$$

则相关系数就是 $\rho \approx -0.2$ 。对于 $T_d = 1$, 包络相关系数的分子为

$$\rho \int_{-\infty}^{\infty} \left| \frac{\sin(t)}{t} \right| \left| \frac{\sin(t-1)}{t-1} \right| dt \approx 2.73 \quad (11.80)$$

则包络相关系数就是 $\rho \approx 0.87$ 。对于 $T_d = 5$, 包络相关系数的分子为

$$\rho \int_{-\infty}^{\infty} \left| \frac{\sin(t)}{t} \right| \left| \frac{\sin(t-5)}{t-5} \right| dt \approx 1.04 \quad (11.81)$$

则包络相关系数就是 $\rho \approx 0.33$ 。

脉冲位置调制在无线通信系统中并不常用, 这是因为其相对较低的频谱效率和时延扩展对系统的影响。但脉冲位置调制仍然在一些新兴技术中得到了应用, 特别是在脉冲无线电中 (见 18.5 节)。

11.3.11 频谱效率小结

表 11.1 总结了不同种调制方法的谱效率, 假定占用带宽分别定义为包含信号总能量的 90% 和 99%。

表 11.1 不同调制的频谱效率和包络变化

调制方式	90% 能量带宽对应的 频谱效率 (bps/Hz)	99% 能量带宽对应的 频谱效率 (bps/Hz)
BPSK	0.59	0.05
BAM ($\alpha = 0.5$)	1.02	0.79
QPSK, OQPSK	1.18	0.10
MSK	1.29	0.85
GMSK ($B_c = 0.5$)	1.45	0.97
QAM ($\alpha = 0.5$)	2.04	1.58

11.4 附录

请访问 www.wiley.com/go/molisch。

深入阅读

不同调制的描述和信号空间图中的表示, 在许多关于数字通信的教科书中都可以找到, 如 Andersen [2005], Barry et al. [2003], Proakis [2005], Sklar [2001], Wilson [1996] 和 Xiong [2006]。专门介绍无线通信中的调制的书是 Burr [2001]。Hanzo et al. [2000] 中深入介绍了正交幅度调制。最小频移键控的多种不同表示形式和对应的解调器结构在 Benedetto and Biglieri [1999] 中可以找到。Murota and Hirade 发明了高斯最小频移键控。Anderson et al. [1986] 对连续相位调制的不同表示形式给出了权威性的表述。

第 12 章 解 调

本章将介绍接收信号的解调方式,并讨论不同解调方案在加性高斯白噪声(AWGN)信道、平坦衰落信道和色散信道中的性能。

12.1 加性高斯白噪声信道中的解调器结构和差错率

本节主要介绍与第 11 章的调制方式相对应的解调器基本结构,以及在加性高斯白噪声信道中误比特率(BER)和符号差错率(SER)的计算。与前面的章节一样,我们侧重于重要结论的介绍,更多详细的推导可参考 Anderson[2005], Benedetto and Biglieri[1999], Proakis[2005]和 Sklar[2001]等文献。

12.1.1 信道模型和噪声

加性高斯白噪声信道使传输信号产生衰减,使信号的相位产生旋转,并会叠加高斯噪声。衰减和相位旋转都是时不变的,所以容易计算。因此,接收信号(复基带形式)为

$$r_{LP}(t) = \alpha s_{LP}(t) + n_{LP}(t) \quad (12.1)$$

其中, α 为(复)衰减系数 $|\alpha| \exp(j\varphi)$, $n(t)$ 是(复)高斯噪声过程。

为了推导噪声的性质,首先考虑带通系统中的噪声。假设在有用带宽中噪声功率谱密度是恒定的,噪声的双边功率谱密度为 $N_0/2$,如图 12.1(a)所示。(复)等效低通噪声功率谱密度[见图 12.1(b)]为

$$S_{n,LP}(f) = \begin{cases} N_0, & |f| \leq B/2 \\ 0, & \text{其他} \end{cases} \quad (12.2)$$

注意, $S_{n,LP}(f)$ 关于 f 是对称的,即 $S_{n,LP}(f) = S_{n,LP}(-f)$ 。

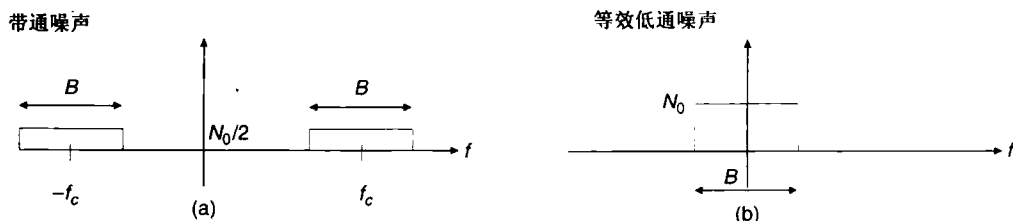


图 12.1 带通噪声及等效低通噪声功率谱

在时域中,噪声用其自相关函数 $R_{LP,nn}(\tau) = (1/2) E \{ n_{LP}^*(t) n_{LP}(t+\tau) \}$ 来描述。对于频带限制在 $[-B/2, B/2]$ 的系统,噪声的自相关函数为

$$R_{LP,nn}(\tau) = N_0 \frac{\sin(\pi B \tau)}{\pi \tau} \quad (12.3)$$

① 自相关函数的定义与第 6 章提到的自相关函数相差一个因子 2。这里用这种修正后的定义,是因为在通信理论中计算误比特率时通常这样使用。

当带宽 B 趋于无穷大时, 自相关函数变为

$$R_{LP,nn}(\tau) = N_0\delta(\tau) \quad (12.4)$$

还应注意, 当噪声的同相分量和正交分量的互相关为零时, 其同相分量和正交分量的相关函数都可以表示为 $R_{LP,nn}(\tau)$ 。

带通信号的自相关函数为

$$R_{BP,nn}(\tau) = \text{Re}\{R_{LP,nn}(\tau)\exp(j2\pi f_c\tau)\} \quad (12.5)$$

12.1.2 信号空间图及最佳接收机

下面推导数字调制的最佳接收机结构。在推导过程中, 还会发现使用信号空间图的另一个目的。推导中, 做如下假设:

1. 所有符号都是等概率发送的;
2. 调制方式是无记忆的;
3. 信道是高斯白噪声信道, 衰减和相位旋转都已知。为了不失一般性, 假设相位旋转已被完全补偿, 这样信道的衰减系数就是实数。

理想判决器称为最大后验概率(MAP)判决器。它可用来解决如下问题: 如果接收到的信号为 $r(t)$, 则最有可能发送的是哪一个信号 $s_m(t)$? 换句话说, 也就是哪个信号 $s_m(t)$ 使下面的概率取最大值:

$$\Pr[s_m(t)|r(t)] \quad (12.6)$$

可以用贝叶斯准则将此表示为(即找到满足下面条件的信号 m):

$$\max_m \Pr[n(t) = r(t) - \alpha s_m(t)] \Pr[s_m(t)] \quad (12.7)$$

由于已经假定所有信号都是等概率发送的, 所以最大后验概率判决实际上变成了最大似然判决:

$$\max_m \Pr[n(t) = r(t) - \alpha s_m(t)] \quad (12.8)$$

第 11 章介绍了表示已调发送信号的信号空间图, 其中仅把信号空间图作为一种较方便的表示方法, 本章中将给出收发信号之间的关联, 以及如何利用信号空间图来计算差错率。

发射信号可以表示为下面的形式^①:

$$s_m(t) = \sum_{n=1}^N s_{m,n} \varphi_n(t) \quad (12.9)$$

其中,

$$s_{m,n} = \int_0^{T_s} s_m(t) \varphi_n^*(t) dt \quad (12.10)$$

此处可发现接收信号能用类似的展开式表示。采用相同的基函数 $\varphi_n(t)$, 可得

$$r(t) = \sum_{n=1}^{\infty} r_n \varphi_n(t) \quad (12.11)$$

其中,

^① 注意, 这些式子对于基带和带通信号的表示形式都有效, 仅需插入正确的展开函数。

$$r_n = \int_0^{T_S} r(t) \varphi_n^*(t) dt \quad (12.12)$$

由于收到的信号中包含噪声,所以初看起来在展开式中需要更多的项,或无穷多项,以使展开式更精确。然而,可以发现把展开式分成两部分很有用:

$$r(t) = \sum_{n=1}^N r_n \varphi_n(t) + \sum_{n=N+1}^{\infty} r_n \varphi_n(t) \quad (12.13)$$

类似地有

$$n(t) = \sum_{n=1}^N n_n \varphi_n(t) + \sum_{n=N+1}^{\infty} n_n \varphi_n(t) \quad (12.14)$$

利用这些展开式,使最大似然接收机最大化的表达式可以写成

$$\max_m \Pr[\mathbf{n} = \mathbf{r} - \alpha \mathbf{s}_m] \quad (12.15)$$

其中,接收信号向量为 $\mathbf{r} = (r_1, r_2, \dots)^T$, $\mathbf{n}$ 可采用类似的表示形式。由于噪声分量是独立的,若发射信号为 $\mathbf{s}_m$,则接收信号向量 $\mathbf{r}$ 的概率密度为

$$p(\mathbf{r}_{\text{LP}} | \alpha \mathbf{s}_{\text{LP},m}) \propto \exp \left\{ -\frac{1}{2N_0} \|\mathbf{r} - \alpha \mathbf{s}_m\|^2 \right\} \quad (12.16)$$

$$= \prod_{n=1}^{\infty} \exp \left\{ -\frac{1}{2N_0} (r_n - \alpha s_{m,n})^2 \right\} \quad (12.17)$$

仅当 $n \leq N$ 时, $s_{m,n}$ 是非零的,最大似然检测器就是寻找:

$$\max_m \prod_{n=1}^N \exp \left\{ -\frac{1}{2N_0} (r_n - \alpha s_{m,n})^2 \right\} \prod_{n=N+1}^{\infty} \exp \left\{ -\frac{1}{2N_0} (r_n)^2 \right\} \quad (12.18)$$

现在的关键是要认识到式(12.18)中的第二项与 s_m 无关,所以该项不影响判决结果。这是不在发射信号空间中的噪声分量(接收信号中的)与检测器判决无关的另一种说法(Wozencraft的不相关理论)。最后,由于 $\exp(\cdot)$ 是单调函数,可以发现使距离最小化是充分的:

$$\mu(\mathbf{s}_{\text{LP},m}) = \|\mathbf{r}_{\text{LP}} - \alpha \mathbf{s}_{\text{LP},m}\|^2 \quad (12.19)$$

在几何上,这表示最大似然接收机把与接收向量 $\mathbf{r}_{\text{LP}}$ 具有最小欧氏距离的发射向量 $\mathbf{s}_{\text{LP},m}$ 判为符号 m 。需要记住,这仅是无记忆未编码系统的最佳判决方法。向量 $\mathbf{r}$ 包含的“软”信息,也就是接收机对它的判决有多大把握。这些软信息,在寻找最近相邻点的判决过程中丢失了。当一个比特不能表达关于其他比特的任何信息时,这些软信息是无用的。然而,对于编码系统和有记忆的系统(见第14章和第16章),这些软信息将是非常有用的。

距离 $\mu(\mathbf{s}_{\text{LP},m})$ 可以重写成

$$\mu(\mathbf{s}_{\text{LP},m}) = \|\mathbf{r}_{\text{LP}}\|^2 + \|\alpha \mathbf{s}_{\text{LP},m}\|^2 - 2\alpha \text{Re}\{\mathbf{r}_{\text{LP}} \mathbf{s}_{\text{LP},m}^*\} \quad (12.20)$$

由于 $\|\mathbf{r}_{\text{LP}}\|^2$ 项与所考察的 $\mathbf{s}_{\text{LP},m}$ 无关,距离函数 $\mu(\mathbf{s}_{\text{LP},m})$ 的最小化等价于下式的最大化:

$$\text{Re}\{\mathbf{r}_{\text{LP}} \mathbf{s}_{\text{LP},m}^*\} - \alpha E_m \quad (12.21)$$

记住, $E_m = \|\mathbf{s}_{\text{LP},m}\|^2/2$ (见第11章)。

判决准则的一个重要条件是接收机需要知道信道的衰减因子 α 。对于无线系统,尤其是信道特性快速变化时,这将是非常困难的(见本书第二部分)。因此最好采用不需要该信息的调

制与解调方法。特别是,当所有信号等能量传输,即 $E_m = E$ 时,不必知道信道的衰减值。如果采用非相干解调或差分解调,就可以忽略信道的相位旋转(α 的角度)。

上述推导的好处是它与实际的调制方案无关。在信号空间图中表示的发射信号,给出了所有的相关信息。图 12.2 所示的接收机结构,对于在信号空间图中表示的任意调制符号,都可实现最佳接收。唯一的前提就是必须满足本章前面提到的条件。这种接收机可以用相关器来实现,或者用与可能的不同发射波形相匹配的匹配滤波器来实现。

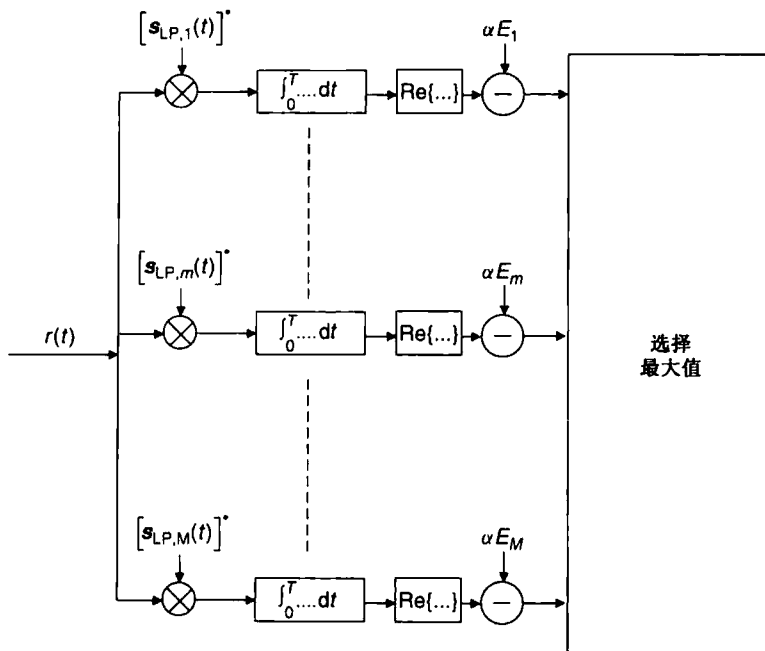


图 12.2 最佳接收机的一般结构

接下来将会发现,在信号空间图中两点之间的欧氏距离是调制方案的重要参数。第 11 章假设等能量传输信号的带通表达式中,欧氏距离与信号分量的能量关系为

$$d_{12}^2 = 2E(1 - \text{Re}\{\rho_{12}\}) \quad (12.22)$$

式中, ρ_k 为相关系数,定义与第 11 章中的相同:

$$\text{Re}\{\rho_{k,m}\} = \frac{s_k s_m}{|s_k| |s_m|} \quad (12.23)$$

12.1.3 差错概率的计算方法

这一节讨论如何计算最佳接收机所能获得的性能。在详细分析之前,先定义几个重要的参数:误比特率(bit error rate),顾名思义是个速率,它描述的是单位时间内的错误比特数,量纲是 s^{-1} ;误比特比率(bit error ratio)是用错误的比特数除以传输的比特数,因此它是无量纲的。当传输无穷多的比特时,就变成了误比特概率。在这本书中,这三种表达趋向于混合表示。特别是很多书的作者常用误比特率表示误比特概率,这种混用非常普遍,本书也将如此。

上一章已经介绍了大量的调制方式,在信号空间图的架构下,它们可以简单地分为如下几类。

- 1. 二进制相移键控(BPSK)是双极性信号。
- 2. 二进制频移键控(BFSK)和二进制脉冲位置调制(BPPM)是正交信号。
- 3. 正交相移键控(QPSK), 差分正交相移键控($\pi/4$ -DQPSK)和偏移正交相移键控(OQPSK)是双正交信号。

相干接收机的差错概率(一般情况)

如前面所讲, 相干接收机通过载波恢复来补偿信道的相位旋转。此外, 假设信道增益 α 是已知的, 并且已经作用于接收信号, 所以在无噪声的情况下, $\mathbf{r} = \mathbf{s}_k$ 。与符号 $\mathbf{s}_k$ 的欧氏距离为 d_{jk} 的信号 $\mathbf{s}_j$ 误判成 $\mathbf{s}_k$ 的概率(成对出错概率)为

$$\Pr_{\text{pair}}(\mathbf{s}_j, \mathbf{s}_k) = Q\left(\sqrt{\frac{d_{jk}^2}{2N_0}}\right) = Q\left(\sqrt{\frac{E}{N_0}(1 - \text{Re}\{\rho_{jk}\})}\right) \tag{12.24}$$

其中, Q 函数定义为

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2) dt \tag{12.25}$$

它与互补误差函数的关系为

$$Q(x) = \frac{1}{2} \text{erfc}\left(\frac{x}{\sqrt{2}}\right) \tag{12.26}$$

和

$$\text{erfc}(x) = \frac{1}{2} \exp(-\gamma_b/2) \tag{12.27}$$

式(12.24)用来计算发送信号为 $\mathbf{s}_j$, 而噪声大到使接收信号在信号空间图中的几何位置接近点 $\mathbf{s}_k$ 时的概率。

相干接收机的差错概率(二进制正交信号)

由第 11 章可知, 许多重要的调制方式都可以看成二进制正交信号。最明显的是二进制频移键控和二进制脉冲位置调制。图 12.3 给出的就是这种情况下的信号空间图。该图也给出了判决界: 如果接收信号点落入阴影区域, 发送符号就判为 +1, 否则判为 -1。

符号的信噪比定义为 $\gamma_s = E_s/N_0$, 可得

$$\Pr_{\text{pair}}(\mathbf{s}_j, \mathbf{s}_k) = Q\left(\sqrt{\gamma_s(1 - \text{Re}\{\rho_{jk}\})}\right) \tag{12.28}$$

$$= Q(\sqrt{\gamma_s}) \tag{12.29}$$

注意, 对于二进制信号, 满足 $\gamma_s = \gamma_b$ 。

相干接收机的差错概率(双极信号)

对于双极信号, 成对出错概率为

$$\Pr_{\text{pair}}(\mathbf{s}_j, \mathbf{s}_k) = Q\left(\sqrt{\gamma_s(1 - \text{Re}\{\rho_{jk}\})}\right) \tag{12.30}$$

$$= Q(\sqrt{2\gamma_s}) \tag{12.31}$$

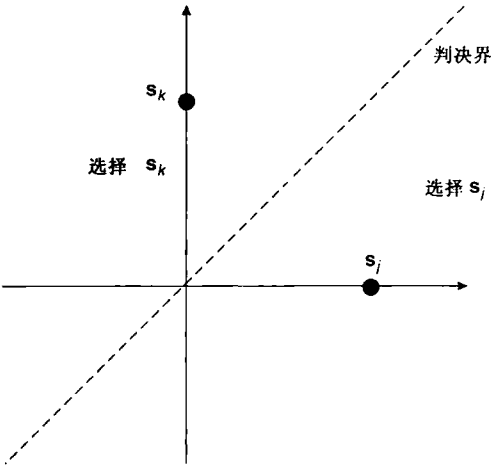


图 12.3 s_j 和 s_k 之间判决界的选取

对等概率发送的二进制信号,成对出错概率等于符号差错率,也等于误比特率。例如二进制相移键控和具有理想相干检测的最小频移键控(见第 11 章)就是这种情况,误比特率由式(12.30)给出。注意,最小频移键控可与频移键控采用类似的检测方法,但它没有利用相位的连续性。在此情况下,最小频移键控可视为正交调制形式,误比特率由式(12.29)给出,这意味着使有效信噪比下降了 3 dB。

联合限和双正交信号

对于 M 进制调制方法,误比特率的精确计算困难得多,所以其误比特率经常是使用联合限方法得到的上界。图 12.4 给出了这种定界方法的原理。在信号空间图中,引起错误判决的区域由局部区域组成,每一个区域表示一对错误,即把正确符号混淆成另一个符号。符号差错率可记为这些成对出错概率之和。因为成对错误区域有重叠,所以它表示的是实际符号错误概率的上边界。尽管重叠区域有一些小的影响,但由于符号差错率主要是由接近判决界的区域决定的,所以随着信噪比的提高,近似值也会得到改善。

使用高阶调制形式的误比特率公式时要多加注意。可能存在以下几点易犯的错误。

- 公式给出的是符号差错率还是误比特率? 例如图 12.4 的 4-QAM,若星座图的编码采用的是格雷码,则 4 个点表示的是比特组合 00, 01, 11 和 10(顺时针方向)。两个相邻的信号星座点之间存在很高的出错率。一个符号错误(在发送符号中)对应于一个比特错误(两个传输比特之间的一个错误),因而误比特率仅为符号差错率的一半^①。

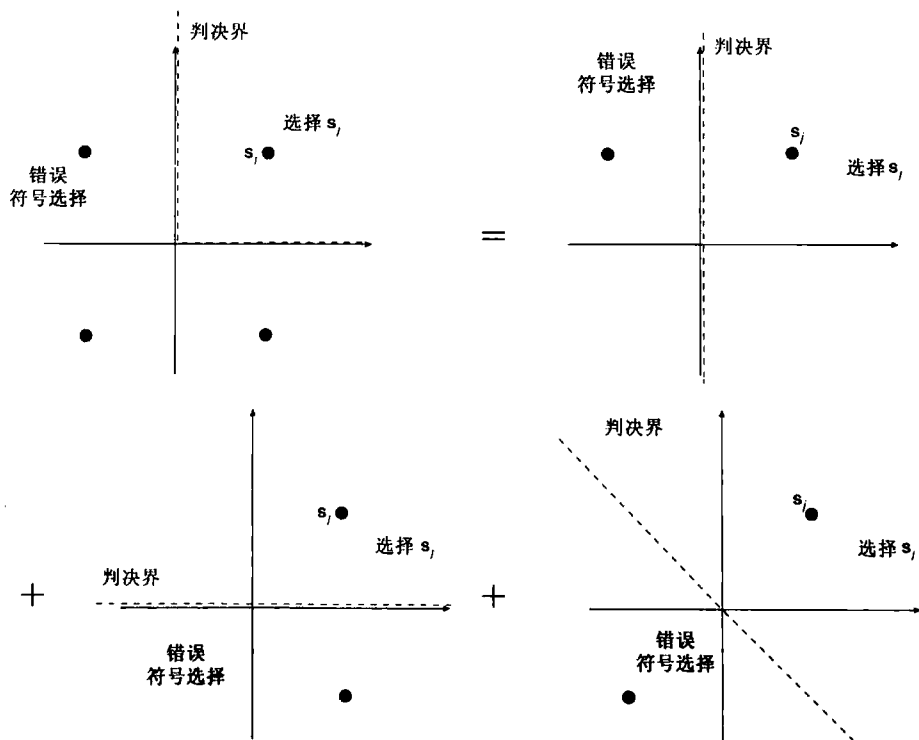


图 12.4 误码率的联合边界

^① 这表明了将“误比特率”和“误比特率”相混淆的缺点。在上面的例子中,误比特率与符号差错率相同,而误比特率仅为符号差错率的一半。

- 信噪比的计算利用的是比特能量还是符号能量? 对不同调制方式进行公平比较, 都应基于 E_B/N_0 。
- 一些书的作者定义原点与信号星座点(等能量)之间的距离不是 $\sqrt{E}$, 而是 $\sqrt{2E}$ 。这与扩展函数的不同归一化方式有关, 可通过不同数值对 n_n 进行补偿。虽然最终结果没有变化, 但这使得不同源的中间结果组合变得复杂得多。

作为例子, 接下来计算 4-QAM 的误比特率。由于它是一个 4 状态调制形式(每符号 2 比特), 所以每个符号需占用 2 比特的能量。信号点与原点之间的欧氏距离为 $d^2 = E_s = 2E_B$; 信号星座图中的点的坐标是 $\pm \sqrt{E_B}(\pm 1 \pm j)$; 两相邻点之间的距离是 $d_{jk}^2 = 4E_B$ 。

下面讨论两种类型的联合界。

1. 对于“全”联合界, 计算所有可能信号星座点的成对出错概率, 然后加起来。如图 12.4 所示。
2. 仅用相邻点计算成对出错概率。这样, 可以在计算中省略图 12.4 中最后的判决区域。可以看到, 前两个区域的联合(最近相邻点的成对出错概率)已经覆盖整个“错误判决区域”。下面将应用这种类型的联合界。

例 12.1 计算正交相移键控的误比特率和符号差错率。

解:

由式(12.24)得到成对出错概率为

$$Q\left(\sqrt{2\frac{E_B}{N_0}}\right) = Q(\sqrt{2\gamma_B}) \quad (12.32)$$

根据图 12.4, 由联合界计算的符号差错率是成对出错概率的 2 倍:

$$\text{SER} \approx 2Q(\sqrt{2\gamma_B}) \quad (12.33)$$

由上面所得, 误比特率是符号差错率的一半:

$$\text{BER} = Q(\sqrt{2\gamma_B}) \quad (12.34)$$

它与二进制相移键控的误比特率相等。

对于正交相移键控, 也可精确计算出其符号差错率: 正确判决的概率(见图 12.4)是位于右半平面(此事件的概率是 $1 - Q(\sqrt{2\gamma_B})$)的概率与位于上半平面(此事件的概率与处于右半平面的概率是独立的, 也是 $1 - Q(\sqrt{2\gamma_B})$)的概率的乘积。因而, 总的符号差错率为 $1 - (1 - Q(\sqrt{2\gamma_B}))^2$, 即

$$\text{SER} = 2Q(\sqrt{2\gamma_B}) \left[1 - \frac{1}{2}Q(\sqrt{2\gamma_B}) \right] \quad (12.35)$$

该式说明了由联合界引起的错误率的量级。相对符号差错率是 $0.5Q(\sqrt{2\gamma_B})$, 随着 γ_B 趋于无穷而趋于零。

高阶调制形式的符号差错率或者误比特率的精确计算是非常复杂的, 而联合界提供了简单且非常精确(尤其对高信噪比)的近似。

差分检测的差错概率

载波恢复是一个非常具有挑战性的问题[Meyr et al. 1997, Proakis 1995], 这使得在许多情

况下很难进行相干检测,因而差分检测是一种替代相干检测的非常有吸引力的选择,因其摒弃了接收信号的绝对相位。在这种方案中,接收机只需比较两个连续符号之间的相位(也可能是幅度)来恢复符号中包含的信息。相位差与绝对相位无关。如果由信道引入的相位旋转是缓慢时变的(因此对两个连续符号也是如此),那么它只改变绝对相位,所以在判决过程中不需要考虑。

对于差分检测的相移键控,在发送端需要进行差分编码。对于二进制对称相移键控,第 i 个比特的发送相位 Φ_i 是

$$\Phi_i = \Phi_{i-1} + \begin{cases} +\frac{\pi}{2}, & b_i = +1 \\ -\frac{\pi}{2}, & b_i = -1 \end{cases} \quad (12.36)$$

比较两个连续采样比特之间的相位差,从而决定传输的比特 b_i 是 $+1$, 还是 -1 ①。

对于连续相位频移键控(CPSK),可避免差分编码。在最小频移键控(无差分编码)中,1比特持续时间内的相位旋转是 $\pm\pi/2$ 。因此可以利用两个连续采样点的相位计算差值,来判决传输的是哪个比特。这也可以解释为发射信号相位是未编码比特序列的积分。计算相位差是求导的最好近似(如果相位是线性变化的,则是精确的),是积分的逆变换。

对于二进制正交信号,差分检测的误比特率为[Proakis 1995]:

$$\text{BER} = \frac{1}{2} \exp(-\gamma_b) \quad (12.37)$$

采用格雷编码的4-PSK,误比特率为

$$\text{BER} = Q_M(a, b) - \frac{1}{2} I_0(ab) \exp\left(-\frac{1}{2}(a^2 + b^2)\right) \quad (12.38)$$

其中,

$$a = \sqrt{2\gamma_b \left(1 - \frac{1}{\sqrt{2}}\right)}, \quad b = \sqrt{2\gamma_b \left(1 + \frac{1}{\sqrt{2}}\right)} \quad (12.39)$$

$Q_M(a, b)$ 是 Marcum Q 函数:

$$Q_M(a, b) = \int_b^\infty x \exp\left[-\frac{a^2 + x^2}{2}\right] I_0(ax) dx \quad (12.40)$$

其级数表示形式见式(5.31)。

非相干检测的差错概率

当载波相位完全未知时,差分检测不是一种好的选择,那么可以采用非相干检测。对于等能量信号;检测器就是尽量使下式最大:

$$|\mathbf{r}_{\text{LP}} \mathbf{s}_{\text{LP},m}^*| \quad (12.41)$$

最佳接收机的结构如图12.5所示。

在实际实现中,当 $r(t)$ 是带通信号时,图12.5中每一个支路将分成两个支路,对所获得的I和Q支路分开处理,对I和Q支路进行“绝对值”处理后输出至“最大值选择”模块。

在这种情况下,误比特率可用式(12.38)来计算,但是参数 a 和 b 的定义有所不同:

$$a = \sqrt{\frac{\gamma_b}{2} (1 - \sqrt{1 - |\rho|^2})}, \quad b = \sqrt{\frac{\gamma_b}{2} (1 + \sqrt{1 - |\rho|^2})} \quad (12.42)$$

① 从理论上讲,非编码信号也可采用差分检测,但是一个误码会引起一连串误码。

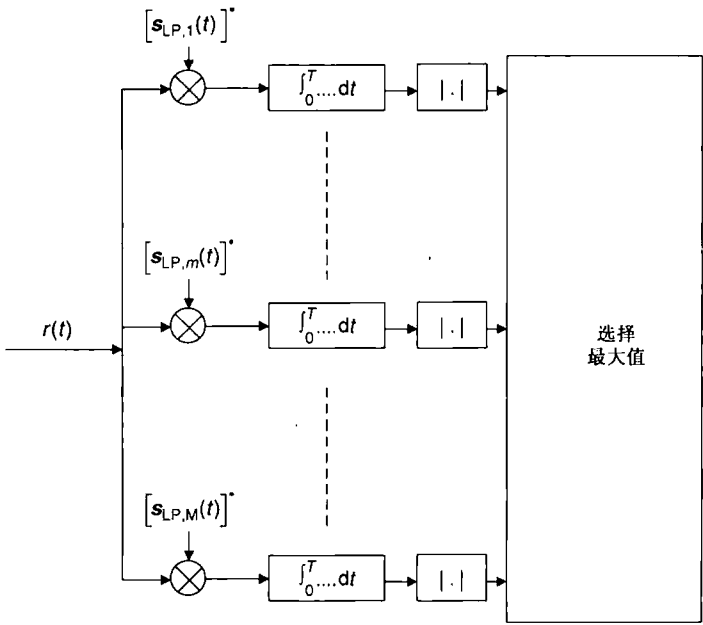


图 12.5 非相干检测的最佳接收机结构

当 $\rho=0$ ，即信号正交时，接收机可以达到最佳性能。当 $|\rho|=1$ 时，对于相移键控信号，包括二进制相移键控和 4-QAM 信号，会出现这种情况，此时误比特率为 0.5。

例 12.2 在 $\gamma_B=5\text{ dB}$ 的加性高斯白噪声信道中，计算二进制频移键控的误比特率，并与差分二进制相移键控和二进制相移键控进行比较。

解：

对于加性高斯白噪声信道中的二进制频移键控，误比特率由式(12.29)得到，且 $\gamma_S=\gamma_B$ 。因此，当 $\gamma_B=5\text{ dB}$ 时，有

$$\begin{aligned} \text{BER}_{\text{BFSK}} &= Q(\sqrt{\gamma_B}) \\ &= 0.038 \end{aligned} \tag{12.43}$$

对于二进制相移键控，其误比特率由式(12.34)决定，因此有

$$\begin{aligned} \text{BER}_{\text{BPSK}} &= Q(\sqrt{2\gamma_B}) \\ &= 0.006 \end{aligned} \tag{12.44}$$

最后，对差分二进制相移键控，由式(12.34)得到结果：

$$\begin{aligned} \text{BER}_{\text{DBPSK}} &= (1/2)e^{-\gamma_B} \\ &= 0.021 \end{aligned} \tag{12.45}$$

12.2 平坦衰落信道中的差错概率

12.2.1 平均误比特率(经典的计算方法)

在衰落信道中，接收信号的功率(和信噪比)不是常数，随着信道衰落的变化而变化。在许多情况下，我们关注的是在不同衰落状态下衰落信道的平均误比特率。在这种信道中，为了从数学上计算其误比特率，需要按照如下 3 个步骤进行。

1. 确定任意信噪比时的误比特率。
2. 确定在此信道中, 某一信噪比出现的概率, 换句话说, 就是确定信道中功率增益的概率密度函数。
3. 利用信噪比的概率分布, 求平均误比特率。

在加性高斯白噪声信道中, 误比特率随着信噪比的提高近似呈指数级迅速减小。对于二进制调制形式, 信噪比为 10 dB 时, 误比特率为 10^{-4} 量级, 信噪比为 15 dB 时, 误比特率在 10^{-8} 以下。相反, 下面可以看到在衰落信道中, 误比特率随 (平均) 信噪比的增加线性减小。初看起来, 这令人惊异: 衰落会使信噪比有时高有时低, 但平均数值不变, 它本该假定这些高低数值能相互补偿。重要的一点是 (瞬时) 误比特率和 (瞬时) 信噪比之间具有很强的非线性关系。因而低信噪比的情况从根本上决定整个系统的误比特率。

例 12.3 两状态衰落信道中的误比特率。

解:

考虑下面一个简单的例子: 一个衰落信道的平均信噪比为 10 dB, 其中衰落使信噪比在一半时间内为 $-\infty$, 而其余时间内为 13 dB。与这两种信道状态相对应的误比特率分别为 0.5 和 10^{-9} 。则平均误比特率为 $\overline{\text{BER}} = 0.5 \times 0.5 + 0.5 \times 10^{-9} = 0.25$ 。而在信噪比为 10 dB 的加性高斯白噪声信道中, 误比特率为 2×10^{-5} 。

在这些直观解释之后, 来研究上述三步中的数学细节。第一步, 确定任意给定信噪比的误比特率, 这在 12.1 节中已经给出。第二步, 要求计算信噪比的分布。第 5 章主要研究幅度 $r = |r(t)|$ 的分布, 现在要将其转换为接收功率 $P_{\text{inst}} = r^2$ 的分布。这种变换可采用雅可比行列式来实现。作为一个例子, 下面给出接收信号幅度为瑞利分布时变换的情况:

$$\text{pdf}_r(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r \geq 0 \quad (12.46)$$

平均功率为 $P_m = 2\sigma^2$ 。利用雅可比式 $|dP_{\text{inst}}/dr| = 2r$, 接收信号功率的概率密度函数为

$$\text{pdf}_{P_{\text{inst}}}(P_{\text{inst}}) = \frac{1}{P_m} \exp\left(-\frac{P_{\text{inst}}}{P_m}\right), \quad P_{\text{inst}} \geq 0 \quad (12.47)$$

由于信噪比是接收功率除以噪声功率, 因而信噪比的概率密度函数为

$$\text{pdf}_{\gamma_B}(\gamma_B) = \frac{1}{\overline{\gamma_B}} \exp\left(-\frac{\gamma_B}{\overline{\gamma_B}}\right) \quad (12.48)$$

其中, $\overline{\gamma_B}$ 是平均信噪比。对于莱斯衰落信道, 有类似的表示式, 但更复杂, 其信噪比概率密度函数为 [Rappaport 1996]

$$\text{pdf}_{\gamma_B}(\gamma_B) = \frac{1 + K_r}{\overline{\gamma_B}} \exp\left(-\frac{\gamma_B(1 + K_r) + K_r \overline{\gamma_B}}{\overline{\gamma_B}}\right) I_0\left(\sqrt{\frac{4(1 + K_r)K_r \gamma_B}{\overline{\gamma_B}}}\right) \quad (12.49)$$

其中, K_r 是莱斯因子。对于其他幅度分布, 可采用类似的计算。

最后一步, 基于信噪比概率分布的平均误比特率为

$$\overline{\text{BER}} = \int \text{pdf}_{\gamma_B}(\gamma_B) \text{BER}(\gamma_B) d\gamma_B \quad (12.50)$$

对于瑞利衰落, 对于多种调制方式可以得到式 (12.50) 的闭式表示形式。利用

$$2 \int_0^\infty Q(\sqrt{2x}) a \exp(-ax) dx = 1 - \sqrt{\frac{1}{1+a}} \quad (12.51)$$

二进制双极信号相干检测的平均误比特率为

$$\overline{\text{BER}} = \frac{1}{2} \left[1 - \sqrt{\frac{\gamma_B}{1 + \gamma_B}} \right] \approx \frac{1}{4\gamma_B} \quad (12.52)$$

而二进制正交信号相干检测的平均误比特率为

$$\overline{\text{BER}} = \frac{1}{2} \left[1 - \sqrt{\frac{\gamma_B}{2 + \gamma_B}} \right] \approx \frac{1}{2\gamma_B} \quad (12.53)$$

这些误比特率的曲线如图 12.6 所示。

对于差分检测, 由于其误比特率曲线是指
数函数, 所以求平均的过程更简单。对于二进
制双极信号为

$$\overline{\text{BER}} = \frac{1}{2(1 + \gamma_B)} \approx \frac{1}{2\gamma_B} \quad (12.54)$$

对于二进制正交信号为

$$\overline{\text{BER}} = \frac{1}{2 + \gamma_B} \approx \frac{1}{\gamma_B} \quad (12.55)$$

对于差分检测, 在莱斯信道中误比特率也可得
到闭合形式。双极信号为:

$$\overline{\text{BER}} = \frac{1 + K_r}{2(1 + K_r + \gamma_B)} \exp\left(-\frac{K_r \gamma_B}{1 + K_r + \gamma_B}\right) \quad (12.56)$$

正交信号为

$$\overline{\text{BER}} = \frac{1 + K_r}{(2 + 2K_r + \gamma_B)} \exp\left(-\frac{K_r \gamma_B}{2 + 2K_r + \gamma_B}\right) \quad (12.57)$$

例 12.4 计算 $\overline{\gamma_B} = 12 \text{ dB}$, $K_r = -3 \text{ dB}$, 0 dB , 10 dB 时, 差分二进制相移键控的误比特率。

解:

由误比特率公式(12.56), 当 $K_r = -3 \text{ dB}$ 时, 可得

$$\begin{aligned} \overline{\text{BER}} &= \frac{1 + K_r}{2(1 + K_r + \gamma_B)} \exp\left(-\frac{K_r \gamma_B}{1 + K_r + \gamma_B}\right) \\ &= 0.027 \end{aligned} \quad (12.58)$$

当 $K_r = 0 \text{ dB}$ 时, $\overline{\text{BER}} = 0.023$; 当 $K_r = 10 \text{ dB}$ 时, $\overline{\text{BER}} = 0.000 56$ 。注意, 当莱斯因子较大时, 即使保持信噪比为常数, 误比特率也会急剧下降。

12.2.2 计算平均差错概率的另一种方法

20 世纪 90 年代末期, 提出了一种新的计算平均误比特率的方法, 并已被证明非常有效。该方法基于 Q 函数的另一种表示形式, 允许对不同的衰落分布进行更简单的平均[Annamalai et al. 2000, Simon and Alouini 2004]。

Q 函数的另一种表示形式

Q 函数的经典定义为

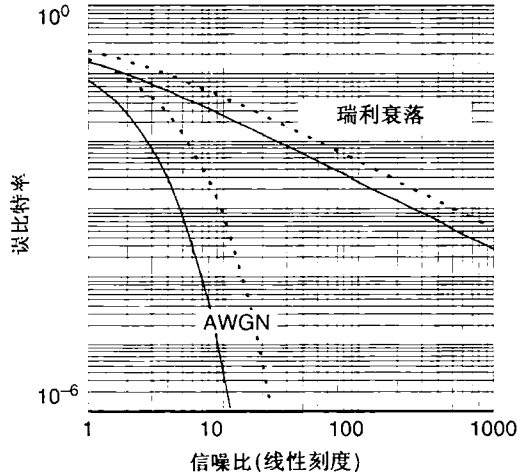


图 12.6 二进制信号相干检测的误比特率: 双极 (实线) 和正交 (虚线) 信号的情况

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-t^2/2) dt \quad (12.59)$$

这个定义的问题在于 Q 函数的变量在积分限上, 而不是在被积函数中, 从而使 Q 函数积分的计算非常困难, 尤其是不能用多重积分中的积分次序交换的策略来求解。这个问题与误比特率的计算密切相关, 可利用 Q 函数的另一种形式来解决:

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta, \quad x > 0 \quad (12.60)$$

现在, 在这个表达式中, 自变量处于被积函数中, 即高斯形式 $\exp(-x^2)$, 并且积分区间是有限的。下面将会看到, 这极大地简化了差错概率的计算。

由式(12.40)定义的 Marcum 的 Q 函数也可以表示为

$$Q_M(a, b) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{b^2 + ab \sin \theta}{b^2 + 2ab \sin \theta + a^2} \exp(-\frac{1}{2}(b^2 + 2ab \sin \theta + a^2)) d\theta & b > a \geq 0 \\ 1 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{b^2 + ab \sin \theta}{b^2 + 2ab \sin \theta + a^2} \exp(-\frac{1}{2}(b^2 + 2ab \sin \theta + a^2)) d\theta & a > b \geq 0 \end{cases} \quad (12.61)$$

加性高斯白噪声信道中的差错概率

为了计算加性高斯白噪声信道中的误比特率, 首先将二进制相移键控的误比特率写为[与式(12.30)比较]

$$\text{BER} = Q(\sqrt{2\gamma_s}) \quad (12.62)$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{\gamma_s}{\sin^2\theta}\right) d\theta \quad (12.63)$$

对于正交相移键控, 符号差错率的计算为(与式(12.35)比较)

$$\text{SER} = 2Q(\sqrt{\gamma_s}) - Q^2(\sqrt{\gamma_s}) \quad (12.64)$$

$$= \frac{1}{\pi} \int_0^{3\pi/4} \exp\left(-\frac{\gamma_s}{\sin^2\theta} \sin^2(\pi/4)\right) d\theta \quad (12.65)$$

更为一般的多进制相移键控为

$$\text{SER} = \frac{1}{\pi} \int_0^{(M-1)\pi/M} \exp\left(-\frac{\gamma_s}{\sin^2\theta} \sin^2(\pi/M)\right) d\theta \quad (12.66)$$

对于二进制正交频移键控, 可以表示为

$$\text{BER} = Q(\sqrt{\gamma_s}) \quad (12.67)$$

$$= \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{\gamma_s}{2\sin^2\theta}\right) d\theta \quad (12.68)$$

例 12.5 在 $\gamma_s = 3$ dB 和 10 dB 时, 比较分别由式(12.66)和联合界确定的 8-PSK 的平均符号差错率。

解:

首先, 由式(12.66)可知, 当 $M=8$, $\gamma_s = 3$ dB 时, 得到

$$\begin{aligned} \text{SER} &= \frac{1}{\pi} \int_0^{(M-1)\pi/M} \exp\left(-\frac{\gamma_s}{\sin^2\theta} \sin^2(\pi/M)\right) d\theta \\ &= 0.442 \end{aligned} \quad (12.69)$$

在 $\gamma_s = 10$ dB 时, $\text{SER} = 0.087$ 。式中积分采用数值积分计算。

8-PSK 星座图中的最小距离为 $d_{\min} = 2 \sqrt{E_s} \sin(\pi/8)$ 。与符号差错率的最近联合界为

$$\begin{aligned} \text{SER}_{\text{union-bound}} &= 2 \cdot Q \left(\frac{d_{\min}}{\sqrt{2N_0}} \right) \\ &= 2 \cdot Q \left(\frac{2\sqrt{E_s} \sin(\pi/8)}{\sqrt{2N_0}} \right) \\ &= 2 \cdot Q \left(\sqrt{2\gamma_s} \sin(\pi/8) \right) \end{aligned} \quad (12.70)$$

因此, 当 $\gamma_s = 3$ dB 时, 可得 $\text{SER}_{\text{union-bound}} = 0.445$, 而 $\gamma_s = 10$ dB 时, $\text{SER}_{\text{union-bound}} = 0.087$ 。在这个例子中, 即使信噪比低到 3 dB, 联合界也给出了很精确的近似。

衰落信道中的差错概率

对于加性高斯白噪声信道来说, Q 函数的另一种表示方法的优势是有限的。它虽然为高阶调制形式提供了更简单的公式, 但是对于实际应用中的各种不同调制, 却并没有绝对的优势。当这种表示形式作为衰落信道中误比特率计算的基础时, $\text{pdf}_\gamma(r)$ 才展现出其真正的优势。如式(12.50)中所述, 平均误比特率需要对信噪比的概率密度函数进行统计平均。现在可以发现, Q 函数的另一种表达式能把符号差错率(对某一给定的信噪比)写成一般形式:

$$\text{SER}(\gamma) = \int_{\theta_1}^{\theta_2} f_1(\theta) \exp(-\gamma f_2(\theta)) d\theta \quad (12.71)$$

因此, 平均符号差错率变为

$$\overline{\text{SER}} = \int_0^\infty \text{pdf}_\gamma(\gamma) \text{SER}(\gamma) d\gamma \quad (12.72)$$

$$= \int_0^\infty \text{pdf}_\gamma(\gamma) \int_{\theta_1}^{\theta_2} f_1(\theta) \exp(-\gamma f_2(\theta)) d\theta d\gamma \quad (12.73)$$

$$= \int_{\theta_1}^{\theta_2} f_1(\theta) \int_0^\infty \text{pdf}_\gamma(\gamma) \exp(-\gamma f_2(\theta)) d\gamma d\theta \quad (12.74)$$

现进一步研究内部积分:

$$\int_0^\infty \text{pdf}_\gamma(\gamma) \exp(-\gamma f_2(\theta)) d\gamma \quad (12.75)$$

可以发现这个积分是 $\text{pdf}_\gamma(\gamma)$ 在点 $-f_2(\theta)$ 的矩生成函数值。矩生成函数定义为 γ 的概率密度函数的拉普拉斯变换(参考 Papoulis[1991]):

$$M_\gamma(s) = \int_0^\infty \text{pdf}_\gamma(\gamma) \exp(\gamma s) d\gamma \quad (12.76)$$

平均信噪比是其在 $s=0$ 处的一阶导数:

$$\bar{\gamma} = \left. \frac{dM_\gamma(s)}{ds} \right|_{s=0} \quad (12.77)$$

总之, 平均符号差错率可以计算如下:

$$\overline{\text{SER}} = \int_{\theta_1}^{\theta_2} f_1(\theta) M_\gamma(-f_2(\theta)) d\theta \quad (12.78)$$

下一步要找出信噪比分布的矩生成函数。不去细究推导过程,可发现当信号幅度服从瑞利分布时,其信噪比分布的矩生成函数为

$$M_Y(s) = \frac{1}{1 - s\bar{\gamma}} \quad (12.79)$$

对于莱斯分布,信噪比分布的矩生成函数为

$$M_Y(s) = \frac{1 + K_r}{1 + K_r - s\bar{\gamma}} \exp\left[\frac{K_r s\bar{\gamma}}{1 + K_r - s\bar{\gamma}}\right] \quad (12.80)$$

对于参数为 m 的 Nakagami 分布,则为

$$M_Y(s) = \left(1 - \frac{s\bar{\gamma}}{m}\right)^{-m} \quad (12.81)$$

已知符号差错率[见式(12.78)]和矩生成函数的一般形式,差错概率的计算就变得简单明了(即使有时有一点冗长)。

例 12.6 计算瑞利衰落信道中二进制相移键控的误比特率。

解:

作为一个例子,下面计算瑞利衰落信道中二进制相移键控的平均误比特率,这是个由式(12.52)已经可以得到结果的问题。根据式(12.63),可以发现 $\theta_1 = 0$, $\theta_2 = \pi/2$:

$$f_1(\theta) = \frac{1}{\pi} \quad (12.82)$$

$$f_2(\theta) = \frac{1}{\sin^2(\theta)} \quad (12.83)$$

考虑瑞利衰落:

$$M_Y(-f_2(\theta)) = \frac{1}{1 + \frac{\bar{\gamma}}{\sin^2(\theta)}} \quad (12.84)$$

所以根据式(12.78),可以得到总的符号差错率为

$$\overline{\text{SER}} = \frac{1}{\pi} \int_0^{\pi/2} \frac{\sin^2(\theta)}{\sin^2(\theta) + \bar{\gamma}} d\theta \quad (12.85)$$

可以证明与式(12.52)中的第一个(精确)表示式一致,即

$$\overline{\text{BER}} = \frac{1}{2} \left[1 - \sqrt{\frac{\bar{\gamma}}{1 + \bar{\gamma}}} \right] \quad (12.86)$$

很多调制方式与衰落分布都可以用类似的方法处理。Simon and Alouini[2004]中列出了在不同类型的信道中,相干检测、非相干检测和部分相干检测的分析结果。

12.2.3 中断概率与平均差错概率

前一节介绍了平均误比特率的计算,而这种平均是对小尺度衰落的平均。类似地,也可以对这种分布在大尺度衰落下进行平均(所用到的数学方法类似)。但是,这些平均值的物理意义是什么呢?为了理解正在计算的量和平均的意义,首先要看一看无线通信系统中不同的工作环境。

第一步,比较系统的“记忆”长度和信道的相干时间。系统的“记忆”由如下原因引起:

1. 人类耳朵的感知(在几毫秒左右或更小数量级的突发错误,通过人类的耳朵被“平滑掉”);
2. 无线链路中传输的典型数据结构的尺寸;
3. 编码可增大记忆长度(分组码的分组长度或卷积码的约束长度决定记忆持续时间),并且交织的长度决定记忆长度(见第14章)。

现在先考虑接收机(或发射机)是移动的情况。信道会在一段有限的时间内发生变化,这样系统记忆就会延伸许多信道实现。这样一个系统记忆通常能经历信道的许多小尺度实现,但是这个时间仍太短了,使得在记忆持续时间内信道不会产生大尺度变化。想要了解文件传输的平均误比特率时,需要在文件传输时间内对误比特率求平均,因而利用该持续时间内的信噪比分布对其求平均^①。其误比特率是单一确定性的数值。

下面再考虑系统记忆时间远小于相干时间的情况。例如,从一个(静止的)微型计算机传输文件,并且环境中的物体都是不动的,那么在传输持续时间内,无线连接的信道可认为是静态的。因此,在解调器中的信噪比为常量,但随着位置变化而随机变化。为了了解在百分之多少的位置上可以得到某一信噪比,研究信噪比的概率密度函数是有意义的。对于每一个位置,即信噪比的每一实现,仅根据其特定的信噪比,采用加性高斯白噪声信道中的误比特率公式来计算其误比特率。以这样一种方式也可得到误比特率的分布,注意这种分布是在发射功率固定的情况下的分布。

由上面这些讨论可以得到中断概率的概念。对于很多实际应用,只要误比特率始终低于某个特定的门限值即可,而其准确值并不重要。例如,只要原始误比特率小到纠错编码可以纠正的范围,换句话说就是只要不超过原始误比特率的某一门限值(典型门限值在百分之几的数量级),则一个文件可以成功传输;因此确定无法成功传输文件的百分比是非常有意义的,这个百分比就是中断概率。

在一些情况,系统有足够长的记忆时间,可以观察到小尺度衰落多个实现,但只能观察到大尺度衰落一个实现。在这种情况下,当小尺度的平均误比特率大于某一门限值时,就产生了中断。这个中断产生的概率仅由大尺度衰落的统计特性决定。

为了使系统正常工作,如果定义的不是最大误比特率,而是最小信噪比 γ_0 , 则中断概率的计算将会简单得多。中断概率为

$$Pr_{out} = P(\gamma < \gamma_0) = \int_0^{\gamma_0} pdf_{\gamma}(\gamma) d\gamma \quad (12.87)$$

中断概率还可以看成确定衰落余量的另一种方法,但需要找到保证特定中断概率的平均信噪比。

12.3 时延及频率色散衰落信道中的差错概率

12.3.1 差错基底的物理原因

在无线传输信道中,传输错误不仅仅是由噪声引起的,而且还与信号失真有关。这些失真一方面由时延色散引起(也就是具有不同到达时延的发射信号回波),另一方面是由频率色散

^① 对于编码长度大于信道相干时间的编码系统,计算有一点棘手(如第14章所讨论的)。

引起的(多普勒效应,也就是具有不同多普勒频移的信号分量)。对于高数据速率,时延色散是信号失真的主要原因;而对于低数据速率,频率色散是主要原因。在任何一种情况下,增大发射信号功率都不会减小误比特率;因此,这种差错通常称为差错基底或不可减小的错误。当然,通过其他方法而不是增加功率(例如均衡,分集等),这些错误也可以被减小或消除。这一节仅研究接收机没有采取任何对策的情况,因而色散将会导致差错率的增加。后面的章节将详细介绍如果采用特殊的接收机结构,那么色散实际上会带来好处。

频率色散

下面首先考察由频率色散引起的错误。对于频移键控,由频率色散引起的错误是很明显的:随机频率调制(见 5.7.3 节)导致接收信号产生频移,可以将 1 比特移至判决门限值之外。假设发送的为 +1(也就是频率为 $f_c + f_{\text{mod}}$),由于随机频率调制效应,接收到的频率为 $f_c + f_{\text{mod}} + f_{\text{inst}}$ 。如果这个频率小于 f_c ,接收机就会将信号判决为 -1。注意,即使随机频率调制是由信道多普勒谱决定的,瞬时频率偏移也可能远大于最大多普勒频移。但是,考虑下面的瞬时频率公式:

$$f_{\text{inst}}(t) = \frac{\text{Im} \left(r^*(t) \frac{dr(t)}{dt} \right)}{|r(t)|^2} \quad (12.88)$$

显然,如果幅度变小,则瞬时频率将变得很大。换句话说,深度衰落导致大的瞬时频率的偏移,因而产生更高的差错概率。

对于差分检测,可以给出稍有不同的解释。前面提到过,差分检测假设信道在相邻信号之间无变化。然而,如果信道存在一个有限的多普勒频移,则信道就会发生变化。多普勒频谱是对信道变化的统计描述。因此,非零多普勒效应意味着差分检测中存在错误的参考相位。如果多普勒效应非常强烈,将会引起错误判决。在深衰落点^①附近通常会出现信道剧烈变化的情况。

对于差分检测的最小频移键控, Hirade et al. [1997] 给出了由于多普勒效应引起的误比特率:

$$\overline{\text{BER}}_{\text{Doppler}} = \frac{1}{2} (1 - \xi_s(T_B)) \quad (12.89)$$

其中, $\xi_s(t)$ 为信道的归一化自相关函数 [$\xi_s(0) = 1$], 也就是归一化多普勒频谱的傅里叶变换。对于小的 $v_{\text{max}} T_B$, 得到的误比特率和多普勒频移与比特周期之积的平方成正比:

$$\overline{\text{BER}}_{\text{Doppler}} = \frac{1}{2} \pi^2 (v_{\text{max}} T_B)^2 \quad (12.90)$$

对于其他多普勒频谱与调制方式,这个基本函数关系也成立,只是比例常数需要进行相应变化。

由这个关系可以发现,频率色散引起的差错对于低数据速率传输的系统具有非常重要的影响。例如,传输速率在 1 kbps 数量级的寻呼系统和传感器网络,当多普勒频移可达到几百赫兹时,其差错基底很容易达到 10^{-2} 数量级。对这些系统设计编码时必须要考虑这个情况。对于高数据速率系统(包括几乎所有当前的蜂窝、无绳和无线局域网系统),由频率色散引起

^① 对于一般的正交幅度调制,参考相位和参考振幅都有关。但是在下面的内容中,将对相移键控和频移键控的条件加以限制,这样就可以忽略幅度的失真。

的误差并不起主要作用^①。即使是日本的 JDC 蜂窝系统,其符号周期为 50 μs,由频率色散引起的误比特率仅为 10⁻⁴数量级,与由噪声引起的差错相比可以忽略不计。

时延色散

与频率色散相反,时延色散对高数据速率系统有非常大的影响。这可从下述现象明显看出:在非均衡系统中,差错由受到码间干扰的符号持续时间与符号未受干扰部分的比值决定。信道脉冲响应的最大附加时延由环境决定,与系统无关。下面,假设最大附加时延为 1 μs。在一个系统中,若符号持续时间为 20 μs,则码间干扰会干扰每个符号的 5%,但如果符号持续时间为 5 μs,码间干扰就会干扰每个符号的 20%。

理论和实验研究表明:由时延色散引起的差错基底可由下式给出:

$$\overline{\text{BER}} = K \left(\frac{S_r}{T_B} \right)^2 \tag{12.91}$$

其中, S_r 为信道的均方根(rms)时延扩展(见第 6 章)。与频率色散类似,差错主要出现在深度衰落(见第 5 章)附近。12.3.2 节将用深衰落附近达到最大值的群时延来解释这个问题。

式(12.91)仅在信道的最大附加时延远小于符号周期,且信道是瑞利衰落的情况下才有效。比例常数 K 与调制方法、发射机和接收机的滤波器、平均脉冲响应的形式及采样时刻的选择有关,下面将讨论这些内容。

采样时刻的选择

在平坦衰落信道中,显然采样时刻应选在判决部分中信噪比最大的时刻,这个时刻通常出现在比特转换或比特转换中间,在这两种情况下都称其为 $t_s = 0$ 。

对于具有时延色散的信道来说,采样时刻的选择就不再那么清晰了。理论上的推导通常假定 $t_s = 0$ (即在最小附加时延处采样)^②或者在平均时延处采样。后者实际上也就是针对某些功率时延谱(PDP,见第 6 章)的最佳采样时刻,如图 12.7 所示。

当根据信道的瞬时状态对采样时刻进行自适应地选择时,可以使误差基底明显减小,而且在无滤波系统中甚至可以消除。

功率时延谱形状的影响

最初不将时延色散的影响仅近似归结为均方根(rms)时延扩展对误比特率的影响。然而进一步的研究表明,功率时延谱的实际形状对误比特率也有影响。因为不同功率时延谱(相同的 S_r)下误比特率的变化因子通常小于 2,所以经常被忽略。图 12.8 表明,矩形功率时延谱的误比特率略大于双峰值功率时延谱的误比特率,差值为 75%。

指数功率时延谱导致更大的差错基底,并且对于这个形状,在平均时延处的采样不是最佳的。

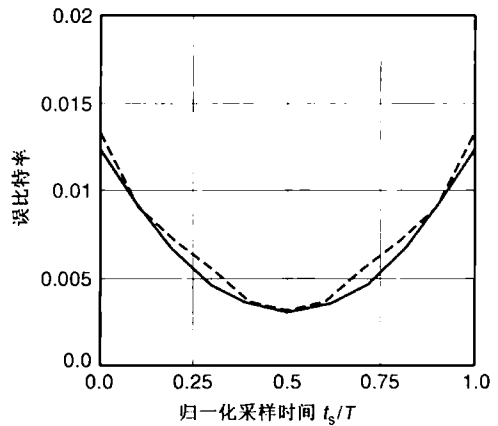


图 12.7 在双峰信道中,由时延色散引起的误比特率与采样时刻选择的关系。
引自 Molisch[2000]© Prentice Hall

① 这并不意味着信道中的时变特性对这样的系统不重要。信道变化也会对编码和信道检测的有效性等产生影响。

② 如果没有其他约束,假设 $\tau_0 = 0$ 。

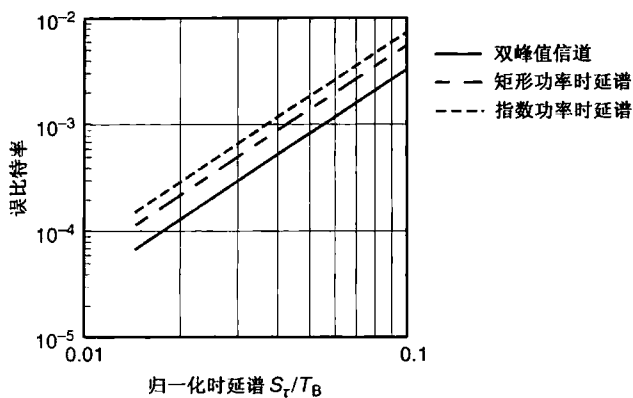


图 12.8 不同功率时延谱形状对时延色散引起的差错基底的影响。调制方法为差分检测的最小频移键控。引自 Molisch [2000] © Prentice Hall

滤波

在发射机和/或接收机中，滤波也会产生信号失真，因而使信号更易受到由信道失真引起的差错的影响。滤波器带宽越窄，差错基底就越大。图 12.9 给出了滤波对升余弦(RC)滤波器正交幅度调制的影响。

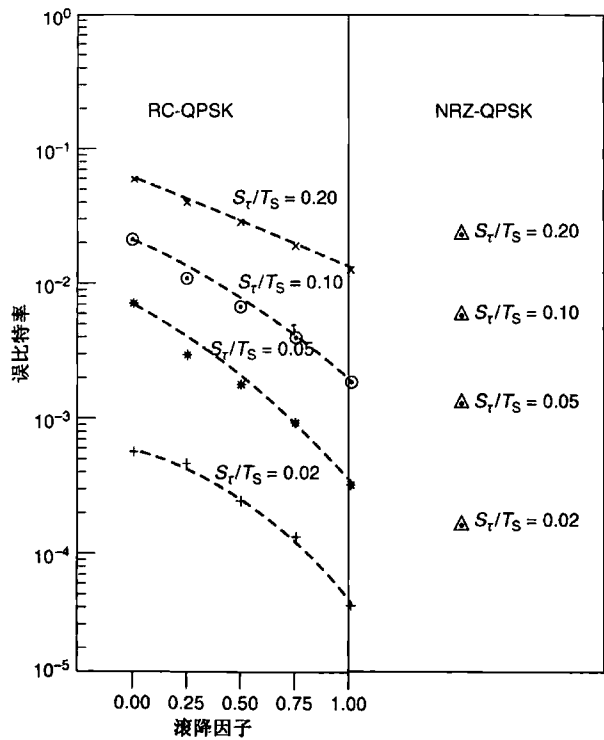


图 12.9 采用升余弦滤波器和相干检测的正交相移键控(QPSK)的差错基底，其中升余弦滤波器是滚降因子的函数。为了比较，图中还给出了常规QPSK的误比特率。注意升余弦(RC)QPSK还原为常规非归零(NRZ)QPSK时无滚降系数。引自 Chuang [1987] © IEEE

现在的问题是找到最佳滤波带宽。窄带滤波器(带宽在码元周期的倒数的数量级或更小)本身会带来很大的码间干扰。在没有其他干扰的情况下，即使所有判决都正确，这类滤波器也

会使系统对信道引入的码间干扰和噪声变得更敏感。而对于宽带滤波器,大部分噪声可以通过滤波器,这也将导致高的差错率。最佳滤波器带宽取决于时延色散与噪声之比。图 12. 10 举例说明了这种折中的一个例子。

调制方法

调制方法对差错基底也有一定的影响: 信号星座图上信号点离得越近, 调制方式对信道引起的失真就越敏感。当均方根时延扩展用同样的符号周期进行归一化时, 正交相移键控的差错基底要高于二进制相移键控的(见图 12. 11)。当比特持续时间假设相同时, 高阶调制形式的性能较好。

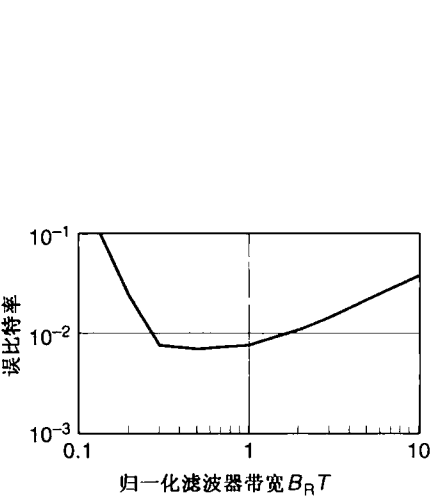


图 12. 10 采用差分检测的滤波最小频移键控的误比特率。归一化均方根时延扩展为0.1, 信噪比为 12 dB。引自 Molisch [2000] © Prentice Hall

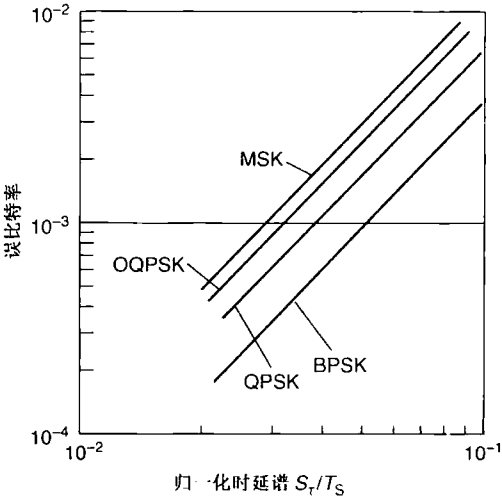


图 12. 11 不同调制方式的差错基底随归一化均方根时延扩展的变化曲线(用符号周期进行归一化)。引自 Chuang [1987] © IEEE

12.3.2 采用群时延法计算差错基底

这一节用一种简单方法来近似计算由时延色散引起的误比特率。Andersen [1991] 中分析相移键控, Crohn et al. [1993] 中分析最小频移键控时都曾介绍过该方法。接下来, 将这种方法用于差分检测的最小频移键控。群时延 T_g 定义为

$$T_g = - \left. \frac{\partial \Phi_c}{\partial \omega} \right|_{\omega=0} \tag{12.92}$$

其中, ω 是角频率。 $\Phi_c(\omega)$ 是信道传输函数的相位, 可以展开为泰勒级数:

$$\Phi_c(\omega) = \Phi_c(0) + \omega \left. \frac{\partial \Phi_c}{\partial \omega} \right|_{\omega=0} + \frac{1}{2} \omega^2 \left. \frac{\partial^2 \Phi_c}{\partial \omega^2} \right|_{\omega=0} + \cdots \tag{12.93}$$

级数的第一项是平均时延, 如果采样时刻是在平均时延处, 则可以省略。略去泰勒级数线性项后面的项, 可得

$$\Phi_c(\omega) = -\omega T_g \tag{12.94}$$

很明显, 采样时刻的相位失真取决于采样时刻的瞬时频率。如果传输的是 +1, 则瞬时频率为 $\pi/(2T_b)$, 否则频率为 $-\pi/(2T_b)$ 。当发送相同的比特(例如, 当前一个传的是 +1, 后一

个也是 +1 时,或当前一个传的是 -1,后一个也是 -1 时),相位差 $\Delta\Phi = 0$;当发送不同的比特时,则

$$\Delta\Phi = \pm \frac{\pi}{T_B} T_g \quad (12.95)$$

当由信道引起的相位失真(或采样时刻引起的相位偏差)大于 $\pi/2$,即

$$|T_g| > T_B/2 \quad (12.96)$$

时,会产生判决错误。众所周知,在瑞利衰落中,群时延的统计特性服从所谓的“学生 t 分布”

$$\text{pdf}_{T_g}(T_g) = \frac{1}{2S_r} \frac{1}{[1 + (T_g/S_r)^2]^{3/2}} \quad (12.97)$$

因此,误比特率可以通过式(12.97)很容易地计算出来,而且对可能出现的不同比特组合求平均时,可得

$$\text{BER} = \frac{4}{9} \left(\frac{S_r}{T_B} \right)^2 \approx \frac{1}{2} \left(\frac{S_r}{T_B} \right)^2 \quad (12.98)$$

例 12.7 考察采用差分检测的最小频移键控系统,其中 $T_B = 35 \mu\text{s}$,工作频率为 900 MHz,以 360 km/h(高速列车)的速度运动,指数功率时延谱, $S_r = 10 \mu\text{s}$ 。计算由频率色散和时延色散引起的误比特率。

解:

计算频率色散引起的误比特率,先要计算最大多普勒频移:

$$v_{\max} = f_c \frac{v}{c} = 9 \times 10^8 \frac{100}{3 \times 10^8} = 300 \text{ Hz} \quad (12.99)$$

对于差分检测的最小频移键控,假设具有经典的 Jake 多普勒谱,频率色散引起的误比特率由式(12.90)给出,因此

$$\begin{aligned} \overline{\text{BER}}_{\text{Doppler}} &= \frac{1}{2} \pi^2 (v_{\max} T_B)^2 \\ &= 5.4 \times 10^{-4} \end{aligned} \quad (12.100)$$

由时延色散引起的误比特率由式(12.98)给出,即

$$\begin{aligned} \overline{\text{BER}}_{\text{Delay}} &= \frac{4}{9} \left(\frac{S_r}{T_B} \right)^2 \\ &= 3.6 \times 10^{-2} \end{aligned} \quad (12.101)$$

12.3.3 一般衰落信道:二次型高斯变量法

计算色散衰落信道的误比特率的一般方法称为二次型高斯变量(QFGV)法^①。该信道存在瑞利和莱斯衰落,并受到频率色散和时延色散的影响,还存在加性高斯白噪声。

从数学上讲,QFGV 法决定了变量 $D < 0$ 的概率, D 为

$$D = A|X|^2 + B|Y|^2 + CXY^* + C^*X^*Y \quad (12.102)$$

其中, X 和 Y 是复高斯变量, A 和 B 是实常量, C 是复常量。按如下方式定义辅助变量:

^① 该方法是基于高斯变量某二次型的计算。该方法在具有奠基性的论文 Bello and Nelin[1963]和 Proakis[1968]之后命名。

$$\left. \begin{aligned} w &= \frac{AR_{xx} + BR_{yy} + CR_{xy}^* + C^*R_{xy}}{4(R_{xx}R_{yy} - |R_{xy}|^2)(|C|^2 - AB)} \\ v_{1,2} &= \sqrt{w^2 + \frac{1}{4(R_{xx}R_{yy} - |R_{xy}|^2)(|C|^2 - AB)}} \mp w \\ \alpha_1 &= 2(|C|^2 - AB)(|\bar{X}|^2R_{yy} + |\bar{Y}|^2R_{xx} - \bar{X}^*\bar{Y}R_{xy} - \bar{X}\bar{Y}^*R_{xy}^*) \\ \alpha_2 &= A|\bar{X}|^2 + B|\bar{Y}|^2 + C\bar{X}^*\bar{Y} + C^*\bar{X}\bar{Y}^* \\ p_1 &= \frac{\sqrt{2v_1^2v_2(\alpha_1v_2 - \alpha_2)}}{|v_1 + v_2|} \\ p_2 &= \frac{\sqrt{2v_1v_2^2(\alpha_1v_1 + \alpha_2)}}{|v_1 + v_2|} \end{aligned} \right\} \quad (12.103)$$

其中, R_{xy} 是二阶中心矩, $R_{xx} = \frac{1}{2}E\{(X - \bar{X})(Y - \bar{Y})^*\}$, 差错概率变为

$$P\{D < 0\} = Q_M(p_1, p_2) = \frac{v_2/v_1}{1 + v_2/v_1} I_0(p_1 p_2) \exp\left(-\frac{p_1^2 + p_2^2}{2}\right) \quad (12.104)$$

其中, Q_M 是式(12.40)定义的 Marcum Q 函数, 如果只有瑞利衰落, 而没有色散, 则 $\alpha_1 = \alpha_2 = 0$, 所以 $P(D > 0) = v_1/(v_1 + v_2)$ 。通常, 最大的问题是如何用公式表示差错概率 $P\{D < 0\}$, 下面将对此进行讨论。

典型的接收机

将不同的接收机结构简化成一个“典型的”接收机 [Suzuki 1982] 通常是很有用的, 其结构如图 12.12 所示。接收信号(经过带通滤波器)与参考信号相乘: 一个与同相参考信号相乘, 另一个与经过 $\pi/2$ 相移得到的正交参考信号相乘。所得到的信号经过低通滤波器后, 被送到采样判决器。在相干解调中, 这一参考信号是经过一个载波恢复电路得到的相干载波, 而在非相干解调中, 这一信号是由输入信号的时延得到的。

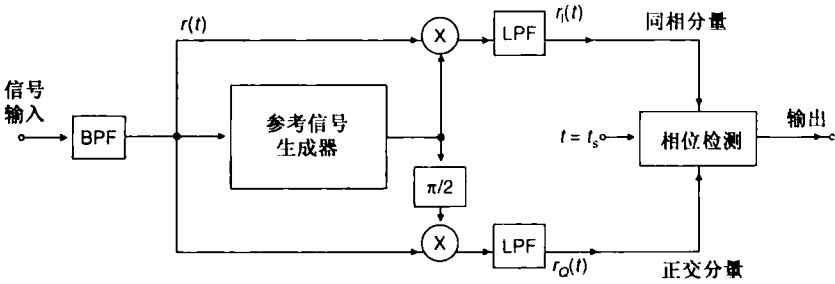


图 12.12 典型的接收机结构

以最小频移键控非相干解调为例, 用 $D < 0$ 方式很容易描述系统的比特错误。如果发送的是 +1, 但在采样时刻, 信号 $X = r(t_s)$, $Y = r(t_s - T)$ 的相位差在 π 和 2π 之间, 这时就会产生一个错码, 因而错误的条件与下式等同:

$$\text{Re}\{b_0XY^*\exp(-j\pi/2)\} < 0 \quad (12.105)$$

其中, b_0 是发送比特。因为 $\text{Re}\{Z\} = (Z + Z^*)/2$, 所以当 $A = B = 0$ 时, 式(12.105)就可以表示成 D 的二次方程。

12.3.4 误比特率

在差分检测中,应用 QFGV 方法的第一步就是计算 $X = r(t_i)$ 和 $Y = r(t_i - T)$ 的相关系数。对于采用鉴频器的检测,定义 X 为采样值 $r(t_i)$, Y 为 $t = t_i$ 时刻的斜率 $dr(t)/dt$, 相关系数的计算公式在 Adachi and Parsons[1989] 和 Molisch[2000] 中给出。

当求出相关系数后,通过式(12.104)求出误比特率的均值。在瑞利衰落条件下,采用二进制频移键控差分检测的平均误比特率可简化为

$$\overline{\text{BER}} = \frac{1}{2} - \frac{1}{2} \frac{b_0 \text{Im}\{\rho_{XY}\}}{\sqrt{\text{Im}\{\rho_{XY}\}^2 + (1 - |\rho_{XY}|^2)}} \quad (12.106)$$

对于 $\pi/4$ -DQPSK 有

$$\overline{\text{BER}} = \frac{1}{2} - \frac{1}{4} \left\{ \frac{b_0 \text{Re}\{\rho_{XY}\}}{\sqrt{(\text{Re}\{\rho_{XY}\})^2 + (1 - |\rho_{XY}|^2)}} + \frac{b'_0 \text{Im}\{\rho_{XY}\}}{\sqrt{(\text{Im}\{\rho_{XY}\})^2 + (1 - |\rho_{XY}|^2)}} \right\} \quad (12.107)$$

其中, b_0 和 b'_0 是构成符号的数据比特。

总结上面的讨论,可以将误比特率的计算步骤归纳如下。

1. 将实际接收机简化为典型结构;
2. 用式子表示 $D < 0$ 情况下出现错误的条件;
3. 计算 X 和 Y 的均值和相关系数;
4. 可以根据一般表示式(12.104)和式(12.103)来计算误比特率,也可以根据简化式(12.107)或式(12.106)来计算瑞利衰落条件下,最小频移键控和 $\pi/4$ -DQPSK 的误比特率。

深入阅读

本章介绍了包含了数百篇论文中的数字调制误比特率的计算。对于加性高斯白噪声信道中误比特率的计算,再次建议读者参考数字通信的教科书 Anderson [2005], Barry et al. [2003], Proakis [2005] 和 Sklar [2001]。在 Gallagher[2008] 中给出了极其严格的数学推导。在 Rice[2008] 中采用了一种时间离散的方法。在 Pawula et al. [1982], Proakis[1968] 和 Stein [1964] 中介绍了适用于非衰落和衰落信道的误比特率的基本计算方法。在很多论文中,对平坦衰落信道中的差错概率计算都进行了介绍。实际上,每种调制方式,与信道的幅度统计特性相结合,都至少有一篇文章给出了其结果。在瑞利衰落信道中有一些重要的例子,包括 Chennakeshu and Saulnier[1993] 分析的 $\pi/4$ -DQPSK, Varshney and Kumar[1991] 和 Yongacoglu et al. [1988] 分析的高斯最小频移键控,以及 Divsalar and Simon[1990] 分析的差分检测的 M 阶相移键控。

误比特率的另一些经典计算方法有 Simon and Alouini [2000] 中给出的通过矩生成函数的计算方法,或 Annamalai et al. [1990] 中给出的通过特征函数的计算方法。Simon and Alouini [2000] 中针对不同的调制方式、信道特性和接收机结构给出了相应的误比特率计算公式。

对于时延色散信道,已经提出了许多不同的计算方法。本章主要介绍了 QFGV 方法[Adachi and Parsons 1989, Proakis 1968] 和群时延方法[Andersen 1991, Crohn et al. 1993]。此外有许多论文采用了 Pawula et al. [1982] 推导的高斯矢量之间角的分布函数来计算。Chuang [1987] 是一篇可读性很强的文章,对不同的调制方式进行了对比。很多论文也致力于研究无均衡系统中能减小时延色散影响的调制方式和检测方法。Molisch[2000] 中总结了这些方法,并给出一些更新的文献。

第13章 分集

13.1 引言

13.1.1 分集的原理

前一章研究了在衰落信道中发送未编码比特流的常规的收发信机。对于加性高斯白噪声信道,这种收发信机的性能很不错,其误比特率随着信噪比的增加呈指数下降,且信噪比为10 dB时,误比特率在 10^{-4} 数量级。然而,在瑞利衰落信道中,误比特率随着信噪比线性减小,因而为了获得 10^{-4} 的误比特率,信噪比需达到40 dB数量级,这显然不实际。这两种情况的性能不同的主要原因是信道的衰落特性:误比特率主要取决于信道中深衰落的概率,因为此时的瞬时信噪比很低。因此改善误比特率的一种途径是改变信道的统计特性,即确保低信噪比的概率比较小。分集就是这样一种技术。

分集的基本原理是使同一信息通过多个统计独立的信道到达接收机。考虑一种简单情况:带有两个天线的接收机,且假设两个天线之间距离足够远,以确保两个天线所收到信号的小尺度衰落是相互独立的。接收机通常选择瞬时接收功率较大的天线支路^①。由于两个天线收到的信号是统计独立的,所以两个天线的信号同时处于深衰落的概率非常小(当然小于一个天线的信号处于深衰落的概率),因此分集改变了检测器输入端信噪比的统计特性。

例 13.1 两状态衰落信道的分集接收。

解:

为了量化分集的作用,考虑一个简单的数值计算的例子:在射频滤波器带宽中的噪声功率为50 pW,平均接收信号功率是1 nW,因而信噪比是13 dB。采用差分检测的频移键控,在加性高斯白噪声信道中的误比特率为 10^{-9} 。在衰落信道中,若90%时间内接收到的信号平均功率为1.11 nW,其信噪比为13.5 dB,其余时间接收到的信号平均功率为0,则在这种衰落信道中,90%时间的误比特率为 10^{-10} ,其余时间的误比特率为0.5,因而其平均误比特率为

$$0.9 \times 10^{-10} + 0.1 \times 0.5 = 0.05 \quad (13.1)$$

采用双天线分集时,两个天线的接收信号功率同时为0的概率为 $0.1 \times 0.1 = 0.01$,同时为1.11 nW的概率为 $0.9 \times 0.9 = 0.81$,一个天线接收信号功率为0,另一个天线接收信号功率为1.11 nW的概率为0.18。在后面的两种情况下,检测器的输入信噪比为13.5 dB,因而总的误比特率为

$$0.01 \times 0.5 + 0.99 \times 10^{-10} = 0.005 \quad (13.2)$$

这近似为单天线系统的误比特率的平方。如采用3个天线进行分集接收,则3个天线接收信号功率同时为0的概率为 0.1^3 ,总的误比特率为 $0.5 \times 0.001 + 0.999 \times 10^{-10} = 0.0005$;近似为单天线系统的误比特率的三次方。

① 在后面会看到这只是众多不同的分集方案之一。

后续章节将给出在瑞利衰落信道中采用分集技术时误比特率的精确计算公式,然而分集概念与上面这个简单的例子是一样的。采用 N_r 个分集天线,可得误比特率与 $\text{BER}_{\infty}^{N_r}$ 呈正比,其中 BER_{∞} 是单天线系统的误比特率^①。

下面首先介绍不同传输信道之间相关系数的特性,然后综述如何实现在多个独立的信道中传输(上面介绍的空间天线分集是一种方法,但当然不是唯一方法)。最后,介绍来自不同信道信号的最佳合并方式和不同合并方式的性能。

13.1.2 相关系数的定义

当不同的信道(又称分集支路)传输衰落特性相互独立的同一信号时,分集是最有效的一种技术。这意味着场强(或功率)的联合概率密度函数 $\text{pdf}_{r_1, r_2, \dots}(r_1, r_2, \dots)$ 等于信道的边缘概率密度函数 $\text{pdf}_{r_1}(r_1)$, $\text{pdf}_{r_2}(r_2)$, $\dots$ 的乘积。信道衰落之间的任何相关都会降低分集的效果。

相关系数表征不同分集支路信号之间的相关性,已有许多不同的定义来描述这个重要的量,如复相关系数或相位相关系数等。最重要的是信号的包络 x 和 y 的相关系数:

$$\rho_{xy} = \frac{E\{x \cdot y\} - E\{x\} \cdot E\{y\}}{\sqrt{(E\{x^2\} - E\{x\}^2) \cdot (E\{y^2\} - E\{y\}^2)}} \quad (13.3)$$

对于两个统计独立的信号,式 $E\{x \cdot y\} = E\{x\}E\{y\}$ 成立,因而相关系数为 0。当相关系数低于某一门限值(典型数值为 0.5 或 0.7)时,信号一般视为被有效地去相关。

13.2 微分集

正如本章引言所提到的,分集的基本原理是接收机收到发射信号的多个副本,每个副本都经历了衰落特性统计独立的信道传输。本节介绍得到这些统计独立副本的几种不同方法,其中重点介绍用于克服小尺度衰落的方法,因而也称为“微分集”。5 种最常用的方法如下。

1. 空间分集:利用空间分离的多个天线。
2. 时间分集:在不同的时刻接收发射信号。
3. 频率分集:在不同载频上传输信号。
4. 角度分集:使用不同天线方向图的多个天线(空间分离或不分离)。
5. 极化分集:多个天线接收不同极化方向的信号(如垂直极化和水平极化)。

当提到天线分集时,指的是在接收机中有多个天线。仅在 13.6 节(和第 20 章)中讨论如何使用多个发射天线来提高系统性能。

在分集中经常会用到的一个重要关系式为:假设两个信号的时间间隔为 τ , 频率间隔为 $f_1 - f_2$, 如附录 13. A (www.wiley.com/go/molisch) 所示,其相关系数为

$$\rho_{xy} = \frac{J_0^2(k_0 v \tau)}{1 + (2\pi)^2 S_f^2 (f_2 - f_1)^2} \quad (13.4)$$

注意,对于正在移动的移动台,时间间隔很容易转化为空间间隔,因此时间分集和空间分集在数学上是等价的。在这种意义上,式(13.4)可用于空间分集、时间分集和频率分集。然而,在推导

① 在瑞利衰落信道中, $\text{BER}_{\infty} \propto \text{SNR}^{-1}$, 可发现,在有 N_r 个独立衰落信道的分集系统中, $\text{BER} \propto \text{SNR}^{-N_r}$ 。在采用分集的衰落信道中,一般情况下, $\text{BER} \propto \text{SNR}^{-d_{\text{div}}}$, 其中 d_{div} 是分集阶数。

该式时进行了一些假设：(i)信道模型是广义平稳非相关散射(WSSUS)模型；(ii)不存在视距传播；(iii)指数功率时延谱；(iv)入射功率各向同性分布；(v)使用全向天线。

例 13.2 在 COST^① 207 信道模型中定义的“典型城市”环境中，分别计算两频率间隔为下列数值时的相关系数：(i)30 kHz；(ii)200 kHz；(iii)5 MHz。

解：

对于零时间间隔，式(13.4)中的 Bessel 函数是 1，因此相关系数只取决于均方根时延扩展和频率间隔。均方根时延扩展可以根据式(6.39)计算，7.6.3 节给出了 COST 207 典型城市信道模型中的功率时延谱。因此，可得

$$S_r = \sqrt{\frac{\int_0^{7 \times 10^{-6}} e^{-\tau/10^{-6}} \tau^2 d\tau}{\int_0^{7 \times 10^{-6}} e^{-\tau/10^{-6}} d\tau} - \left(\frac{\int_0^{7 \times 10^{-6}} e^{-\tau \cdot 10^{-6}} \tau d\tau}{\int_0^{7 \times 10^{-6}} e^{-\tau \cdot 10^{-6}} d\tau} \right)^2}$$
$$= 0.977 \mu s$$

(13.5)

相关函数为

$$\rho_{\Delta f} = \frac{1}{1 + (2\pi)^2 (0.977 \cdot 10^{-6})^2 (f_1 - f_2)^2}$$
$$= \begin{cases} 0.97, & f_1 - f_2 = 30 \text{ kHz} \\ 0.4, & f_1 - f_2 = 200 \text{ kHz} \\ 1 \times 10^{-3}, & f_1 - f_2 = 5 \text{ MHz} \end{cases}$$

(13.6)

可以看出，相邻间隔为 30 kHz 的两个信道(例如在 IS-136^② 时分多址蜂窝系统中采用 30 kHz 的信道间隔)的相关系数很大；相邻间隔超过 200 kHz(在全球移动通信系统中两相邻信道间隔 200 kHz，见第 24 章)时，相关系数仍然较大。但在这种环境中，宽带码分多址的两相邻信道(间隔 5 MHz)是不相关的。此外，在全球移动通信系统中，基站-移动台和移动台-基站之间采用频分双工，收发间隔为 45 MHz，收发载波之间的信道是完全不相关的(见第 17 章和第 24 章)。

13.2.1 空间分集

空间分集是最早并且最简单的分集方式，尽管(或因为)如此，它还是应用最广泛的一种分集方式。在空间分集中，发射信号被多个天线接收，然后天线接收到的这些信号会根据将在 13.4 节中介绍的方法被进一步处理。但是，无论使用何种处理方法，其性能都会受到各天线接收信号之间的相关性的影响。我们不希望接收信号之间具有很强的相关性，因为这会降低分集的效果。设计分集天线的第一个重要步骤就是建立天线间距和相关系数之间的关系。对于基站天线与移动台天线而言，这种关系是不同的，因此必须分别对待。

1. 蜂窝和无绳系统中的移动台。通常，假设电波从各个方向进入移动台(见第 5 章)。因此，建设性的和破坏性的多径分量干扰的位置(也就是接收功率高和接收功率低的点)大约相距 $\lambda/4$ ，这是实现接收信号之间去相关所需的距离。这种直观的认识与通过严格的数学推导得出的结果是一致的[式(13.4)中， $f_2 - f_1 = 0$ 时]。在图 13.1 中，若定义 $\rho = 0.5$ 时去相关，则当天线间距为 $\lambda/4$ 时，可以去相关，也可与例 5.3 进行比较。

上面的讨论意味着在全球移动通信系统(900 MHz)中，各个天线之间的最小距离约

① 欧洲科技领域研究合作组织。
② IS-136 是使用时分多址技术的第二代蜂窝系统(现已基本不用)。

为 8 cm, 而对于工作在 1800 MHz 频段的各种无绳和蜂窝系统, 最小距离约为 4 cm。对于无线局域网(在 2.4 GHz 和 5 GHz 频段), 此距离会更小。显然, 在蜂窝系统中的一个移动台上放置两个天线是可行的。

2. 无绳系统和无线局域网中的基站。首先近似认为, 在室内基站, 入射波在各个角度均匀分布。也就是说, 来自各个方向的入射波是等强度的, 这对于移动台同样适用。
3. 蜂窝系统中的基站。对于蜂窝系统中的基站, 在各个方向均匀入射的假设不再成立。相互作用体集中在移动台周围(如图 13.2 所示, 见第 7 章)。由于所有波基本上从一个方向入射, 相关系数(对于给定的天线之间的距离 d_a)会大得多。因为表现出的特性不同, 所以实现充分去相关所需的天线间距会增加。

为了得到直观的认识, 从一个简单的例子开始: 只有两个多径分量, 多径分量的波向量夹角为 α (如图 13.3 所示, 见第 5 章)。显然, α 越小, 干涉图中最大值和最小值之间的距离就越大。对于很小的 α , 天线之间的连线落在干涉图的“脊”上, 天线之间则完全相关。相关系数的数值与天线之间距离的函数关系如图 13.4 所示。第一列表示功率谱均匀分布的结果; 高斯分布的结果如第二列所示。可以看出, 角度扩展在 $1^\circ \sim 5^\circ$ 之间变化时, 天线之间的距离需 2 ~ 20 个波长, 才能达到去相关。还可以发现, 主要是均方根角扩展决定了天线之间所需的距离, 而角度功率谱的形状只产生极小的影响。

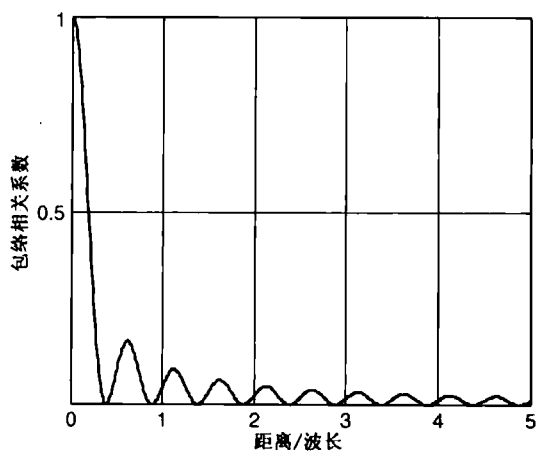


图 13.1 包络相关系数与天线之间距离的函数关系

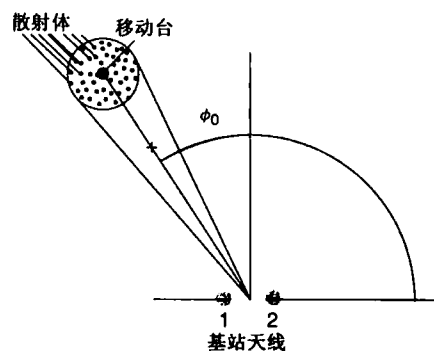


图 13.2 集中在移动台周围的散射体

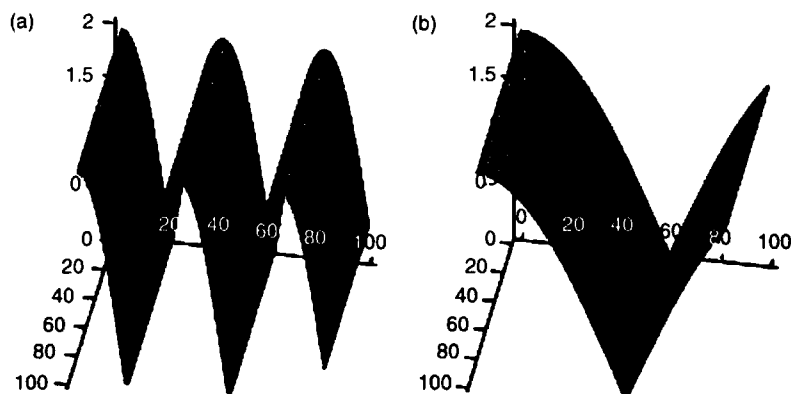


图 13.3 两波夹角为 45° (a) 和 15° (b) 时的干涉图形

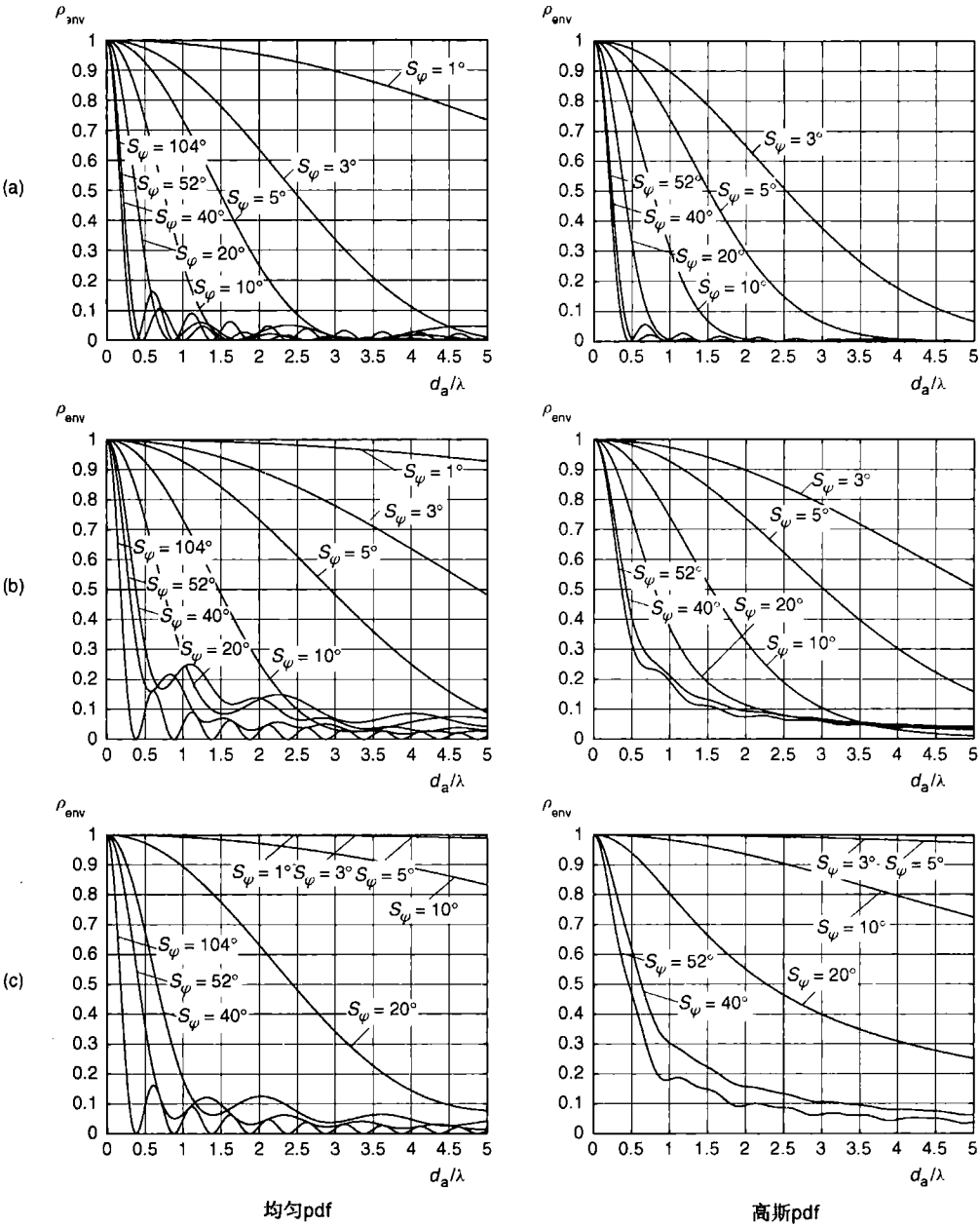


图 13.4 到达方向的概率密度函数(pdf)分别为均匀(左)和高斯(右)分布时,基站信号的包络相关系数, ϕ_0 为到达角, (a) $\phi_0 = 90^\circ$, (b) $\phi_0 = 45^\circ$, (c) $\phi_0 = 10^\circ$, S_ϕ 为角度扩展。引自 Fuhl et al. [1998] © IEEE

13.2.2 时间分集

由于无线传输信道是时变的,所以在不同时刻接收到的信号是不相关的。要想“充分”地去相关,时间间隔必须至少为 $1/2\nu_{\max}$, 其中 $\nu_{\max}$ 为最大多普勒频移。在静态信道中,无论是发射机、接收机还是相互作用体都是静止的,因此信道状态在任何时刻都是恒定的;例如,对于无线局域网,就会出现这种情况。此时,对于所有的时间间隔,相关系数 $\rho = 1$, 所以时间分集是无效的。

时间分集可以采用以下几种不同的实现方法。

1. 重复编码。这是最简单的一种形式。信号重复发送多次,而重复时间间隔要足够大,以达到去相关。这显然能获得分集,但也会大大降低带宽效率。频谱效率降低的因子就等于信号重复发送的次数。
2. 自动重传请求(ARQ)。这种情况下,接收机会给发射机发送一个信息,表明它是否以足够的质量收到了数据,参阅附录 14. A (www.wiley.com/go/molisch)。如果不满足要求,发射机将会重新发送数据(在经过一段可以达到去相关的等待时间后)。自动重传请求的频谱效率要比重复编码好,因为只有在第一次传输遇到严重衰落时,才需要多次发送,而重复编码则会一直重复发送。在下行信道中,自动重传请求需要反馈信道。
3. 交织和编码的结合。一种更高级的重复编码方案是带交织的前向纠错编码。一个码字的不同符号在不同的时间传输,这样其中一些符号以较高信噪比到达接收机的概率就会增加。发送码字在接收端会被重建。更详细的内容可参阅第 14 章。

13.2.3 频率分集

在频率分集中,同一信号在两个(或以上)不同的频率上传输。如果这些频率之间的间隔大于信道的相关带宽,则这些频率上的衰落近似是相互独立的,因此信号在两个频率上同时出现深度衰落的概率很小。对于指数形式的功率时延谱,通过设置式(13.4)中的分子为 1,可以得到同一时刻的两个不同频率信号的相关系数,即

$$\rho = \frac{1}{1 + (2\pi)^2 S_f^2 (f_2 - f_1)^2} \quad (13.7)$$

这再次证明了两信号之间必须至少间隔一个相关带宽。图 13.5 给出了 ρ 与两个频率间隔之间的函数关系。关于频率相关性更全面的讨论可参阅第 6 章。

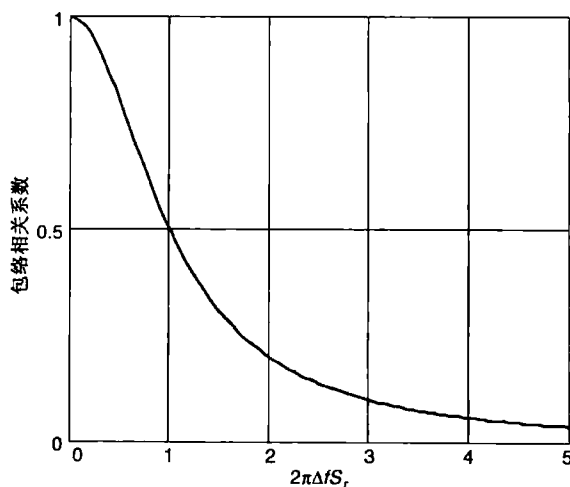


图 13.5 包络相关系数与归一化频率间隔的函数关系

在两个不同频率上重复传输相同的信息实际上并不常用,因为这会大大降低频谱效率。然而,可将信息扩展到较大的带宽上,这样各部分信息就能通过不同的频率分量进行传输。然后,接收机会综合不同频率分量上的信息,从而恢复原始信息。

这种扩展有以下几种不同的实现方法。

- 在时间上压缩信息,也就是发送占用很大带宽的短突发,即时分多址(见第 17 章);
- 码分多址(CDMA)见 18.2 节;
- 多载波 CDMA(见 19.9 节)和编码正交频分复用(见 19.4.3 节);
- 带编码的跳频,即在不同的载波频率上传送码字的不同部分(见 18.1 节)。

这些方法可以传输信息,且不浪费带宽,第 17 章至第 19 章将进行更详细的介绍。目前,我们只强调了采用频率分集要求信道是频率选择性信道。换句话说,频率分集(时延色散)可使系统更健壮,降低衰落的影响。这似乎与第 12 章的结果矛盾,因为在第 12 章中提到频率选择性会增大误比特率,甚至会产生差错基底。产生这种矛盾的原因是,第 12 章只考虑了一种简单的接收机,它不采取任何措施来减小频率选择性的影响。

13.2.4 角度分集

当来自不同方向的多径分量之间发生相消干涉时,将产生深度衰落。如果这些电波中的一部分被衰减或消除,则深度衰落的位置会改变。也就是说,两个方向图不同但在同一位置的天线对多径分量的增益不同,因此多径分量对两个天线的干扰不同。这就是角度分集的原理(又称模式分集)。

角度分集经常与空间分集结合使用;角度分集可增强距离很近的天线接收信号之间的去相关性。不同的天线方向图是很容易实现的。当然,不同类型的天线具有不同的方向图。但是,当两个天线安装得很近时,即使相同的天线也可以有不同的方向图(见图 13.6)。产生这种结果的原因是两个天线之间的相互耦合:天线 B 作为天线 A 的反射器,天线 A 的方向图因此向左倾斜^①。与之类似,由于来自天线 A 的反射,天线 B 的方向图向右倾斜。因此,两个天线的方向图是不同的。

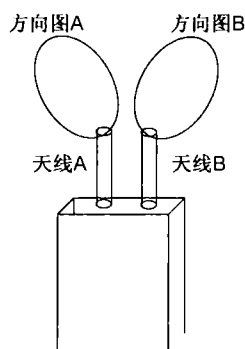


图 13.6 距离很近的天线的角度分集

当天线固定在机壳的不同部分时,方向图的不同更明显。偶极天线一般限于机壳的顶部,而贴片天线和倒 F 形天线(见第 9 章)能放置在机壳的任何部分(见图 13.7)。在所有这些情况下,即使天线彼此放置得很近,也会有很好的去相关特性。

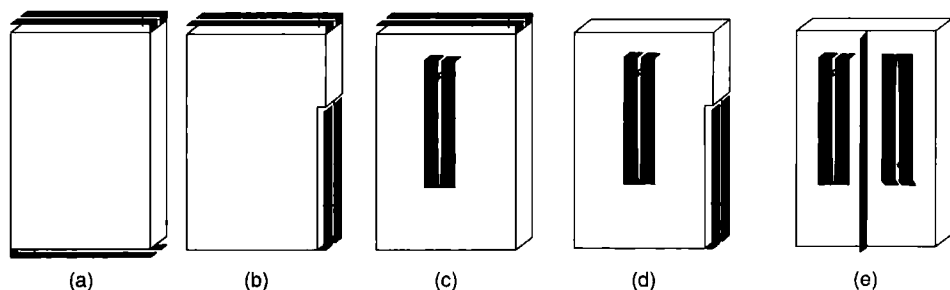


图 13.7 移动台分集天线的结构。引自 Erätuuli and Bonek [1997] © IEEE

^① 这种排列也可以看成 Yagi 天线,无论天线 B 是导向器还是反射器,以及方向图向哪个方向倾斜,它都取决于两个天线之间的距离。

13.2.5 极化分集

水平和垂直极化的多径分量在无线信道中的传播是不同的^①，这是由于反射和绕射过程与极化方式有关(见第 4 章和第 7 章)。虽然发射天线只发射单一极化的信号，但信道传播特性的影响会导致去极化，因此会有两个极化方式的信号到达接收机。极化方式不同的信号，其衰落是统计独立的。因此，用双极化天线接收两个极化方式的信号，并分别对信号进行处理，就能实现分集。这种分集的实现对于天线之间的最小间距没有任何要求。

考虑一种更具体的情况，在垂直极化方向上发射信号，而在垂直和水平极化两个方向上接收。在这种情况下，两个接收信号之间的衰落是统计独立的，但在两个分集支路上的平均接收信号强度是不同的。由于环境的影响，水平极化方向(即横向极化)的分量的强度要比垂直极化方向(即互极化)的分量低 3 ~ 20 dB。后面将会看到，这对分集方案的有效性有很重要的影响。目前已提出多种天线排列方式来解决这个问题。

已经证明极化分集的分集阶数可达到 6 阶：E 域的 3 个分量和 H 域的 3 个分量都可以利用[Andrews et al. 2001]^②。然而，电波的传播特性和实际情况会使分集达不到其最高阶数，尤其是在室外环境。对于分集系统来说，这不是一个严重的限制，后面将会看到，从 1 阶分集(即无分集)到 2 阶分集所得到的性能改善要优于从 2 阶分集到更高阶数的分集所得到的改善。但是，对于多输入多输出系统(MIMO)，这是一个重要的问题(见 20.2 节)。

13.3 宏分集和同播

前一节介绍了克服小尺度衰落(也就是由相互干扰影响引起的衰落)的分集方法。然而，并不是所有分集方法都适合用来克服阴影效应引起的大尺度衰落。阴影对于发射频率和极化基本上是独立的，因此频率分集或极化分集无效。而空间分集(或者相当于有移动发射机/接收机的时分集)可以使用，但需要知道大尺度衰落的相关距离是在十米还是百米的数量级上。也就是说，如果发射机和接收机之间有一座山，在基站或移动台增加天线也不能消除这座山的阴影影响。相反地，可以利用一个独立的基站(BS2)，且 BS2 的位置应该使这座山不在移动台和 BS2 之间的连线上。这就意味着 BS1 和 BS2 之间会有很大的距离，这就是“宏分集”的由来。

最简单的宏分集方法是使用频率中继器，中继器用来接收信号，进行放大后再发送出去。同播与这种方法很相似：不同的基站同时发送相同的信号。在蜂窝应用中，这两个基站必须同步，使发送给某一用户的两个信号波几乎同时到达接收机(定时超前)^③。必须指出，只有信号从两个基站到移动台的传播时间已知，才能实现同步。总的来讲，希望同步误差不要大于接收机所能处理的时延色散。更重要的是，在接收机所在的区域，来自两个基站的信号的强度必须近似相等。

在广播中，尤其是数字电视，同播同样得到了广泛应用。在这种情况下，所有可能的接收机实现精确同步是不可能的，因为每个接收机需要从发射机中得到不同的定时超前。

① 为了简便，称其为水平极化和垂直极化，但是这对于任何两个正交的极化都是成立的。

② 注意，这个结果存在一定的争议，对于能否仅仅通过极化分集达到 6 阶分集，至本书落笔时科学界还没有达成统一的认识。

③ 如果接收机不易处理时延色散，就要求两个信号到达接收机的时间是严格相同的。对于高级的接收机(见第 16 章至第 19 章)，可以允许有少量的时延色散；如果基站定时和信号从发射机到接收机的时间能准确得到，这也可以实现。

同播的一个缺点是大量的信令信息需要通过通信电缆传输。同步信息及发送数据也需要通信电缆(或者微波链路)传输到基站。在数字移动电话的早期,这曾经是一个比较严重的问题,但目前光纤链路的广泛应用使这个问题很容易解决。

由于不需要同步,所以频率中继器的使用比同播简单一些。但另一方面,使用中继器会导致时延色散较大,因为,(i)信号从基站到中继器和从中继器到移动台的传输时间较长(与从另一个基站到移动台的传输时间相比);(ii)由于电子器件、滤波器等存在时延,中继器本身也会产生附加的时延。

13.4 信号的合并

下面重点讨论如何利用分集信号来提高检测信号的总质量。为了简化问题,这里只介绍接收机如何合并来自不同天线的信号。当然,这些数学方法对于其他类型的分集信号仍然有效。总的来讲,可将多个分集支路信号的处理方法分为如下两种。

1. 选择式分集。选择并处理(解调和解码)“最佳”信号副本,其余的副本全部丢弃。对于“最佳”信号的判断有多个不同的标准。
2. 合并分集。合并所有的信号副本(在解调器之前或之后),再对合并的信号进行解码。对于信号的合并也有多种不同的算法。

由于利用了全部可用的信息,所以合并分集具有更好的性能。在下行链路中,合并分集比选择式分集需要更复杂的接收机。在大多数接收机中,所有的处理都是在基带完成的。因此,采用合并分集的接收机需要对所有可用信号都进行下变频,然后在基带进行适当的合并,因此它需要 N_r 个天线分量和 N_r 个完整的射频(下变频)链路。而采用选择式分集的接收机在一个时刻只处理单个接收信号,所以它只需要一个射频链路。

下面会更详细地介绍选择(合并)规则和算法。假设不同信号副本经历了统计独立的衰落,这会大大简化关于信号合并的直观解释和数学分析。关于有限相关系数的影响将在 13.5 节讨论。

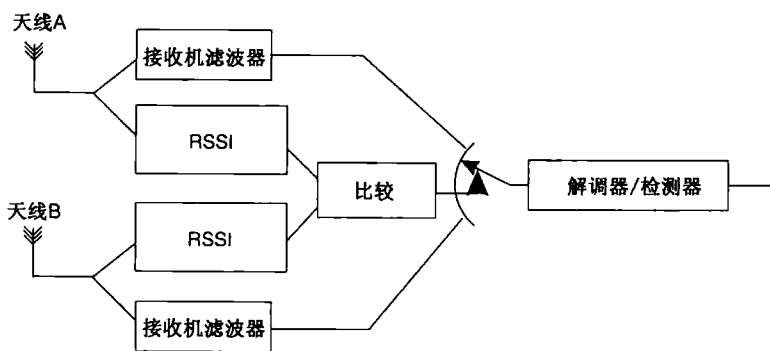
除了这些,还必须记住多个天线的增益有两部分:分集增益和波束成形增益。分集增益反映了多个天线接收的信号同时经历深度衰落是几乎不可能的;因此通过使用多个天线,信号出现很低电平的概率会减小。波束成形增益反映了合并器对不同天线上的噪声进行平均的效果(对于合并分集)。因此,即使所有天线上的信号电平都一样,合并器的输出信噪比要比只有单个天线时的信噪比大。

13.4.1 选择式分集

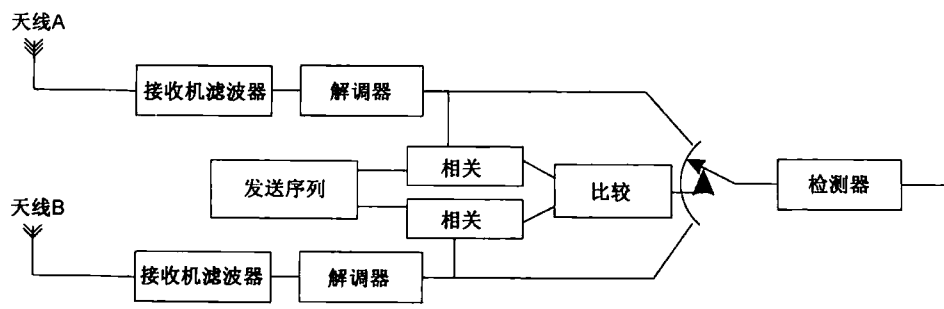
接收信号强度指示驱动分集

在这种方法中,接收机选择瞬时接收功率(或接收信号强度指示, RSSI)最大的信号,并做进一步处理。这种方法需要 N_r 个天线, N_r 个接收信号强度指示传感器,以及一个 $N_r:1$ 转换开关,但只需要一个射频链路(见图 13.8)。即使在快衰落信道中,此方法也可采用简单的选择准则进行跟踪。因此,只要接收信号强度指示变得较大,就可以切换到一个较好的天线。

1. 如果误比特率由噪声决定, 那么接收信号强度指示驱动分集是最好的选择式分集方法, 因为接收信号强度指示的最大值也是信噪比的最大值。
2. 如果误比特率由信道之间的相互干扰决定, 那么接收信号强度指示不再是一个好的选择准则。高电平干扰会产生高的接收功率, 因此接收信号强度指示准则会使系统选择信干比较小的支路。当干扰主要由一个占主导作用的干扰源决定时, 这种情况尤为严重, 这是频分多址或时分多址系统的一种典型情况。
3. 类似地, 如果差错是由信道的频率选择性引起的, 那么接收信号强度指示驱动分集是次最佳的。但它仍是一种合理的近似方法, 因为, 在第 12 章讲过, 由信号失真引起的差错主要出现在信道的深度衰落中。但这只是一种近似方法, 可以证明, 与最佳分集(误比特率驱动分集)相比, 接收信号强度指示驱动分集系统(无编码、无均衡)的误比特率仅高了一个常数因子。



(a) 接收信号强度指示驱动分集



(b) 误比特率驱动分集

图 13.8 选择式分集的原理

为了得到准确的性能估计(见 13.5 节), 确定选择器的输出信噪比分布是很重要的。假设信号瞬时幅度是瑞利分布, 则第 n 个分集支路的信噪比 γ_n 为[见式(12.48)]

$$\text{pdf}_{\gamma_n}(\gamma_n) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma_n}{\bar{\gamma}}\right) \quad (13.8)$$

其中, $\bar{\gamma}$ 为支路平均信噪比(假设对所有分集支路是相同的)。那么累积分布函数(cdf)为

$$\text{cdf}_{\gamma_n}(\gamma_n) = 1 - \exp\left(-\frac{\gamma_n}{\bar{\gamma}}\right) \quad (13.9)$$

根据定义, 累积分布函数为瞬时信噪比低于给定门限值的概率。由于接收机选择信噪比最大的支路, 所以所选信号信噪比低于给定门限值的概率等于每个支路的信噪比低于给定门限值

的概率的乘积。也就是说,被选信号的累积分布函数是每个支路的累积分布函数的乘积:

$$\text{cdf}_\gamma(\gamma) = \left[1 - \exp\left(-\frac{\gamma}{\bar{\gamma}}\right) \right]^{N_r} \quad (13.10)$$

例 13.3 一个选择式分集系统,当使用的天线个数分别为 $N_r = 1, 2, 4$ 时,计算其输出功率低于每个支路平均功率 5 dB 的概率。

解:

门限 $\gamma_{\text{dB}} = \bar{\gamma}_{\text{dB}} - 5 \text{ dB}$, 即线性表示为 $\gamma = \bar{\gamma} \times 10^{-0.5}$ 。利用式(13.10), 输出功率低于 $\bar{\gamma} \times 10^{-0.5}$ 的概率为

$$\begin{aligned} \text{cdf}_\gamma(\bar{\gamma} \times 10^{-0.5}) &= [1 - \exp(-10^{-0.5})]^{N_r} \\ &= \begin{cases} 0.27, & N_r = 1 \\ 7.4 \times 10^{-2}, & N_r = 2 \\ 5.4 \times 10^{-3}, & N_r = 4 \end{cases} \end{aligned} \quad (13.11)$$

例 13.4 假设 $N_r = 2$, 两个支路的平均功率分别为 $1.5\bar{\gamma}$ 和 $0.5\bar{\gamma}$, 结果会有什么变化?

解:

在此条件下, 概率为

$$\text{cdf}_\gamma(\bar{\gamma} \times 10^{-0.5}) = \left[1 - \exp\left(-\frac{1}{1.5} \times 10^{-0.5}\right) \right] \left[1 - \exp\left(-\frac{1}{0.5} \times 10^{-0.5}\right) \right] \quad (13.12)$$

$$= 8.9 \times 10^{-2} \quad (13.13)$$

由此可以证明, 当各支路的平均功率不同时, 分集的性能会降低。

误比特率驱动分集

对于误比特率驱动分集, 首先发送一个训练序列, 即接收机已知的一个比特序列。然后接收机分别解调每个天线接收到的信号, 并与发射信号进行比较。信号具有最小误比特率的天线被判为“最佳”, 并用于后面数据信号的接收。另一种类似的方法是利用“软判决”解调信号的均方差, 或者利用发射信号和接收信号的相关性。

如果信道是时变的, 那么训练序列需要定期重复发送, 并且需要重新选择最佳天线。所需的重复速率由信道的相干时间决定。

误比特率驱动分集有如下几个缺点。

1. 接收机需要 N_r 个射频链路和解调器(这会使接收机变得复杂), 或训练序列需要重复发送 N_r 次(这会降低频谱效率), 这样所有天线上的信号质量才能估计出来。
2. 如果接收机只有一个解调器, 就不可能持续监测所有分集支路的选择标准(即误比特率)。特别是当信道变化很快时, 这一点尤其严重。
3. 由于训练序列的持续时间是有限的, 所以选择的标准(误比特率)是不能精确确定的。当训练序列的持续时间增加时, 误比特率围绕其真实均值的方差会减小。因而, 在选择标准的错误判决导致的性能损失与较长训练序列导致的频谱效率降低之间, 需要进行折中考虑。

13.4.2 开关分集

选择式分集的主要缺点是, 为了确定选择另一个不同天线的时刻, 需要监视所有分集支路

的选择标准(例如功率,误比特率等)。如上所述,这会增加硬件的工作量或降低频谱效率。避免这些缺点的解决方案之一就是开关分集。在开关分集中,只需监视有效分集支路的选择标准。当所选支路的质量低于某一门限值时,接收机则转向另一个天线支路^①。开关转换只取决于有效支路的质量,与其余支路是否提供更好质量的信号无关。

如果两个支路的信号质量都低于门限值,开关分集会出现问题。这时,接收机会在两个支路之间来回切换。利用一段滞后或保持时间可以避免这种情况,因此新的分集支路通常被使用一段时间,这段时间与实际的信号质量无关。这里有两个自由参数:开关门限值和滞后时间。这些参数需要认真选取,如果门限值选得太低,那么当选择某一个分集支路时,另一个天线也许能提供更好的质量;如果门限值选得太高,那么接收机切换到的支路提供的信号质量可能低于当前的有效支路。如果滞后时间选得太长,就有可能在很长时间内使用较差的分集支路;如果滞后时间选得太短,接收机就会浪费全部时间在两个支路之间来回切换。

总之,开关分集的性能要低于选择式分集的性能,因此我们不再深入地进行讨论。

13.4.3 合并分集

基本原理

选择式分集丢弃了 $N_r - 1$ 个接收信号的副本,浪费了信号的能量。而合并分集克服了这个缺点,利用了全部可用的信号副本。每个信号副本乘以一个 w_n^* (复值)权重,然后相加。每个复值权重可看成由校正相位^②和幅度的(实)权重组成。

- 相位校正使信号幅度相加,但是另一方面,噪声也会非相干地相加,从而噪声功率也会加起来。
- 对于幅度加权,应用较广泛的有两种方法:最大比值合并(MRC)将所有信号副本的幅度进行加权,(在一定的假设下)可以证明这是最佳的合并策略。另一种方法是等增益合并(EGC),其所有的幅度权重都相同(换句话说,就是不对信号进行加权,只进行相位校正)。这两种方法如图 13.9 所示。

最大比值合并

最大比值合并方法是根据不同天线支路信号的信噪比,对其进行相位补偿和加权。如果几个假设都满足,这是合并不同分集支路的最佳方法。假设传输信道是平坦慢衰落信道,那么唯一的干扰是加性高斯白噪声。在这些假设下,每个信道实现可看成一个时不变滤波器,其脉冲响应为

$$h_n(\tau) = \alpha_n \delta(\tau) \quad (13.14)$$

其中, α_n 是分集支路 n 的(瞬时)衰减。信号的不同支路乘以不同的权重 w_n^* 并相加,因此信噪比变为

$$\frac{\left| \sum_{n=1}^N w_n^* \alpha_n \right|^2}{P_n \sum_{n=1}^N |w_n|^2} \quad (13.15)$$

① 这种方法主要用于有两个分集支路可用的情况。

② 注意,在我们的注释中,信号的权重是 w 的复共轭;这将会简化后续的注释。

其中, P_n 是每个支路的噪声功率(假定每个支路的都相同)。根据柯西-施瓦兹不等式 $\left| \sum_{n=1}^N w_n^* \alpha_n \right|^2 \leq \sum_{n=1}^N |w_n|^2 \sum_{n=1}^N |\alpha_n|^2$, 其中当且仅当 $w_n = \alpha_n$ 时等式成立。从而, 当选择权重如下时信噪比有最大值:

$$w_{\text{MRC}} = \alpha_n \tag{13.16}$$

例如, 信号相位校正(注意接收信号要乘以 w^*)并且幅度加权。很容易看出, 分集合并器输出的信噪比是各个支路信噪比之和:

$$\gamma_{\text{MRC}} = \sum_{n=1}^{N_r} \gamma_n \tag{13.17}$$

如果各支路是统计独立的, 那么总的信噪比的矩生成函数可看成各支路信噪比的特征函数的乘积。此外, 如果各个支路的信噪比服从指数分布(对应于瑞利衰落), 并且所有支路有相同的平均信噪比, $\bar{\gamma}_n = \bar{\gamma}$, 通过计算可以得出:

$$\text{pdf}_{\gamma}(\gamma) = \frac{1}{(N_r - 1)!} \frac{\gamma^{N_r - 1}}{\bar{\gamma}^{N_r}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right) \tag{13.18}$$

合并器输出的平均信噪比是支路平均信噪比与分集支路个数的乘积:

$$\bar{\gamma}_{\text{MRC}} = N_r \bar{\gamma} \tag{13.19}$$

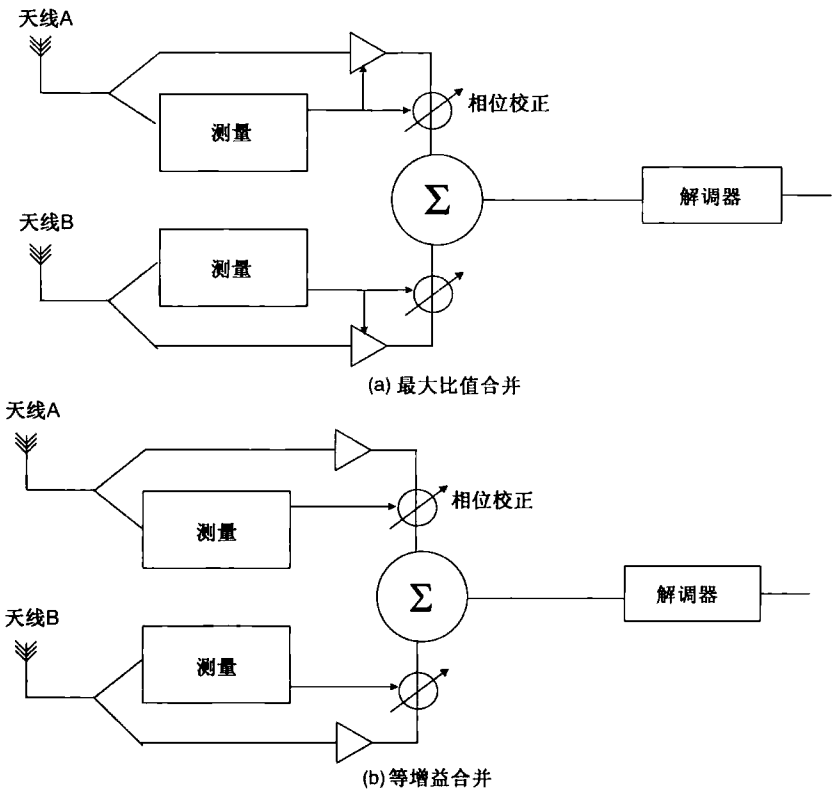


图 13.9 合并分集的原理

图 13.10 比较了接收信号强度指示驱动选择式分集和最大比值合并的信噪比的统计特性。当 $N_r = 1$ 时, 没有采用分集, 两者自然没有区别。从图中可进一步看出, 对于最大比值合并和选择式分集, 信噪比分布的斜率是一样的, 但随着 N_r 的增加, 均值之间的差值就会增大。由

于选择式分集丢弃了 $N_r - 1$ 个信号副本 ($N_r - 1$ 的增加与 N_r 是一致的), 所以这种区别是很明显的。对于 $N_r = 3$, 两种分集之间仅仅相差大约 2 dB。

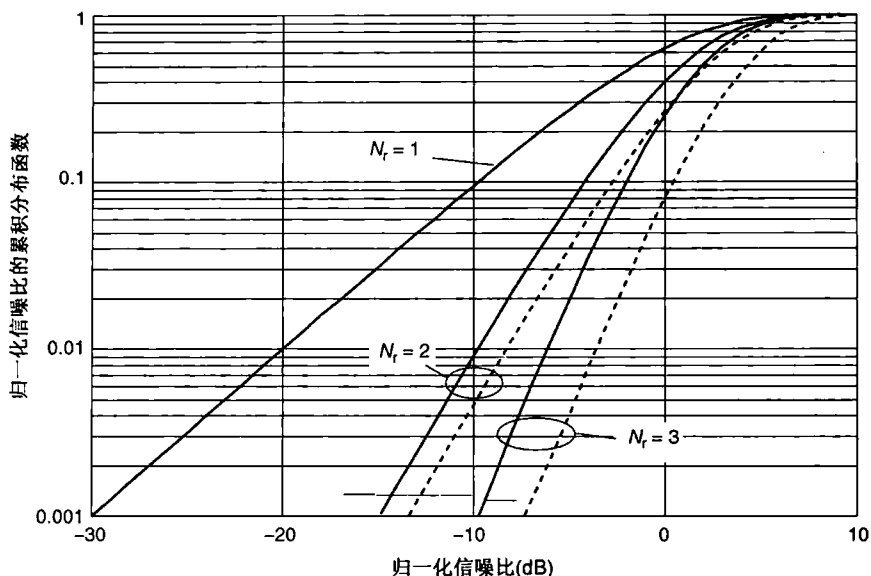


图 13.10 $N_r = 1, 2, 3$ 时, 归一化瞬时信噪比 $\gamma/\bar{\gamma}$ 的累积分布函数, 实线表示接收信号强度指示驱动选择式分集, 虚线表示最大比值合并。对于 $N_r = 1$, 两者无区别

等增益合并

对于等增益合并, 合并器的输出信噪比为

$$\gamma_{\text{ECG}} = \frac{\left(\sum_{n=1}^{N_r} \sqrt{\gamma_n} \right)^2}{N_r} \quad (13.20)$$

这里假设所有分集支路的噪声电平是相同的。如果所有支路都经历了平均信噪比为 $\bar{\gamma}$ 的瑞利衰落, 可证明合并器输出的平均信噪比为

$$\bar{\gamma}_{\text{ECG}} = \bar{\gamma} \left(1 + (N_r - 1) \frac{\pi}{4} \right) \quad (13.21)$$

这里只假设所有支路的平均信噪比相同, 而支路的瞬时信噪比 (表示不同的信道实现) 可以不同。值得注意的是, 等增益合并的性能只比最大比值合并差一个因子 $\pi/4$ (用平均信噪比)。当各支路的平均信噪比也不同时, 等增益合并和最大比值合并之间的性能差别会更大。

信噪比的概率密度函数可以由式 (13.20) 导出, 但是变得很复杂 (结果可在 Lee [1982] 中找到)。

最佳合并

在最大比值合并的推导中, 均假设信号只受加性高斯白噪声的干扰。如果干扰成为决定信号质量的因素, 那么最大比值合并不再是最佳方案。为了使信干噪比 (SINR) 有最大值, 这时权重需根据一种称为最佳合并的方案确定, 该方案最早在标志性论文 Winters [1984] 中推导出来。第一步先确定不同天线的干扰和噪声的相关矩阵:

$$\mathbf{R} = \sigma_n^2 \mathbf{I} + \sum_{k=1}^K E\{\mathbf{r}_k \mathbf{r}_k^\dagger\} \quad (13.22)$$

其中,期望是在信道保持不变的一段时间内的取值, $\mathbf{r}_k$ 是第 k 个干扰源的接收信号向量。此外,还需 N_r 个分集支路的复传输函数(由于假定信道平坦衰落,为复衰减);这些都包含在向量 $\mathbf{h}_d$ 中。包含最佳接收权重的向量为

$$\mathbf{w}_{\text{opt}} = \mathbf{R}^{-1} \mathbf{h}_d \quad (13.23)$$

当信道改变时,这些权重也要进行相应的调整。很容易看出,对于一个噪声受限的系统,相关矩阵就变成了(成比例的)单位矩阵,最佳合并就简化为最大比值合并。

更为有趣的是,最小化均方误差等同于最大化信干噪比,可以写成下式:

$$\text{SINR} = ((\mathbf{w}^\dagger \mathbf{h}_d \mathbf{h}_d^\dagger \mathbf{w}) / (\mathbf{w}^\dagger \mathbf{R} \mathbf{w})) \quad (13.24)$$

这个信干噪比值是归一化瑞利商,可以通过最大化下式中的最大归一特征向量对应的特征值得到:

$$\mathbf{h}_d \mathbf{h}_d^\dagger \mathbf{w} = \lambda \mathbf{R} \mathbf{w} \quad (13.25)$$

通过归一化特征值问题得到的权重向量与式(13.22)中得到的相同。

N_r 个分集支路信号的最佳合并能提供 N_r 个自由度。这可以消除 $N_r - 1$ 个干扰源的干扰,也就是说,能消除 $N_s \leq N_r - 1$ 个干扰,剩余 $N_r - N_s$ 个“正常”的天线可用来降低噪声。这看似简单的描述对无线系统的设计产生了巨大影响!如第3章所述,许多无线系统是干扰受限的。通过适当的分集合并,至少能消除部分干扰源的干扰,这极大地提高了此类系统的链路质量,或提高了系统的容量(详见20.1节)。

例 13.5 假设一个期望的二进制相移键控信号在频率平坦信道中传播, $\mathbf{h}_d = [1, 0.5 + 0.7j]^\top$, 有一个同步二进制相移键控干扰信号, $\mathbf{h}_{\text{int}} = [0.3, -0.2 + 1.7j]^\top$ 。噪声方差为 $\sigma_n^2 = 0.01$ 。试说明对接收信号用式(13.23)计算的权重进行加权就能够减轻干扰。

解:

令期望的发射信号为 $s_d(t)$, 干扰信号为 $s_{\text{int}}(t)$, $n_1(t)$ 和 $n_2(t)$ 是相互独立的零均值噪声。假设期望信号和干扰信号是不相关的,并且 $E\{s_d s_d^*\} = 1$, $E\{s_{\text{int}} s_{\text{int}}^*\} = 1$ 。

根据式(13.22),计算噪声加干扰的相关矩阵为

$$\mathbf{R} = \sigma_n^2 \mathbf{I} + \sum_{k=1}^K E\{\mathbf{r}_k \mathbf{r}_k^\dagger\} \quad (13.26)$$

$$= 0.01 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0.3 \\ -0.2 + 1.7j \end{pmatrix} (0.3 \quad -0.2 - 1.7j) \quad (13.27)$$

$$= \begin{pmatrix} 0.1 & -0.06 - 0.51j \\ -0.06 + 0.51j & 2.94 \end{pmatrix} \quad (13.28)$$

利用式(13.23),并对权重进行归一化,可得

$$\mathbf{w} = \frac{\mathbf{R}^{-1} \mathbf{h}_d}{|\mathbf{R}^{-1} \mathbf{h}_d|} \quad (13.29)$$

$$= \begin{pmatrix} 0.979 + 0.111j \\ 0.041 - 0.165j \end{pmatrix} \quad (13.30)$$

把这个值代入式(13.24)中的 $\mathbf{w}$,可以得到 SINR 值为 78。这个值可以与 $\text{SNR} = 100$ 相比较,这个值在接收端不需要抑制干扰时可以得到,而在单天线时 $\text{SINR} = 0.6$ 。

混合选择/最大比值合并

选择式分集和全信号合并之间折中的方案是所谓的混合选择方案,即先选择 N_r 个天线信号中最佳的 L 个,然后进行下变频和处理。这使所需的射频链路由 N_r 个减少到 L 个,这样可以大大节约成本。与全信号合并系统相比,这种节约是以(通常很小)性能降低为代价的。这种方法称为混合选择/最大比值合并(H-S/MRC),有时也称为广义选择合并(GSC)。

众所周知,最大比值合并的输出信噪比是各个不同接收天线的信噪比之和。对于 H-S/MRC,瞬时输出信噪比与最大比值合并看似很像,即

$$\gamma_{\text{H-S/MRC}} = \sum_{n=1}^L \gamma_{(n)} \quad (13.31)$$

与最大比值合并的主要不同是 $\gamma_{(n)}$ 为有序的信噪比,即 $\gamma_{(1)} > \gamma_{(2)} > \cdots > \gamma_{(N_r)}$ 。这会导致不同的性能,并对性能分析提出了新的数学挑战。特别是需引入顺序统计的概念。选择式分集(只选择 N_r 个天线之一)和最大比值合并分别是 H-S/MRC 对应于 $L=1$ 和 $L=N_r$ 的特殊情况。

由于 H-S/MRC 方案选择最好的天线支路用于合并,因而它能提供好的分集增益。实际上,这种方案得到的分集阶数与 N_r 成比例,而不是与 L 成比例。然而,这并不能提供全部的波束形成增益。如果所有天线的信号完全相关,那么 H-S/MRC 的信噪比增益只有 L ,而最大比值合并方案的增益为 N_r 。

H-S/MRC 的分析是基于已选的有序支路的,但由于有序支路的信噪比的统计值不是相互独立的,这就显得很复杂。但是,通过把有序的支路变量转换为一个新的随机变量集合,就可以简化这个问题的复杂性。可以找到一种能产生独立分布的随机变量(称为虚支路变量)的变换,使合并器的输出信噪比可以采用独立同分布(iid)的虚支路变量来表示,这会大大简化系统性能的分析。例如,对于非编码 H-S/MRC 系统,符号差错概率(SEP)的推导通常需要计算 N 重积分,实质上可简化为计算有限区间的单重积分。

对于 H-S/MRC,输出信噪比的均值和方差分别为

$$\bar{\gamma}_{\text{H-S/MRC}} = L \left(1 + \sum_{n=L+1}^{N_r} \frac{1}{n} \right) \bar{\gamma} \quad (13.32)$$

和

$$\sigma_{\text{H-S/MRC}}^2 = L \left(1 + L \sum_{n=L+1}^{N_r} \frac{1}{n^2} \right) \bar{\gamma}^2 \quad (13.33)$$

13.5 在衰落信道中分集接收的差错概率

这一节计算接收机采用分集接收时,在衰落信道中的符号差错率(SER)。从平坦衰落信道开始,计算接收功率和误比特率的统计值,然后再研究色散信道,分析分集如何减轻色散信道对简单接收机的有害影响。

13.5.1 平坦衰落信道中的差错概率

经典计算方法

与第 12 章类似,可以由信噪比的分布通过求平均条件差错概率(在一定信噪比条件下)来计算分集系统的差错概率:

$$\overline{\text{SER}} = \int_0^{\infty} \text{pdf}_{\gamma}(\gamma) \text{SER}(\gamma) d\gamma \quad (13.34)$$

下面以计算有 N_r 个分集支路最大比值合并的二进制相移键控的性能为例, 在加性高斯白噪声信道中, 二进制相移键控的符号差错率为(见第 12 章):

$$\text{SER}(\gamma) = Q(\sqrt{2\gamma}) \quad (13.35)$$

将此原理用于最大比值合并的情形。把式(13.18)和式(13.35)代入式(13.34), 可得到一个用于计算分析的式子:

$$\overline{\text{SER}} = \left(\frac{1-b}{2}\right)^{N_r} \sum_{n=0}^{N_r-1} \binom{N_r-1+n}{n} \left(\frac{1+b}{2}\right)^n \quad (13.36)$$

其中, b 定义为

$$b = \sqrt{\frac{\bar{\gamma}}{1+\bar{\gamma}}} \quad (13.37)$$

对于较大的 $\bar{\gamma}$, 可以近似为

$$\overline{\text{SER}} = \left(\frac{1}{4\bar{\gamma}}\right)^{N_r} \binom{2N_r-1}{N_r} \quad (13.38)$$

由此可以看出, 误比特率(系统有 N_r 个分集天线)随着信噪比的 N_r 次方增加而下降。

通过矩生成函数计算

前一节利用 Q 函数的典型表示, 利用信噪比的分布对误比特率求平均。第 12 章已经介绍过, Q 函数还有另一种定义, 可以很容易与信噪比的矩生成函数 $M_{\gamma}(s)$ 联合使用。在给定的信噪比条件下, 符号差错率可表示为(见第 12 章):

$$\text{SER}(\gamma) = \int_{\theta_1}^{\theta_2} f_1(\theta) \exp(-\gamma_{\text{MRC}} f_2(\theta)) d\theta \quad (13.39)$$

由于

$$\gamma_{\text{MRC}} = \sum_{n=1}^{N_r} \gamma_n \quad (13.40)$$

符号差错率可写为

$$\text{SER}(\gamma) = \int_{\theta_1}^{\theta_2} f_1(\theta) \prod_{n=1}^{N_r} \exp(-\gamma_n f_2(\theta)) d\theta \quad (13.41)$$

通过对不同支路的信噪比求平均, 可得

$$\overline{\text{SER}} = \int d\gamma_1 \text{pdf}_{\gamma_1}(\gamma_1) \int d\gamma_2 \text{pdf}_{\gamma_2}(\gamma_2) \cdots \int d\gamma_{N_r} \text{pdf}_{\gamma_{N_r}}(\gamma_{N_r}) \int_{\theta_1}^{\theta_2} d\theta f_1(\theta) \prod_{n=1}^{N_r} \exp(-\gamma_n f_2(\theta)) \quad (13.42)$$

$$= \int_{\theta_1}^{\theta_2} d\theta f_1(\theta) \prod_{n=1}^{N_r} \int d\gamma_n \text{pdf}_{\gamma_n}(\gamma_n) \exp(-\gamma_n f_2(\theta)) \quad (13.43)$$

$$= \int_{\theta_1}^{\theta_2} d\theta f_1(\theta) \prod_{n=1}^{N_r} M_{\gamma}(-f_2(\theta)) \quad (13.44)$$

$$= \int_{\theta_1}^{\theta_2} d\theta f_1(\theta) [M_{\gamma}(-f_2(\theta))]^{N_r} \quad (13.45)$$

由此,可写出瑞利衰落信道中二进制相移键控的差错率:

$$\overline{\text{SER}} = \frac{1}{\pi} \int_0^{\pi/2} \left[\frac{\sin^2(\theta)}{\sin^2(\theta) + \bar{\gamma}} \right]^{N_r} d\theta \quad (13.46)$$

例 13.6 8-PSK 系统有 4 个天线,信噪比为 10 dB,比较 H-S/MRC 中, $L=1, 2, 4$ 时的符号差错率。

解:

对于 H-S/MRC, M 进制相移键控的符号差错率为

$$\overline{\text{SER}}_{e, \text{H-S/MRC}}^{\text{MPSK}} = \frac{1}{\pi} \int_0^{\pi(M-1)/M} \left[\frac{\sin^2 \theta}{\sin^2(\pi/M)\bar{\gamma} + \sin^2 \theta} \right]^L \prod_{n=L+1}^{N_r} \left[\frac{\sin^2 \theta}{\sin^2(\pi/M)\bar{\gamma} \frac{L}{n} + \sin^2 \theta} \right] d\theta \quad (13.47)$$

把 $M=8$, $\bar{\gamma}=10$ dB, $N_r=4$, $L=1, 2, 4$ 分别代入式(13.47), 可得

L	$\overline{\text{SER}}_{e, \text{H-S/MRC}}^{\text{MPSK}}$
1	0.0442
2	0.0168
4	0.0090

13.5.2 频率选择性衰落信道中的符号差错率

现在研究在受时间色散和频率色散影响信道中的符号差错率。假设频移键控用差分相位检测,分析时利用了第 12 章讨论过的两个采样时刻信号之间的相关系数 ρ_{XY} 。

对于采用选择式分集的二进制频移键控信号:

$$\overline{\text{SER}} = \frac{1}{2} - \frac{1}{2} \sum_{n=1}^{N_r} \binom{N_r}{n} (-1)^{n+1} \frac{b_0 \text{Im}\{\rho_{XY}\}}{\sqrt{(\text{Im}\{\rho_{XY}\})^2 + n(1-|\rho_{XY}|^2)}} \quad (13.48)$$

其中, b_0 为发射比特。这可近似为

$$\overline{\text{SER}} = \frac{(2N_r - 1)!!}{2} \left(\frac{1 - |\rho_{XY}|^2}{2(\text{Im}\{\rho_{XY}\})^2} \right)^{N_r} \quad (13.49)$$

其中, $(2N_r - 1)!! = 1 \cdot 3 \cdot 5 \cdots (2N_r - 1)$ 。

对于采用最大比值合并的二进制频移键控:

$$\overline{\text{SER}} = \frac{1}{2} - \frac{1}{2} \frac{b_0 \text{Im}\{\rho_{XY}\}}{\sqrt{1 - (\text{Re}\{\rho_{XY}\})^2}} \sum_{n=0}^{N_r-1} \frac{(2n-1)!!}{(2n)!!} \left(1 - \frac{(\text{Im}\{\rho_{XY}\})^2}{1 - (\text{Re}\{\rho_{XY}\})^2} \right)^n \quad (13.50)$$

可近似为

$$\overline{\text{SER}} = \frac{(2N_r - 1)!!}{2(N_r!) } \left(\frac{1 - |\rho_{XY}|^2}{2(\text{Im}\{\rho_{XY}\})^2} \right)^{N_r} \quad (13.51)$$

上式说明,与选择式分集相比,最大比值合并将符号差错率提高了 $N_r!$ 倍。一个更重要的结论是,由时延色散和随机频率调制引起的误差和由噪声引起的误差同样被降低了。圆括号内表达式的 N_r 次方就说明了这个问题。这些项把误差都归因于各种不同的影响。分集时的符号差错率大约是无分集时符号差错率的 N_r 次方(见图 13.11 至图 13.13)。

对于采用选择式分集的差分正交相移键控,平均误比特率为

$$\overline{\text{BER}} = \frac{1}{2} - \frac{1}{4} \sum_{n=1}^{N_r} \binom{N_r}{n} (-1)^{n+1} \left[\frac{b_0 \text{Re}\{\rho_{XY}\}}{\sqrt{(\text{Re}\{\rho_{XY}\})^2 + n(1-|\rho_{XY}|^2)}} + \frac{b'_0 \text{Im}\{\rho_{XY}\}}{\sqrt{(\text{Im}\{\rho_{XY}\})^2 + n(1-|\rho_{XY}|^2)}} \right] \quad (13.52)$$

其中, b_0 和 b'_0 为发射符号。

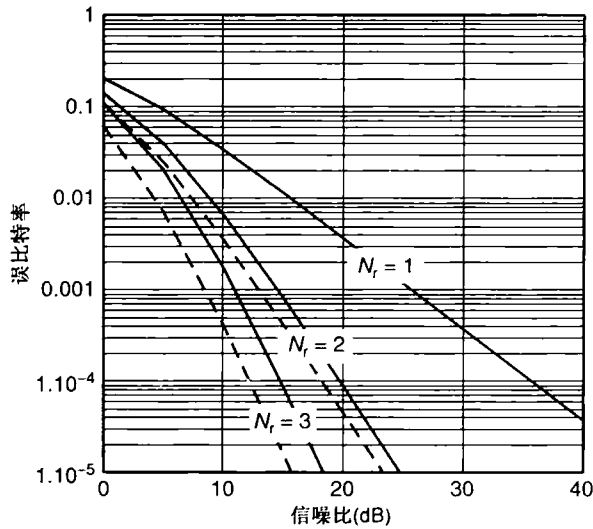


图 13.11 采用 N_r 个分集天线时, 最小频移键控的误比特率与信噪比的函数关系, 其中实线表示接收信号强度指示驱动选择式分集, 虚线表示最大比值合并。引自 Molisch[2000] © Prentice Hall

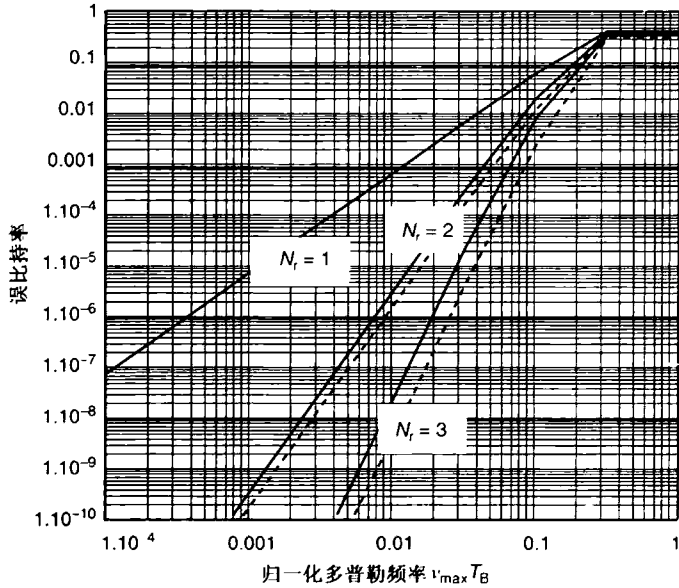


图 13.12 使用 N_r 个分集天线时, 最小频移键控的误比特率与归一化多普勒频率的函数关系, 其中实线表示接收信号强度指示驱动选择式分集, 虚线表示最大比值合并。引自 Molisch[2000] © Prentice Hall

对于采用最大比值合并的差分正交相移键控, 平均误比特率为

$$\begin{aligned} \overline{\text{BER}} = & \frac{1}{2} - \frac{1}{4} \sum_{n=0}^{N_r-1} \frac{(2n-1)!!}{(2n)!!} \left[\frac{b_0 \text{Re}\{\rho_{XY}\}}{\sqrt{1-(\text{Im}\{\rho_{XY}\})^2}} \left(\frac{1-|\rho_{XY}|^2}{1-(\text{Im}\{\rho_{XY}\})^2} \right)^n \right. \\ & \left. + \frac{b'_0 \text{Im}\{\rho_{XY}\}}{\sqrt{1-(\text{Re}\{\rho_{XY}\})^2}} \left(\frac{1-|\rho_{XY}|^2}{1-(\text{Re}\{\rho_{XY}\})^2} \right)^n \right] \end{aligned} \tag{13.53}$$

更一般的情况, 可采用二次型高斯变量(QFGV)方法(见 12.3.3 节)处理, 其中式(12.103)可由下式替换:

$$\begin{aligned}
 P(D < 0) = & Q_M(p_1, p_2) - I_0(p_1 p_2) \exp\left[-\frac{1}{2}(p_1^2 + p_2^2)\right] \\
 & + \frac{I_0(p_1 p_2) \exp\left[-\frac{1}{2}(p_1^2 + p_2^2)\right]}{(1 + v_2/v_1)^{2N_r-1}} \sum_{n=0}^{N_r-1} \binom{2N_r-1}{n} \left(\frac{v_2}{v_1}\right)^n \\
 & + \frac{\exp\left[-\frac{1}{2}(p_1^2 + p_2^2)\right]}{(1 + v_2/v_1)^{2N_r-1}} \cdot \sum_{n=1}^{N_r-1} I_n(p_1 p_2) \sum_{k=0}^{N_r-1-n} \binom{2N_r-1}{k} \\
 & \times \left[\left(\frac{p_2}{p_1}\right)^n \left(\frac{v_2}{v_1}\right)^k - \left(\frac{p_1}{p_2}\right)^n \left(\frac{v_2}{v_1}\right)^{2N_r-1-k} \right]
 \end{aligned} \quad (13.54)$$

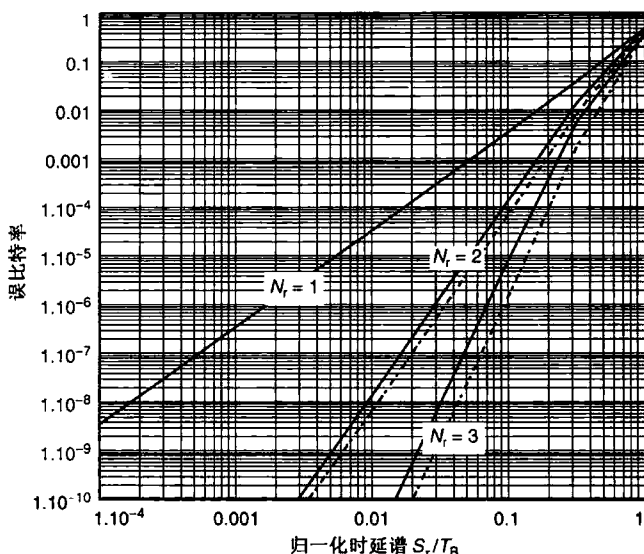


图 13.13 采用 N_r 个分集天线时, 最小频移键控的误比特率与归一化均方根时延扩展的函数关系, 其中实线表示接收信号强度指示驱动选择式分集, 虚线表示最大比值合并。引自 Molisch[2000]© Prentice Hall

当误差主要由频率选择性和时间选择性引起时, 接收信号强度指示驱动选择式分集不是最佳的选择策略。因为它只关注幅度最大的信号, 而不是失真最小的信号^①。在这些情况下, 误比特率驱动选择式分集更可取。对于 $N_r = 2$, 差分检测的最小频移键控的误比特率为

$$\overline{\text{BER}} \approx \left(\frac{\pi}{4}\right)^4 \left(\frac{S_r}{T_B}\right)^4 \quad (13.55)$$

接收信号强度指示驱动选择式分集的误比特率为

$$\overline{\text{BER}} \approx 3 \left(\frac{\pi}{4}\right)^4 \left(\frac{S_r}{T_B}\right)^4 \quad (13.56)$$

13.6 发射分集

对于许多情况, 只在链路的一端(通常在基站)安装多个天线。对于从移动台到基站的上行链路传输, 多个天线相当于接收分集支路。对于下行链路, 任意可能的分集都源于发射机。

① 由于类似的原因, 最大比值合并不是最佳合并方案。

本节将讨论从多个发射机天线发射信号和获得分集的方法。本节专门讨论天线分集(包括空间分集、模式分集和极化分集)。时间分集和天线分集是发射机固有的,所以在此不必再讨论。

关于多个天线的发射分集如何在接收机合并的问题将在第 20 章中讨论。

13.6.1 具有信道状态信息的发射机分集

我们分析的第一种情况是发射机完全知道信道信息。这些信息可从接收机的反馈或互易原理中获得;关于此问题的更详细讨论见 20.1.6 节。在这种情况下,可以发现发射分集和接收分集之间是完全等价的(至少对于噪声受限的情况)。换句话说,最佳发射方案根据从发射天线到单个接收天线之间的信道传输函数的复共轭,对从不同天线发射的信号进行线性加权。这种方法称为最大比值发射。

13.6.2 无信道状态信息的发射机分集

在许多情况下,信道状态信息(CSI)在发射机中是不可知的,因而不能简单地用不同发射天线发射同一信号的加权副本,因为不知道这些副本在接收机中是如何叠加的。不同分量的叠加是增强还是减弱是等可能的,换句话说,我们只是将具有不同随机相位的多径分量叠加起来,这就会产生瑞利衰落。因此,就不会得到任何分集(或波束成形)。

为了获得益处,不同天线的发射信号应该使接收机能区分出不同的发射信号分量。一种方式是延迟分集。在这种方案中,不同天线的发射信号是不同延迟的信号副本。这样,即使信道本身是平坦衰落的,也可以确保有效脉冲响应是时延色散的。因而在平坦衰落信道中,每个发射天线经过一个符号持续时间的时延(与前面的天线有关)再发射数据流。信道的有效脉冲响应为

$$h(\tau) = \frac{1}{\sqrt{N_r}} \sum_{n=1}^{N_t} h_n \delta(\tau - nT_s) \quad (13.57)$$

其中, h_n 为从第 n 个发射天线到接收天线之间的衰减,脉冲响应已被归一化,因此总的发射功率与天线的数量无关。从不同发射天线到接收机的信号相当于经时延的多径分量。如果天线之间的距离足够远,这些系数就具有独立衰落特性。对于时延色散信道,采用恰当的接收机,例如第 16 章介绍的均衡器,或第 18 章介绍的 Rake 接收机,就能得到与天线数量相等的分集阶数。

如果从单个发射天线到接收机之间的信道已经是时延色散的,那么这种方案仍然有效,但为不同的天线选择时延的时候应该注意,不同天线的发射信号之间的时延应至少与信道的最大附加时延一样大。

另一种方法是相位扫描分集。当只有两个天线时,这种方法非常有用,两个天线发射相同的信号。但是,其中一个天线的信号需要经历时变的相移。这意味着,在接收机接收信号时需要以时变的方式叠加;也就是说,我们在信道中人为地引入了时变特性。这样做的原因是,即使发射机、接收机和相互作用体是静止的,在深度衰落中信号也不是保持不变的。如果此方案与合适的编码或交织相结合,性能就会提高。

另外一种实现发射分集的方法是空时编码。这种方法将在第 20 章中进行讨论。

13.7 附录

请访问 www.wiley.com/go/molisch。

深入阅读

空间(天线)分集是一种最早的分集方式,在有关无线传输信道的书中都进行了讨论,如 Vaughan and Anderson [2003]。对于不同的角度谱,天线相关系数的估计可在 Fuhl et al. [2001] 及 Roy and Fortier [2004] 中找到。关于手机天线的相应情况, Ogawa et al. [2001] 给出了结果。Taga [2004] 也介绍了相互耦合对模式分集和相关系数的影响,在 MIMO 系统中的相关内容有一系列相关的论文 [Waldschmidt et al. 2004, Wallace and Jensen 2004]。Narayanan et al. [2004] 和 Shafi et al. [2006] 对极化分集做了更详细的讨论; Dietrich et al. [2001] 介绍了联合天线分集、极化分集及模式分集(角度分集)。Proakis [2005] 和许多基础文献中介绍了天线信号的不同合并策略,以及引起的信道统计特性。这种情况与第 12 章提到的是类似的:每一种衰落特性和每一种合并策略都至少有一篇论文。在利用这些衰落特性来计算误比特率方面,已经发表了大量论文。Simon and Alouini [2004] 针对不同的调制方式、不同的衰落特性和天线合并策略介绍了一种较好的统一处理方法。对于瑞利衰落或莱斯衰落,可采用 Proakis [1968] 中的二次型高斯变量的经典方法; Adachi and Ohno [1991] 和 Adachi and Parsons [1989] 分别计算了差分正交相移键控和频移键控的性能。Win and Winters [1999] 介绍了对于 HS-MRC 的虚路径方法; Simon and Alouini [2004] 也研究了这种情况。有关有序统计的一般数学特性可在 David and Nagaraja [2003] 中找到。对于频率选择性信道中的系统,分集影响的讨论可在 Molisch [2000] 的第 13 章中找到。Lo [1999] 最早提出了最大比值发射。Winters [1994] 和 Wittneben [1993] 提出了延迟分集。多种发射分集技术的概述可在 Hottinen et al. [2003] 中找到。

第14章 信道编码

14.1 编码与信息论基础

14.1.1 编码历史与动机

第12章证明了在无线系统中的典型信噪比情况下,可以发生 10^{-2} 数量级的误比特率。高的误比特率大多是由多径传播效应引起的。先进的接收机结构有助于减小误比特率:分集技术可以抗衰落骤降,而均衡和Rake接收机(见第16章和第18章)可以改进频率选择性信道下的性能。但是,即使采用这些先进的接收机,对误比特率的降低可能仍然不够。数据通信经常需要误比特率达到 $10^{-6} \sim 10^{-9}$ 数量级。这么低的值只有采用数据的编码才能实现,即在传输中引入冗余。纠错码^①的作用是减少误比特率,或者说产生编码增益 G_c ,即与未编码系统相比,我们可以发送小 G_c dB的功率,达到目标误比特率^②。

编码的历史开始于克劳德·香农[Shannon 1948]关于通信的数学理论的开创性工作。他指出只要比特率小于信道容量,无误的传输数据是可能的。无误码可以通过使用“适当”的编码来实现。香农证明了(无限长的)随机码可以达到容量。遗憾的是,这种编码在实际中无法采用,因为其解码需要庞大的工作量。50多年来,编码理论家的主要工作就是寻找逼近香农限的实用码,即以接近信道容量的信息速率进行通信。

本章的后续部分会对纠错码进行简单的概述。基本的编码理论对时不变信道和时变信道均适用,所以14.2节至14.7节并不区分这两种情况,而是给出了很多重要类别的编码/译码的理论背景:分组码、卷积码、网格编码调制(TCM)、Turbo码及低密度奇偶校验(LDPC)码。14.8节和14.9节专门讲了衰落信道的特性,并描述了编码结构如何适应这种情况。

14.1.2 信息论的基础概念

香农开创性的工作是信息论的基础,探索了最佳通信系统理论上的性能限。信息理论限的知识对系统设计者是有用的,因为它标示着一个真实的系统有多大的提升空间。我们面对的核心概念是:(i)关于信息的数学意义上的严格定义;(ii)信道容量,它描述了在给定的信道条件下,信息最多能以多大的速率传输。我们将会发现通信需要合适的编码来趋近信道容量。

作为第一步,首先定义离散无记忆信道(DMC)。假设有一个发射符号的符号集合 $\mathcal{X}$ (集合的大小是 $|\mathcal{X}|$),集合中的每一个元素的发射概率是已知的。此外还有一个接收符号的符号集合 $\mathcal{Y}$,大小为 $|\mathcal{Y}|$ 。最后,也是非常重要的一点,离散无记忆信道定义了从每一个发射符号到每一个接收符号的转移概率。最常见的形式是二元对称信道,特征为 $|\mathcal{X}| = |\mathcal{Y}| = 2$,转移概率为 p 。例如,如果发送的是 $\mathcal{X} = +1$,那么接收符号 $\mathcal{Y} = +1$ 的概率为 $1 - p$,接收符号 $\mathcal{Y} = -1$

① 本章的余下部分使用编码表示纠错码。注意,这与第15章和第23章不同,那里编码将指信源编码。

② 注意,编码增益取决于目标差错率。编码导致误比特率-信噪比曲线形状的变化。

的概率为 p ; 类似地, 如果发送的是 $\mathcal{X} = -1$, 那么接收 -1 的概率为 p , 接收 $+1$ 的概率为 $1-p$ 。无记忆信道具有的另一个特征为, 如果发送符号序列 $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$, 观察输出符号序列 $\mathbf{y} = \{y_1, y_2, \dots, y_N\}$, 那么

$$\Pr(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^N \Pr(y_n|x_n) \quad (14.1)$$

换句话说, 没有符号间干扰(也称码间干扰)。

离散无记忆信道可以用来描述, 比如, 二进制相移键控调制器、加性高斯白噪声信道, 以及硬判决输出的解调器的串联。那么, 变量 X 对应于发送符号 $+1/-1$, 变量 Y 对应于解调器/判决器的输出, 也是 $+1/-1$ 。噪声的大小决定了符号差错概率(如第12章中计算得到), 在这种情况下等同于转移概率 p 。如果想要避免硬判决解调器的限制, 一个有用的信道模型是离散时间加性高斯白噪声信道(如第11章至第13章所采用的), 为

$$y_n = x_n + n_n \quad (14.2)$$

其中, n_n 是高斯分布随机变量, 方差为 σ^2 , 输入变量受限于平均功率约束: $E\{X^2\} \leq P$ ①。

现在定义两个离散随机变量 X, Y 之间的互信息。如果观察 Y 的某个特定实现, 比如 $Y = y$, 那么互信息是通过这次观察所知关于出现事件 $X = x$ 的信息的量度, 实现 x 和 y 之间的互信息定义为

$$I(x; y) = \log \frac{\Pr(x|y)}{\Pr(x)} \quad (14.3)$$

其中, $\Pr(x|y)$ 是 y 出现的条件下 x 的概率。上式中的对数既可以 2 为底, 这种情况下的互信息以比特为单位; 也可以 e 为底, 这种情况下的 I 的单位为奈特。随机变量 X 和 Y 之间的互信息是 $I(x; y)$ 的均值, 即

$$I(X; Y) = \sum_x \sum_y \Pr(X = x, Y = y) \log \frac{\Pr(x|y)}{\Pr(x)} \quad (14.4)$$

$$= \sum_x \sum_y \Pr(X = x, Y = y) \log \frac{\Pr(x, y)}{\Pr(x) \Pr(y)} \quad (14.5)$$

当 y 和/或 x 是连续变量时, 求和以积分代替。

对互信息的直观解释依据离散无记忆信道得出: 再次将 X, Y 分别与发送符号和接收符号相关联。如果 X 和 Y 是统计独立的, 也就是接收一个特定的符号并未给出有关发送哪个符号的任何信息, 那么互信息为 0; 这种情况发生在信噪比非常低的信道中, 在一个二元对称信道中, 这对应于 $p = 0.5$ 。如果接收到一个特定符号, 比如 $+1$, 以很大的概率告诉我们发送的是哪个符号, 那么互信息很大。这种情况发生在具有小的转移概率 p 的信道中, 或者换句话说, 在几乎没有噪声的信道中。注意, 互信息永远不可能大于 $\min[\log|\mathcal{X}|, \log|\mathcal{Y}|]$ 。如果 $X = Y$, 那么对于离散符号集

$$I(X; Y) = - \sum_x \sum_y \Pr(X = x, Y = y) \log[\Pr(x)] \quad (14.6)$$

$$= - \sum_x \Pr(X = x) \log[\Pr(x)] \quad (14.7)$$

① 注意 P 是功率, p 是转移概率。

被称为熵 $H(X)$ 。熵表示随机变量 X 的信息量,对于离散无记忆源,如果符号集中的所有符号等概,则熵取最大值。

一般的互信息也可以写成接收符号的熵 $H(Y)$ 和条件熵 $H(Y|X) = -\sum_{x,y} \Pr(x,y) \log[\Pr(y|x)]$ 的形式:

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad (14.8)$$

14.1.3 信道容量

对于任何经过离散无记忆信道的通信,以下定理都成立:可靠的通信(也就是差错概率趋于零的通信)是可能的,当且仅当通信速率 $R < C$, C 是信道容量

$$C = \max_{\Pr(x)} I(X;Y) \quad (14.9)$$

也就是 X 和 Y 的互信息在所有输入分布下的最大值。如果速率高于容量,则可靠的通信是不可能的。

到达或者至少趋近信道容量的关键是对长码字编码的使用。如同本章引言中提到的,编码引进了冗余,以便即使码字中的一些比特被错误地接收,仍有可能还原原始的码字。考虑长度为 N 的码字在二元对称信道的传输:当 $N \rightarrow \infty$, 接收约等于 Np 个错误的概率趋于 1。

$$\binom{N}{Np} = \frac{N!}{(Np)!(N(1-p))!} \quad (14.10)$$

$$\approx \frac{\sqrt{2\pi N} N^N e^{-N}}{[\sqrt{2\pi Np} (Np)^{Np} e^{-Np}] [\sqrt{2\pi N(1-p)} (N(1-p))^{N(1-p)} e^{-N(1-p)}]} \quad (14.11)$$

$$\approx \frac{1}{2^{N[p \log(p) + (1-p) \log(1-p)]}} \quad (14.12)$$

$$= 2^{N H_b(p)} \quad (14.13)$$

共有上式给出的这么多个不同的码字具有 Np 个错误;第一个近似依据 Sterling 准则得出,而第二个是用 $p = 2^{\log(p)}$ 重新排列所得。最后一个等号则简单地为二进制熵函数的定义, $H_b(p) = -p \log(p) - [(1-p) \log(1-p)]$ 。长度为 N 的所有序列的总数为 2^N 。那么可确切分辨的序列的数目为 $M = 2^{N(1-H_b(p))}$ 。所以,用 N 个符号可以传输 M 种不同的信息,进而有 $\log(M)$ 位信息比特。可能的码率为

$$R = \frac{\log(M)}{N} = 1 - H_b(p) \quad (14.14)$$

对于加性高斯白噪声信道,可以进行类似的证明。对于长的码字,可知以高概率有

$$\frac{1}{N} \sum_n |n_n|^2 \rightarrow \sigma^2 \quad (14.15)$$

于是接收的信号向量 $\mathbf{y}$ 位于一个围绕发射信号向量 $\mathbf{x}$ 的接近半径为 $\sqrt{N\sigma^2}$ 的球体(称为噪声球体)的表面。只要与不同的码字相关联的球体不相互重叠,即每个接收信号点可以唯一地对应于一个特定的发射信号点,可靠通信就是可能的。另一方面,由于平均功率 $\sqrt{N(P + \sigma^2)}$ 的限制,可知所有接收信号点必须位于一个半径为 $\sqrt{N(P + \sigma^2)}$ 的球体内部。可以得出结论,能够被可靠译码的不同接收序列的数目等于装进半径为 $\sqrt{N(P + \sigma^2)}$ 的球体

的噪声球体的个数。因为半径为 ρ 的 N 维球体体积与 ρ^N 成正比, 因此能够用长度为 N 的码字进行通信的不同信息的个数为

$$M = \frac{[N(P + \sigma^2)]^{N/2}}{[N\sigma^2]^{N/2}} \quad (14.16)$$

于是通信的可能速率为

$$R = \frac{\log(M)}{N} = \frac{1}{2} \log \left[1 + \frac{P}{\sigma^2} \right] \quad (14.17)$$

这等于加性高斯白噪声信道的容量。省却推导, 这里声明这个容量是在当传输符号集是高斯的(也是连续的)时得到的。

值得注意的是, 式(14.17)是对于一个实的调制符号表和信道的每次信道使用(也就是每个传输的基本符号)的容量。当使用复调制时, 每单位带宽的容量为

$$C_{\text{AWGN}} = 2 \times \frac{1}{2} \log \left[1 + \frac{P}{N_0 B} \right] \text{ bps/Hz} \quad (14.18)$$

它可以用信噪比的形式表示为

$$C_{\text{AWGN}} = \log [1 + \gamma] \text{ bps/Hz} \quad (14.19)$$

值得注意的是, 容量随信噪比仅以对数形式增长。这意味着对于加性高斯白噪声信道, 提高发射功率仅能获得有限的信道容量提升。还记得这一表述与第12章和第13章中所获得的见解是不同的吗? 在那里讲过对于一个给定的传输速率, 提高发射功率对于降低加性高斯白噪声信道中未编码传输的误码率是非常有效的。此处要说明, 对于一个理想的编码系统(根据定义, 没有符号错误), 提高发送功率对于提高传输速率不是一种有效途径。然而, 很清楚, 可接受的传输速率随着有效带宽的增加而线性提高。

14.1.4 功率与带宽的关系

从容量公式可知, 也可以在一个传输的频谱效率和功率效率之间取折中。已知发送速率必须小于归一化带宽与投入带宽之积

$$R < B \log \left[1 + \frac{P}{N_0 B} \right] \quad (14.20)$$

进一步, 每传输信息比特消耗的能量为

$$E_b = \frac{P T_{\text{seq}}}{\log(M)} = \frac{P}{\frac{\log(M)}{T_{\text{seq}}}} = \frac{P}{R} \quad (14.21)$$

其中, T_{seq} 是长度为 N 的符号序列的持续时间。 R/B 表示为频谱效率 r , 重写式(14.20)为

$$r < \log \left(1 + r \frac{E_b}{N_0} \right) \quad (14.22)$$

于是

$$\frac{E_b}{N_0} > \frac{2^r - 1}{r} \quad (14.23)$$

这个公式表示了每比特能量与频谱效率之间的折中。当谱效率降低时, 对 E_b/N_0 的要求也会降低。以频谱效率极限零为例, $E_b/N_0 \rightarrow \ln(2) = -1.59 \text{ dB}$, 由洛必达法则所得的这个 -1.59 dB 是

可能可靠通信的 E_b/N_0 的最小值。注意, 具有很大维数的正交调制, 如频移键控或者脉冲位置调制) 近似于这种情况。

14.1.5 与实际系统的关系

以上对信道容量的推导均使用无限长的码。而且, 以上产生的结果并没有给出如何得到具有上述特点的好码的构造性方法。香农的推导证明随机码能达到信道容量; 然而, 真正随机码的唯一译码方法是列表译码, 当 N 很大时, 列表很快变得难以操纵。下面将讨论的实际编码方法虽然性能差于随机码, 但是具有通过有限的代价而实际上可译码的好处。对于分组码和卷积码尤其如此, 它们已经被应用多年, 但是由于相对短的码长, 并不接近理论的性能限。20 世纪 90 年代终于有了实际上可译码, 同时又具有大的码字有效长度, 进而接近最佳性能的方法: Turbo 码和 LDPC 码距离香农限均不足 1 dB (见 14.6 节和 14.7 节)。

14.1.6 实际编码的分类

编码的一种分类方法是分为分组码和卷积码, 前者是在一组数据后添加冗余, 而后者则是连续地添加冗余。分组码可以很好地纠正突发错误, 这在无线通信中经常出现, 但是突发错误也可以通过交织技术转变成随机错误。卷积编码的优点是其解码通过维特比译码很容易实现, 用同样的算法还能进行联合译码和均衡。Turbo 编码和 LDPC 编码可以很容易地归入这种分类。在 14.6 节将会看到, Turbo 码使用了两个并行交织的卷积码对信息进行编码。但是, 由于编码器的记忆长度, 采用了不同的译码器结构。LDPC 编码也类似: 正如其名字所示, 它属于分组编码, 但是其分组太大, 所以需要不一样的译码结构。正因为这些原因, 将其单列一节进行讨论。

最近几年的研究实际上“模糊”了分组码和卷积码的界限。已经证明维特比译码器(卷积码的经典解决方案)可以用于检测分组码[Wolf 1978], 而置信传播算法[Loeliger 2004] (常用来对分组码进行译码)可以推广成对卷积码进行译码。然而, 这都属于比较复杂的论题, 这里不进行深入讨论, 感兴趣的读者可以参考所引论文。

编码的另一种分类方法基于其编码器输入/输出是硬信息还是软信息。硬信息只是表达对一个比特进行检测时的(二元的)值, 而软信息表达判决这个比特时的置信信息。

例 14.1 循环码的软/硬信息。

解:

考虑一个简单的例子: 重复码重复一个比特 3 次, 设发送比特序列为 1 1 1, 接收端的解调信号为 $-0.05, -0.1, 1.0$ 。现在, 一种方法是译码前对每个比特做硬判决, 判决后的比特序列为 $-1 -1 1$, 译码器根据这些硬译码器的输入, 由多数判决确定发送的是 -1 , 但是软译码器可能会将解调信号相加^①, 得到 0.85 作为软译码器输入, 对这个值做硬判决会得出发送的是 $+1$ 。

对于有些应用(如迭代和级联译码器), 译码器不仅使用软输入信息, 输出也会在硬比特信息之外, 输出软信息。无论如何, 与硬输入信息相比, 使用软输入信息总是会提高系统性能。

^① 这纯粹是为了演示目的。我们将在后面讨论更智能的软合并策略。

14.2 分组编码

14.2.1 引言

分组编码是把数据源进行分组, 根据分组内的比特值计算出一个更长的码字进行传输。码率(消息序列长度与码字长度之比)越低, 冗余度越高, 错误被纠正的概率也越高。最简单的码是分组长度为1的重复码: 输入“分组”为 x , 输出分组为 xxx (码率为 $1/3$ 的重复码)。

在这样的直观介绍之后, 下面给出更精确的描述。首先定义一些重要的术语和概念。

- 分组码。数据源每 K 个符号分成一组, 每个未编码的数据块对应一个长为 N 个符号的码字。
- 码率。比值 K/N 称为码率 R_c (假设编码前后符号集相同)。
- 二进制码。当符号集是二进制值时产生的是二进制码, 取值只能是“0”和“1”。几乎所有实用的分组码都是二进制码, RS码例外(见下文)。如无特别声明, 后面章节中提到的都是二进制码, 因此后面的“和”意味着“模2和”, “+”意味着“模2加”。
- 汉明距离。两个码字间的汉明距离 $d_H(\mathbf{x}, \mathbf{y})$ 是指它们之间不同的比特数:

$$d_H(\mathbf{x}, \mathbf{y}) = \sum_n |x_n - y_n| \quad (14.24)$$

其中 $\mathbf{x}$ 和 $\mathbf{y}$ 为码矢量, $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_N]$, $x_n \in \{0, 1\}$ 。例如, 码字01001011和11101011之间的汉明距离为2。注意, 在编码理论中, 经常用到的是行矢量, 而通信理论中经常用列矢量。为了简化对其他编码书籍的交叉参考, 本章中也遵循这一用法。

- 欧氏距离。两个码字之间的平方欧氏距离是指码矢量 $\mathbf{x}$ 和 $\mathbf{y}$ 之间的几何距离的平方和:

$$d_E^2(\mathbf{x}, \mathbf{y}) = \sum_n |x_n - y_n|^2 \quad (14.25)$$

- 最小距离。码字的最小距离 $d_{\min}$ 是指最小汉明距离(d_H), 这里的“最小”是对码中任意两个码字的所有可能组合求最小。注意, 这里的最小距离与奇偶校验矩阵(见后面)中线性独立的列数相等。
- 权重。码字的权重是指码字中“1”的个数, 例如码字01001011的权重为4。
- 系统码。在系统码中, 原始信息位明显地出现在固定位置, 校验(冗余)位由信息位计算出来, 在不同(也是固定)的位置。理想传输时(无噪声非失真信道), 不需要奇偶校验位的任何信息就可以确定码字。一个 $(7, 4)$ 系统分组码按如下方式生成:

k	k	k	k	m	m	m

其中, k 表示信息符号, m 表示奇偶校验符号。

- 线性码(群码)。对于这些码, 任意两个码字的和都是一个有效码字, 因此有如下重要性质:
 - (i) 全0码是一个有效码字;
 - (ii) 所有码字(全0码除外)的权重都大于或等于 $d_{\min}$;
 - (iii) 距离(即有效码字之间的汉明距离)的分布与码权重的分布相同;
 - (iv) 所有码字都可以由基本码字(生成码)的线性组合表示。

- 循环码(见 14.2.6 节)。循环码是一种特殊的线性码，其任一码字的循环移位依然是个有效码字，循环码可以由码矢量表示，也可以由码字多项式(多项式次数 $\leq N-1$ ， N 为码字长度)表示，其中非零系数对应码矢量中的非零元素， x 是虚拟变量。

例如，码字 011010 的两种表示分别为

$$\mathbf{x} = [0 \ 1 \ 1 \ 0 \ 1 \ 0]$$

(14.26)

$$X(x) = 0 \cdot x^5 + 1 \cdot x^4 + 1 \cdot x^3 + 0 \cdot x^2 + 1 \cdot x^1 + 0 \cdot x^0$$

(14.27)

- 伽罗华域。一个伽罗华域 $GF(p)$ 是有 p 个元素的一个有限域，其中 p 为素数，域中定义了元素之间的加法和乘法运算，且这些运算都具有封闭性(即两个元素的和与积仍是有效元素)；这两种运算具有幺元和逆元，并满足结合律、交换率和分配律。最重要的例子是 $GF(2)$ ，它包括元素 0 和 1，是最小的有限域，其加法和乘法表如下：

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

经常使用 $GF(2)$ 编码，是因为在计算机中很容易只用 1 比特表示，也可以定义扩域 $GF(p^m)$ ，其中 p 为素数， m 为任意整数。

- 本原多项式。定义 N 次不可约多项式为不能被任何次数小于 N 且大于 0 的多项式整除的多项式。如果一个 m 次多项式 $g(x)$ 是不可约的，且能被 $g(x)$ 整除的 $(x^N + 1)$ 中整数 N 的最小值为 $N = 2^m - 1$ ，则它被定义为本原多项式。

14.2.2 编码

最直接的编码是通过映射表：任意长为 K 的消息对应一个长为 N 的码字，表格只需要核对输入并输出相应的码字。但是，这种方法效率很低，因为它需要存储 2^K 个码字。

对于线性码，任意码字都可以通过其他码字的线性组合生成，这样只需存储码字的一个子集就足够了。例如，一个长为 K 的消息码， 2^K 个码字中只有 K 个是线性独立的，因此只需存储 K 个码字。最好选择前 K 个位置只有一个 1 的那些码字。这样的选择会自动生成系统码，则编码过程表示为矩阵相乘：

$$\mathbf{x} = \mathbf{uG}$$

(14.28)

其中， $\mathbf{x}$ 为 N 维码矢量， $\mathbf{u}$ 为 K 维信息矢量， $\mathbf{G}$ 为 $K \times N$ 维的生成矩阵。对于系统码，生成矩阵中左边 K 列为一个 $K \times K$ 的单位阵，右边 $N - K$ 列表示奇偶校验比特。 $\mathbf{x}$ 中开始的 K 比特与 $\mathbf{u}$ 相同。要注意的是，前面已经说过，我们使用行矢量来表示码矢量，将行矢量左乘矩阵得到该矢量与矩阵的乘积。

例 14.2 (7, 4) 汉明码的编码。

解：

为了解得更具体，来看一个(7, 4)码的例子。对信源码字[1 0 1 1]编码，生成矩阵为

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(14.29)

计算 $\mathbf{x} = \mathbf{uG}$ ：

$$\mathbf{x} = \begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (14.30)$$

则码字为

$$\mathbf{x} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \quad (14.31)$$

14.2.3 译码

为了判断接收码字是否为一个有效码字,可用奇偶校验矩阵 $\mathbf{H}$ 与码字相乘,得到一个 $N-K$ 维的伴随矢量 $\mathbf{s}_{\text{synd}}$ 。如果该矢量有全0项,则接收码字有效。

例 14.3 (7, 4) 汉明码的伴随式。

解:

利用上面给出的例子演示伴随式的计算过程。设接收到的比特序列(硬判决之后)为

$$\hat{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad (14.32)$$

例 14.2 中,码的奇偶校验矩阵为(下面将描述从 $\mathbf{G}$ 得到 $\mathbf{H}$ 的构造方法)

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \quad (14.33)$$

现在计算表达式 $\mathbf{s}_{\text{synd}} = \hat{\mathbf{x}}\mathbf{H}^T$ 。逐个写出表达式的各个组成部分,得到 3 个奇偶校验方程,对应 3 个奇偶校验比特:

$$\mathbf{s}_{\text{synd}}^T = \hat{\mathbf{x}}\mathbf{H}^T = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \quad (14.34)$$

这样,伴随式的计算可以解释如下:

- 接收到信息,即 1000;
- 计算出奇偶校验比特,即 110(校验位由接收到的系统比特计算得到);
- 接收到的奇偶校验比特为 101;
- 得到伴随式 011。

下面讨论如何求 $\mathbf{H}$ 矩阵。由于生成矩阵的每一行都是一个有效码字,与奇偶校验矩阵相乘的结果为 $\mathbf{0}$ ^①,所以关系式 $\mathbf{H} \cdot \mathbf{G}^T = \mathbf{0}$ 成立,将 $\mathbf{G}$ 表示成

$$\mathbf{G} = (\mathbf{I} \ \mathbf{P}) \quad (14.35)$$

关系 $\mathbf{H} \cdot \mathbf{G}^T = \mathbf{0}$ 可简化为

$$\mathbf{H} \cdot \mathbf{G}^T = (\mathbf{H}_1 \ \mathbf{H}_2) \begin{pmatrix} \mathbf{I} \\ \mathbf{P}^T \end{pmatrix} = (\mathbf{H}_1 + \mathbf{H}_2\mathbf{P}^T) = \mathbf{0} \quad (14.36)$$

如果现在选择 $\mathbf{H}_2 = \mathbf{I}$, $\mathbf{H}_1 = -\mathbf{P}^T$, 则上式完全满足:两个相同的矩阵相减结果为全0矩阵。注意,对每个生成矩阵可以存在不同的奇偶校验矩阵,上述构造性方法只给出一种可能方案。

^① 换句话说, $\mathbf{G}$ 是码空间, $\mathbf{H}$ 是对应的零空间。

例 14.4 计算汉明码的奇偶校验矩阵。

解：

(7, 4) 汉明码(事实上可为任何系统码)的奇偶校验矩阵的计算过程可描述如下。首先得到生成矩阵中最右边 $N - K$ 列：

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

(14.37)

并进行转置运算，注意在模 2 运算里加和减没有区别，然后添加一个 $(N - K) \times (N - K)$ 的单位矩阵得到 $\mathbf{H}$ ：

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

(14.38)

线性码中著名的一类是汉明码。汉明码可以简单地由其奇偶校验矩阵来定义。 $\mathbf{H}$ 中的各列包括了除全 0 码字外的 2^{N-K} 个所有可能的长度为 K 的位组合。因而 $\mathbf{H}$ 中所有的列是独特的。容易验证，式(14.38)中的奇偶校验矩阵满足该条件。汉明码的大小为 $(2^m - 1, 2^m - 1 - m)$ ，其中 m 为一个正整数。

14.2.4 差错识别和纠正

由于码的线性特性，每个接收码字都可以表示成码字 $\mathbf{x}$ 和错误图样 $\boldsymbol{\epsilon}$ 的和，这意味着伴随式仅与错误图样有关，而与发送码字无关。设接收矢量为 $\hat{\mathbf{x}}$ ，则在接收端计算伴随矢量得：

$$\mathbf{s}_{\text{synd}} = \hat{\mathbf{x}}\mathbf{H}^T = (\mathbf{x} + \boldsymbol{\epsilon})\mathbf{H}^T = \mathbf{x}\mathbf{H}^T + \boldsymbol{\epsilon}\mathbf{H}^T = \mathbf{0} + \boldsymbol{\epsilon}\mathbf{H}^T = \boldsymbol{\epsilon}\mathbf{H}^T$$

(14.39)

伴随式不为 0 表示发生了错误($\boldsymbol{\epsilon} \neq \mathbf{0}$)。译码就是需要找出“正确”的 $\boldsymbol{\epsilon}$ ，从 $\hat{\mathbf{x}}$ 中减去 $\boldsymbol{\epsilon}$ 即可得到正确的码字。然而，很显然相同的伴随式对应有很多组错误图样 $\boldsymbol{\epsilon}$ ，这些错误图样称为一个陪集。这样，目标就是找出与所得伴随式相对应的最有可能的 $\boldsymbol{\epsilon}$ 。在“合理的”信道中，一个比特正确到达的概率比错误到达的概率大，因此在给定陪集中具有最小权重的 $\boldsymbol{\epsilon}$ 就是最有可能的 $\boldsymbol{\epsilon}$ 。

另一种解释是估计可检测与可修正错误的个数，以码空间的表示(见图 14.1)开始：二进制 (N, K) 分组码中共有 2^K 个可能的比特组合，每一个对应码空间中的一个点。其中 2^K 种组合对应着有效码字，码字之间的最小距离为 $d_{\min}$ 。其他比特组合可以解释为位于有效码字周围的“校正球”内的点。校正球的最小尺寸 t 值决定了可纠正的错误个数。

对于小的分组码可以通过查表法进行译码，也就是说，所有可能的伴随式及其相应的最小权重错误图样都存储在要查找的表中。对于每个接收码字，利用这个表确定发送的码字。遗憾的是，小的分组码的性能往往很差。正是由于这个原因，可以用于更大码的译码技术得到很好的研究。最重要的一个是置信传播算法，在 14.7 节将会详细描述。而且，一些允许简化译码的特殊的线性码，如循环码，可能更优越。

综上所述，可得出以下结论：最小距离为 $d_{\min}$ 的码总是允许码字中检测到 $d_{\min} - 1$ 个错误。只有影响到 $d_{\min}$ 比特的错误才会导致另一个有效码字。或者，这种码字可以纠正 $\lfloor (d_{\min} - 1)/2 \rfloor$ 个错误，也就是说，接收到的比特序列必须位于真实码字的校正球内($\lfloor x \rfloor$ 表示小于 x 的最大整数)。如果所有的比特组合可以被唯一划入一个校正球内，则称该码是完备(perfect)的。

对于一个汉明码, 单个错误的位置可被唯一地确定。由于奇偶校验矩阵的所有列互不相同且线性独立, 并且汉明码可以正好纠正一个错误, 所以伴随式表明了错误的位置。

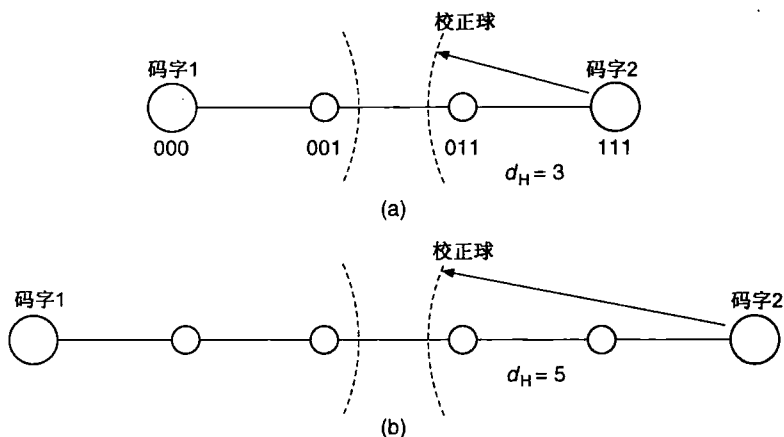


图 14.1 $d_H = 3$ 的码字(a)码空间和 $d_H = 5$ 的码字(b)码空间切分及校正球略图

14.2.5 级联码

使用两个码可以增强差错防护能力。一个是所谓的内编码, 按平常的方法保护数据, 如前面所述。但是仍然会存在一些错误(当接收码位于错误的校正球内时)。可以把信道和内编码结合起来成为超信道, 它比原来的信道有更低的误比特率, 外编码为这些错误提供保护(见图 14.2)。超信道的错误往往是突发产生的; 如果内编码是 (N, K) 分组码, 则突发的长度为 K 比特, 因此外编码一般是特别适合于抗突发错误的码。RS 码是这种应用中有效的码^①。

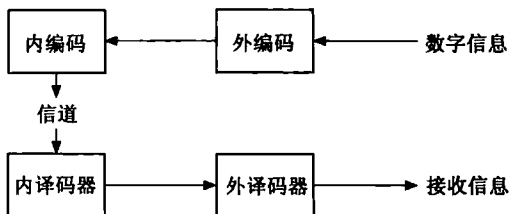


图 14.2 码的级联

将码进行恰当的级联是项相当困难的工作, 特别是两层码的相对性能需要仔细地权衡。内编码如果不够强就没有用, 而且甚至可能提高需要由外编码处理的误比特率。

14.3 卷积码

14.3.1 卷积码原理

卷积码并不会将(源)数据流分组, 而是以连续的方式添加冗余。卷积编码器包括一个有 L 个存储单元的移位寄存器和 N 个模 2 加法器(见图 14.3)。假设开始时明确定义了存储单元的状态, 即全为 0。当第一个数据比特进入编码器时, 它被放入移位寄存器的第一个存储单元(其他 0 值右移, 最右边的 0 移出寄存器)。由多路复用器读出所有加法器 $n = 1, 2, 3$ 的输出, 从而对每个输入比特可得到 3 个输出比特。然后下一个数据比特被放入寄存器(所有存储

^① 码级联也可以应用于内码为卷积码的情况, 见下文。

单元内的值右移一个单元), 加法器产生新的输出, 并由多路复用器读出。这个过程一直继续到最后一个源数据比特进入寄存器, 接着, 0 又作为寄存器的输入, 直到寄存器输出最后一个源数据比特, 存储单元又再次进入明确定义的(全零)状态。

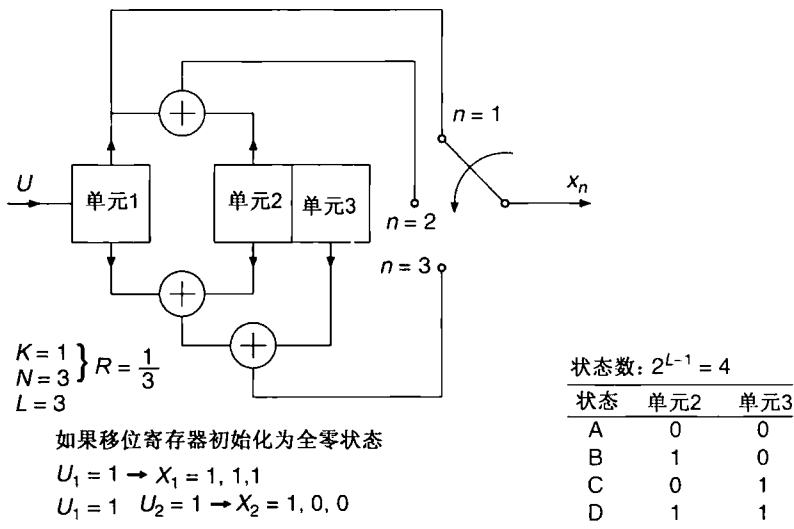


图 14.3 卷积码的例子。引自 Oehrvik [1994]© Ericsson AB

所以, 一个卷积编码器可以由一组移位寄存器和加法器表示, 加法器由其与存储单元的连接来表示。在图 14.3 的例子中, 只有元素 $l=1$ 与输出 $n=1$ 相连, 以使源数据比特直接映射到编码器输出; 第二个输出为存储单元 $l=1, 2$ 内的值相加; 第三个输出为 $l=1, 2, 3$ 存储单元内的值的组合。

编码器结构可以用不同方法表示, 一种方法是通过发生器序列, 产生 N 个长为 L 的矢量。如果第 l 个移位寄存器的元素与第 n 个加法器有连接, 则第 n 个矢量的第 l 个元素的值为 1, 否则为 0。发生器序列可以直接地解释编码器的构造过程。

对于译码器, 格图是一种更有用的描述方式。在该表示中, 编码器的状态由存储单元的内容表征, 格图表明输入了什么值使移位寄存器进入了什么状态, 结果又产生了什么输出值。作为例子, 图 14.4 所示为图 14.3 的卷积编码器的格图。只有单元 2, ..., L 的状态需要描述, 因为单元 1 的内容和输入(信息)比特相同。因为这个原因, 需要区分的状态数量为 $2^{L-1} = 4$ 。每个状态引出两条线, 上面的代表源数据比特为 0, 下面的代表源数据比特为 1, 每个状态直接进入所有其他状态是不可能的。例如, 状态 A 只能到达状态 A 或 B(但是不能到达 C 或 D)。译码过程中的冗余信息可以减少误码率。还可以看出, 格图是周期重复的, 因此没有必要画出无限长的格图, 尽管输入数据序列是无限长的。

如果接收到的数据序列与发送的数据相同, 译码就会很简单。知道了接收到的已编码序列后, 只需要沿着格图从它产生的地方追溯源数据即可。当有加性噪声时, 在接收比特序列中会存在误差, 需要从受干扰的接收数据中找出最有可能发散的序列, 即最大似然序列估计 (MLSE)。实际中, 这种估计经常用到维特比算法, 这将在下一节进行讨论。还有一些其他实现方法, 如 Fano 算法、栈式算法及反馈算法[Lin and Costello 2004, McEliece 2004]。

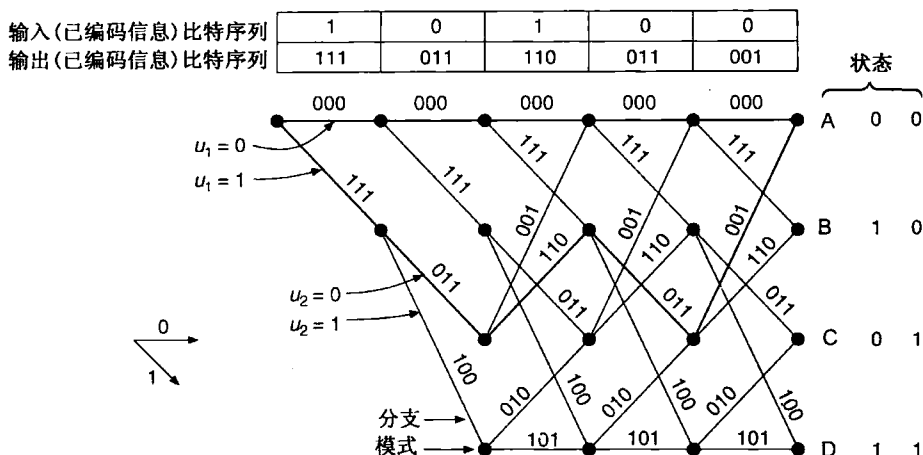


图 14.4 图 14.3 的卷积码的格图。引自 Oehrvik [1994] © Ericsson AB

14.3.2 维特比译码(经典表示)

维特比算法[Viterbi 1967]是最大似然序列估计的最常用的算法。算法的目的是在接收序列为 r 的情况下找到最有可能发送的序列 $\hat{s}$ ^①:

$$\hat{s} = \max_s \Pr(r|s) \quad (14.40)$$

即在所有可能的发送序列 s 范围内求最大值。如果噪声对接收符号的干扰统计独立^②, 则序列的概率 $\Pr(r|\hat{s})$ 可以分解为每个符号概率的乘积:

$$\hat{s} = \max_s \prod_i \Pr(r_i | s_i) \quad (14.41)$$

现在, 不求上式乘积的最大值, 而是求其对数的最大值也可以找到最佳序列, 因为对数运算是严格的单调函数:

$$\hat{s} = \max_s \sum_i \log[\Pr(r_i | s_i)] \quad (14.42)$$

对数转移函数 $\log[\Pr(r_i | s_i)]$ 也称为分支量度; 在这种情况下, 译码器仅输出“硬判决”(估计发送的是哪个编码比特), $r_i = \pm 1$, 分支量度为 r_i 与 s_i 之间的汉明距离。

最大似然序列估计确定格图上的所有可能路径(即所有可能的输入)的总量度 $\sum_i d_H(r_i - s_i)$, 最后选择具有最小量度的路径。这种优化过程需要很大的计算量, 因为可能路径的数目随输入比特数呈指数增加。维特比译码的核心思想是: 不是计算格图上所有可能路径的量度(自顶向下工作), 而是采取“自左向右”通过格图的方式工作。准确地说, 从移位寄存器的一组可能的状态(A_i, B_i, C_i, D_i , 其中 i 表示时刻或输入比特)出发。现在考虑所有进入状态 A 的路径(从左边)。对所有的可能路径 $s^{(1)}$, 如果在状态 A_i 遇到了比它具有更小量度的路径 $s^{(2)}$, 则将 $s^{(1)}$ 丢弃。路径经过格图上的同一状态时, 从后面状态的角度无法区分这些路径, 因此挑选一个具有最好性能的。同理, 选择经过状态 B_i, C_i 和 D_i 的最佳路径。确定了状态 i 的幸存路径

① 这是最大后验(MAP)检测的定义。但是, 对于等概率序列, 它等价于最大似然序列估计(见第12章)。

② 如果是有色噪声, 即不同样本之间相关, 接收机就必须采用所谓的“白化滤波器”(见第16章)。

后,继续进行到状态 $i + 1$ (或者进入状态组 $A_{i+1}, B_{i+1}, C_{i+1}, D_{i+1}$) 并重复该过程。最终,格图上的所有路径合并为一个定义明确的点,即全零状态^①。在这一点只有一条幸存路径,即概率最大的发送序列。

例 14.5 Viterbi 译码举例。

解:

图 14.5 给出了该算法的一个例子。格图的基本结构如图 14.5(a) 所示,要发送的比特序列如图 14.5(b) 所示。显示的量度是接收序列的汉明距离,接收序列为与格图中理论上可能的比特序列进行比较并做硬判决之后的序列。

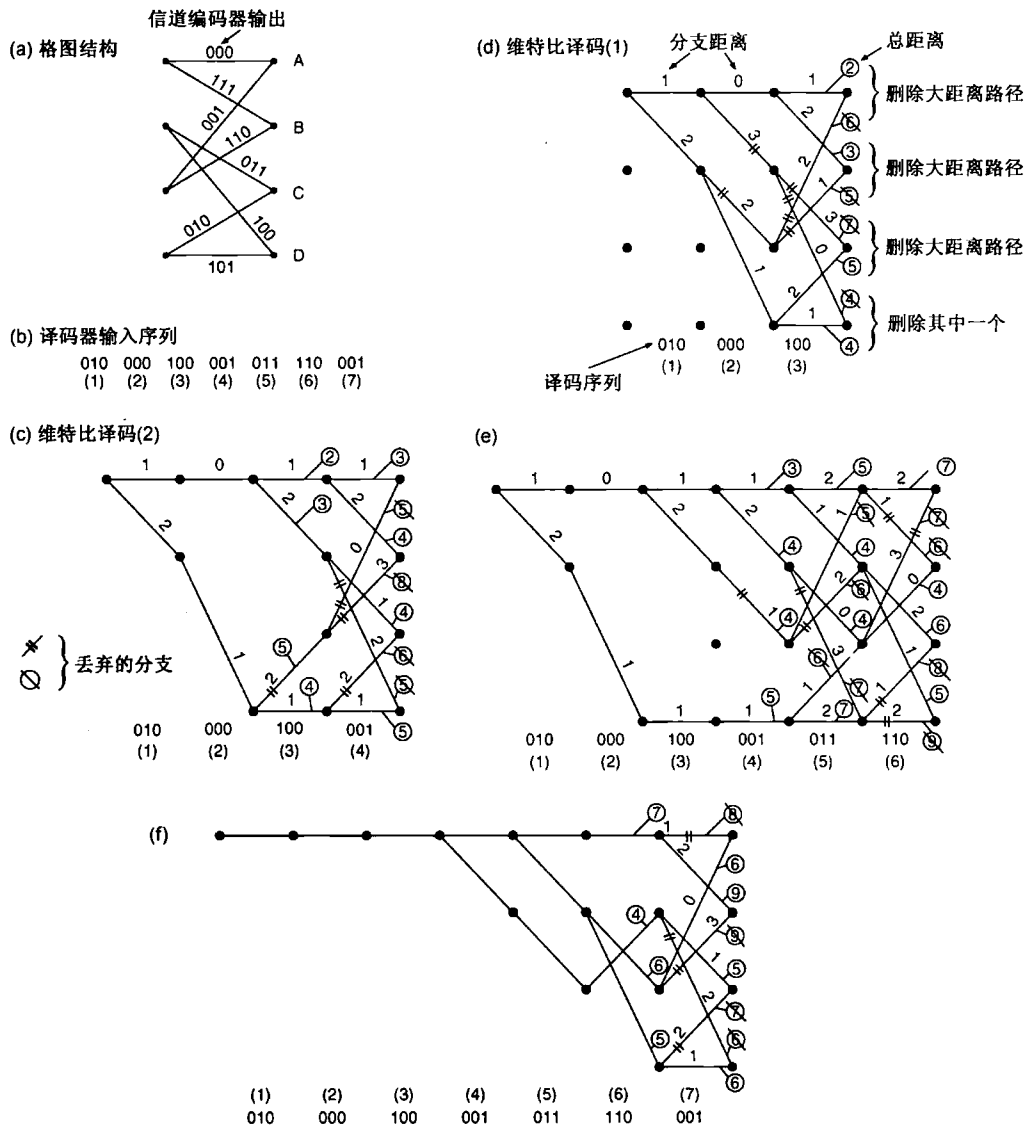


图 14.5 Viterbi 译码举例。引自 Oehrvik[1994]© Ericsson AB

① 实际上,这只发生在对卷积编码器的源数据序列添加了足够多的零(尾比特)以迫使编码器进入规定状态的情况下。如果不进行这种处理,译码器就不得不考虑所有的可能结束状态,并比较结束在每一个状态的路径的量度值。

假设开始时移位寄存器处于全零状态。图 14.5(c) 给出了最初 3 比特的格图, 从状态 A_0 到状态 A_4 有两种可能: 发送的源数据序列为 0, 0, 0 (相应地编码后的序列为 000, 000, 000, 经过状态 A_2 和 A_3); 或者发送的源数据序列为 100 (编码后的序列为 111, 011, 001, 通过状态 B_2 和 C_3)。前者中分支量度为 2, 后者为 6, 于是立刻排除了第二种可能性。同理, (根据最大似然度) 可发现从状态 A_0 到 B_4 是通过发送序列 0, 0, 1 产生的, 而不是 1, 1, 0。图 14.5 后面的子图给出了该过程是如何随输入比特而重复进行的。

通过排除非幸存路径, 维特比算法极大地降低了存储需求, 但该需求依然相当可观。因此要等到源数据的最后比特才来判断发送的是哪个序列显然是不合需求的。更合适的是, 该算法对“充分”过去的比特进行判决。更确切地说, 在考虑状态 A_i, B_i, C_i, D_i 期间, 对状态组 $A_{i-L_T}, B_{i-L_T}, C_{i-L_T}, D_{i-L_T}$ 的符号进行判决, 其中 L_T 为截尾长度。该原理如图 14.6 所示。位于长为 L_T 的窗中的数据被存储起来, 当移到格图中的下一组时, 最左边的状态组移出窗口, 这时需要做出发送的比特是哪一个的最后判决。

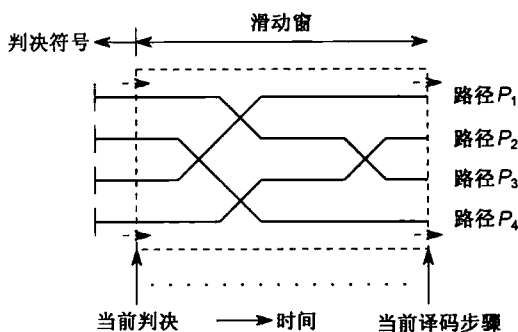


图 14.6 采用有限长度滑动窗的判决原理。引自 Mayr[1996] © B. Mayr

这种判决是依据当前观察状态(位于窗的最右边的状态)中包含最小量度路径的状态做出的。虽然这个过程是次优的, 但是明智地选择窗的长度可以使性能损失保持很小。在实际中, 长度

$$L_T = 6L \quad (14.43)$$

是个很好的折中。

14.3.3 维特比算法的改进

软译码

上面的例子中将接收符号与可能符号之间的汉明距离作为量度, 这意味着接收符号在判决过程之前首先要经过硬判决, 所以这种算法利用了码的冗余度(并不是所有的比特组合都是有效码字), 但是没有利用接收比特的可靠性知识。在有些情况下, 比特可能与判决边界很接近, 它们对最终所选序列的影响应该比离判决边界较远的时候小一些。在维特比算法中, 通过利用另一种量度, 可以将“软信息”考虑在内。可以证明, 在加性高斯白噪声信道和平坦衰落信道下, 它们与信号空间图中的欧氏距离 $d_i^2 = |r_i - s_i|^2$ 成正比。比例系数是多少并不重要, 因为它们并不影响寻找最佳量度:

$$\hat{s} = \min_s \sum_i |r_i - s_i|^2 \quad (14.44)$$

该式假设信道对所有接收符号的衰减相同, 且信道衰减在接收机中已得到补偿, 否则需要按下式求最小值:

$$\min_s \sum_i |r_i - \alpha_i s_i|^2 \quad (14.45)$$

其中, α_i 为第 i 个符号的衰减。第 16 章给出了利用软信息的维特比算法的具体例子(可看到最大似然序列估计均衡也能通过该算法实现)。

尾比特

正如前面提到的,如果编码器在发送数据序列的最后回到固定状态,经常是全零状态,那是最好不过的了。这样,选择哪条幸存路径是很明显的。图 14.7 表明这种方法大大减小了序列中最后部分比特的错误概率。该方法的缺点是损失了频谱效率,因为要传输不携带信息的符号(结尾添加的零)。这可能成为很短的分组持续时间系统中的一个问题。

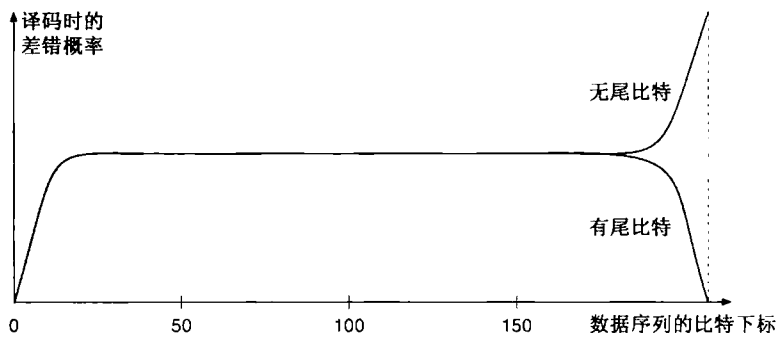


图 14.7 当采用尾比特时的差错概率。引自 Oehrvik [1994]© Ericsson AB

如果要传输的数据块很长,尾比特原理也可以用于块内:在预先定义的特定位置处发送一串零,从而迫使移位寄存器(及格图)进入一个固定状态。格图上的这种腰状结构减小了错误概率。

删余

一般的卷积编码只能达到一定的码率,如 $R_c = 1/2, 1/3, 1/4$ 。但是,在很多应用中,需要由信源速率和可用带宽决定的不同码率,例如, WiMedia 超宽带标准采用 $11/32$ 的码率作为默认模式。

调整码率的一种简单方法是对码进行删余。如果码率不够要求,则删除已编码序列中的某些比特。因为码中包含大量的冗余信息,所以这并不是个问题,这与实际中因衰落导致一些码比特不见了的情况类似。删除的比特在整个码字中只需有适当的分布即可,不要把某个比特的所有信息都删除掉。

14.4 网格编码调制

14.4.1 基本原理

编码的一个主要问题是降低了频谱效率。因为我们需要传输更多的比特,添加了校验位,带宽需求也更大了。这个问题可以通过使用在相同带宽内传输更多比特数的高阶调制方式来避免,也就是说,使用 $1/3$ 的码率,同时将调制方式从二进制相移键控改为 8-PSK,单位时间内传输的符号数是不变的。

正如第 12 章讨论过的,增加符号集会导致差错率增加;但是另一方面,引入编码可以减少差错概率。解决频谱效率问题的一种简单方法是给数据比特添加奇偶校验位,并将编码后的数据映射到高阶调制符号。但是,这往往得不到好的结果。不同的是,网格编码调制(TCM)通过增加信号空间的维数来添加码的冗余,在扩展的信号空间内不接受某些符号序列。这里最重要的方面是调制和编码作为一个联合过程来设计,这允许设计一种调制器加编码器,它在相同频谱效率的情况下比未编码系统有更好的抗噪声能力。

例 14.6 简单的网格编码调制。

解:

为了理解网格编码调制的基本原理,先看加性高斯白噪声信道的一个简单例子。比较一个未编码系统和网格编码调制系统,二者在每个符号持续时间内都传输 2 比特。在未编码系统中,使用的调制方式为正交相移键控(见图 14.8),每 2 比特的组合对应信号星座图中的一个点,每比特能量为 E_B ,星座图中点之间的距离为 $2\sqrt{E_B}$ 。这样,加性高斯白噪声信道下的误比特率可以表示为(见第 12 章)

$$\text{BER} \approx Q(\sqrt{2\gamma_B}) \quad (14.46)$$

对于网格编码调制,使用更大的符号集,即使用高阶调制方式(这里用 8-PSK)。由于每个符号可以传输 3 比特,所以码率为 $R_c = 2/3$ 。可用符号序列由移位寄存器的结构决定,与卷积码的情况是类似的。与任何其他卷积码一样,(编码过的)传输符号由当前比特及存储器的当前状态决定。在这个例子中,存储器有 2 比特,因此有 4 种可能的状态:00, 01, 10, 11。传输的是由状态和源比特决定的不同的 8-PSK 符号。

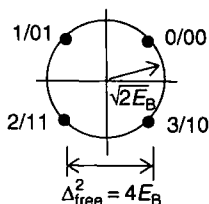


图 14.8 正交相移键控信号星座图

图 14.9 给出了可能的状态转移图。例如,若存储器在状态 11,要传输的信息符号为 01,则传输的 PSK 符号是 3,存储器结束于状态 10。从图中可以看到所谓的并行转移也是可能的。可以通过传输不同的 8-PSK 符号从状态 A 转到状态 B。但是,不是所有状态和 8-PSK 符号的组合都是可能的:当在状态 11 时,只有符号 1, 3, 5, 7 才有可能传输。图 14.10 是这个系统的编码器结构和网格图。

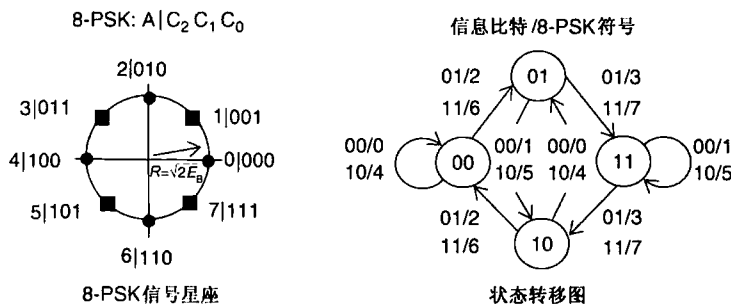


图 14.9 8-PSK 信号星座及 2/3 码率 8-PSK 网格编码调制的状态转移图。引自 Mayr[1996] © B. Mayr

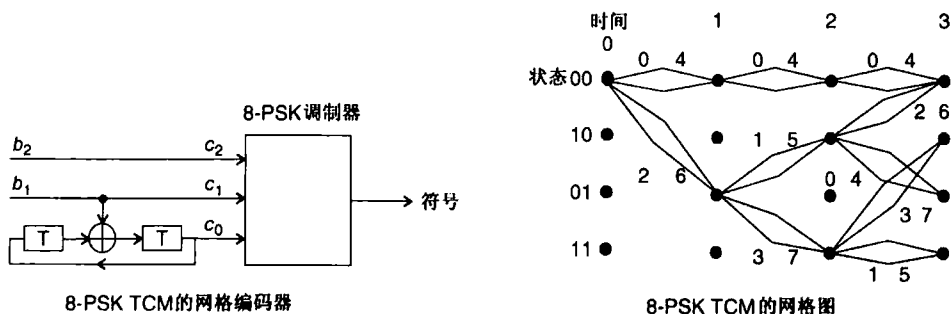


图 14.10 2/3 码率, 具有 4 个网格状态的 8-PSK 网格编码调制的编码器和网格图。引自 Mayr[1996] © B. Mayr

网格图也可以确定网格编码调制的误比特率。第一步,确定可能导致错误的符号序列之间的最小平方欧氏距离。看网格图的一部分,可以找到0时刻从00状态出发,在3时刻到达状态00的所有允许的路径。可能的第一对是使用PSK符号0和4的并行传输(从时刻0的00状态到时刻1的00状态;后续符号不需要不同),这两个符号之间的平方欧氏距离为 $d^2 = 8E_B$ 。网格中只向右移动一步的其他转移都是不可能的,而是需要整个序列。例如,通过传输PSK符号0-0-0或者2-1-6都可以从状态 $A_0(00)$ 转移到状态 $A_3(00)$ 。符号0与2的平方欧氏距离为 $d^2 = 4E_B$,0与1之间是 $2E_B[2\sin(\pi/8)]^2$,0与6之间也是 $4E_B$ 。两条路径之间的距离之幅度平方和为 $(4 + 1.17 + 4) = 9.17$ 。可能导致错误的其他序列也可用类似方法找到。可以发现,在同一状态开始和结束的序列之间的最小平方欧氏距离为 $d^2_{\text{coded}} = 8E_B$,这是未编码系统中该距离的两倍。

一般地,可以定义渐近编码增益,即高信噪比时的编码增益:

$$G_{\text{coded}} = 10 \log_{10} \left(\frac{d^2_{\text{coded}}}{d^2_{\text{uncoded}}} \right) \tag{14.47}$$

上例中的渐近编码增益可达到3 dB。也就是说,对于相同的带宽需求,相同的源数据速率,相同的目标误比特率,系统的发送功率需少3 dB。对有限的目标错误概率,编码增益会稍微小一些;对 10^{-5} 的误比特率,编码增益只有2.5 dB(见图14.11)。但是,同样需要注意的是,在低信噪比时,系统性能要比未编码系统差。

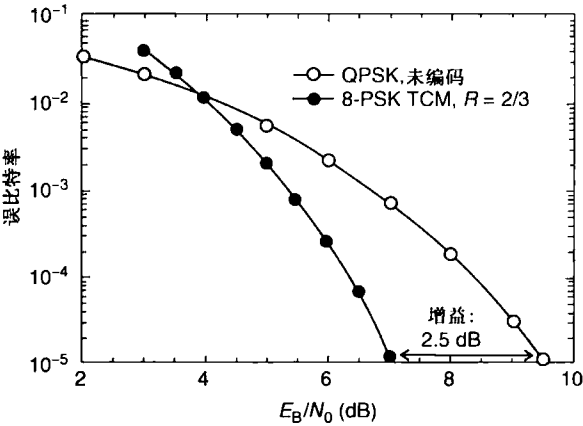


图 14.11 加性高斯白噪声信道下未编码正交相移键控和2/3 码率8-PSK 网格编码调制的仿真差错概率。引自Mayr [1996] © B. Mayr

14.4.2 集分割

上文给出了网格编码调制的优点,但其所用的码是无规则的。对于8-PSK,这种无规则构建方法还是可行的,并能得到好的结果,但是对高阶调制方式就变得不可能了。Ungerboeck [1982]提出了一种构造好码的启发式方法,即所谓的集分割。该方法的基本原理为:

- 1. 将调制符号集加倍;
- 2. 选择允许的转移,使序列之间的最小平方欧氏距离极大化。

为了距离的极大化,符号集分几步分割,在每一步中,分割集中的符号之间的最小距离都要有最大的增加。

例 14.7 集分割。

解：

这个原理有点抽象,最好通过例子来解释。考虑前一节中的 8-PSK 星座图: 信号星座图中相邻两点之间的距离为 $d = \sqrt{E_B} [2\sqrt{2}\sin(\pi/8)] = 1.08\sqrt{E_B}$ 。第一步,将现有的符号集分为两个集合(见图 14.12)。为了使元素之间的距离最大,每个集合是一个正交相移键控星座,其中一个由另一个旋转 45° 得到,这使得集内的欧氏距离增加到 $2\sqrt{E_B}$,再也没有比该星座具有更大最小距离的其他星座。下一步,正交相移键控星座被分割成两个相位相差 90° 的二进制相移键控星座,其欧氏距离为 $2\sqrt{2E_B}$,至此完成了集分割。

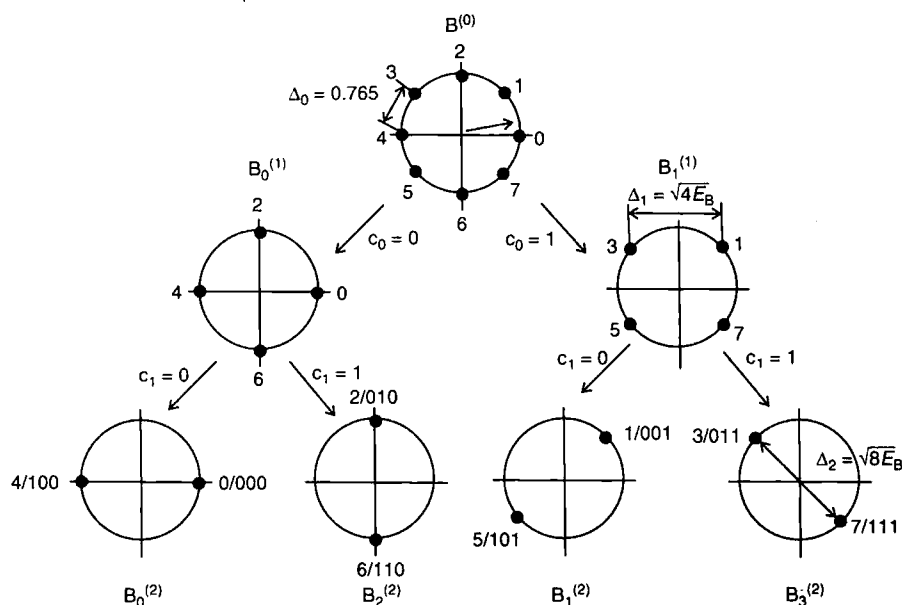


图 14.12 8-PSK 的分割

图 14.13 给出了采用这种集分割的网格编码调制编码器的结构。我们要区分 N_{syms} 与 $\tilde{N}$, 前者为每个符号传输的比特数,后者为经速率 $R_c = \tilde{N}/(\tilde{N} + 1)$ 的卷积编码器映射后的编码比特数。在编码过程中,源比特 $u_1 \cdots u_{\tilde{N}}$ 映射到编码比特 $x_0 \cdots x_{\tilde{N}}$ 。剩下的源比特 $u_{\tilde{N}+1} \cdots u_{N_{\text{syms}}}$ 直接映射到“未编码”比特 $x_{\tilde{N}+1} \cdots x_{N_{\text{syms}}}$ 。未编码比特用来选择信号在子集内的星座点,这些点之间有很大的欧氏距离,从而不会出现错误。其他比特用于选择子集。未编码的比特数越少,并行传输的可能性也就越小。译码通过维特比译码器实现(见 14.3.2 节)。

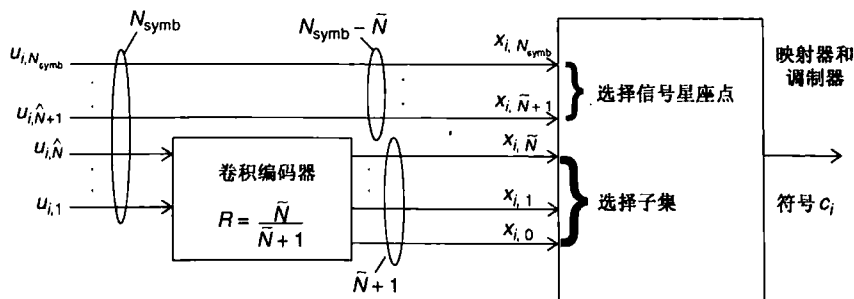


图 14.13 Ungerboeck 给出的网格编码调制编码器的结构。引自 Mayr [1996] © B. Mayr

14.5 比特交织编码调制

另一种结合编码与更高阶调制的方法是比特交织编码调制(BICM)。在这种方法中,信道编码器和调制器被交织器分开。因此,经过编码处理的比特被映射到暂时分开的符号(这可能提高分集,见14.8节),而且能够提供对于噪声毛刺更强的鲁棒性。

图14.14给出了比特交织编码调制传输的框图。比特进行编码(通过任何二进制编码器),由交织器进行交织 Π 。经过编码和交织的比特最后被映射到属于符号表 $\mathcal{X}$ 的调制符号上。二进制的标号方案(后面会详细讨论)决定了什么样的比特组合(正如在交织器之后所看到的)与哪个调制符号相对应。接下来,这些符号经过信道传输,接收到的信号被送入解映射器,它计算估计信号可靠性量度 L [对数似然比(LLR),见下文]。解交织器提供“软信息”输入到译码器。可选择地,译码器的输出可以被再次交织并反馈,以帮助符号解映射器改善其对对数似然比的估计。这种迭代的译码提高了性能,但提高了复杂度。其精神实质与将在14.6节讨论的Turbo码相似,本节不再赘述。

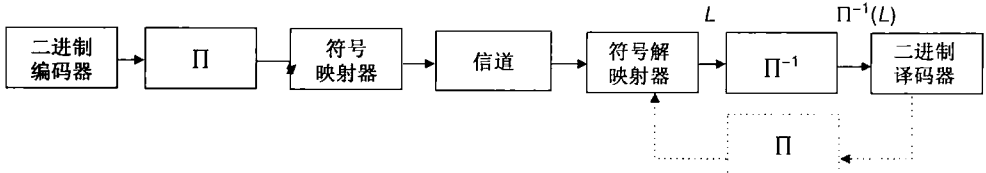


图 14.14 比特交织编码调制器框图。译码器中的反馈为可选项

二进制标号对于比特交织编码调制的性能至关重要。假定下面使用高阶调制(至少采用正交相移键控)。数值研究和理论论证均表明对于中等到高的信噪比值,二进制反射格雷码映射是最佳的。对于低信噪比,集分割映射是最佳的。

解映射器计算每个特定(编码过的)比特的软估计;更精确地说,是对数似然比

$$L_i = \log \left[\frac{\Pr(b_i = +1|r)}{\Pr(b_i = -1|r)} \right] \quad (14.48)$$

$$= \log \left[\frac{\Pr(r|b_i = +1)}{\Pr(r|b_i = -1)} \right] + \log \left[\frac{\Pr(b_i = +1)}{\Pr(b_i = -1)} \right] \quad (14.49)$$

即信号 r 被接收的条件下,比特 b_i 为 $+1$ 或者 -1 的概率比的对数值。如果进一步假定等先验概率,及一个加性高斯白噪声信道模型,则对数似然比可以计算为

$$L_i = \log \left[\frac{\sum_{\mathbf{s} \in X_{i,1}} \exp(-\gamma|\mathbf{r} - \alpha\mathbf{s}|^2)}{\sum_{\mathbf{s} \in X_{i,0}} \exp(-\gamma|\mathbf{r} - \alpha\mathbf{s}|^2)} \right] \quad (14.50)$$

α 是(复)信道增益 $\mathbf{s} \in X_{i,1}$,表示对应二进制表示(在二进制标号中)的第 i 个比特为 $+1$ 的调制符号。

已经证明,比特交织编码调制尤其在与迭代译码结合使用时具有优越的性能。它因而广泛应用于实际的无线系统中。编码与调制的分离允许非常灵活的设计(实现起来比网格编码调制更简单灵活),并使自适应调制和编码的简单实现成为可能(比如,调制星座随信号发生的信道的质量变化)。

14.6 Turbo 码

14.6.1 引言

Turbo 码是自编码领域建立以来编码理论中最重要的发展之一。正如 14.1 节已经提到的, 一个非常长的码可以达到香农极限(信道容量)。但是, 对如此长的码进行强力译码的复杂度超乎想象。Turbo 码是实际应用中第一个只需做适当的努力就能接近香农极限的码。Berrou et al. [1993] 通过将几个并行的简单码组合起来生成很长的码, 这些码通过伪随机交织器进行交织。最重要的技巧在于译码器: 因为这个码是几个短码的组合, 译码器也可以分解成几个简单的译码器, 它们交换关于译码比特的软信息, 并如此反复迭代得到结果。

码中的随机交织器可以近似地实现一个随机码的思想, 从而整个码字几乎没有结构。其优点如下。

- 交织增加了组合码的有效码长。也就是说, 决定码长的是交织器(其逆操作也很简单), 而不是分量码。这就使得用简单的编码器结构可以构造出很长的码。
- 整个码的特殊结构(即分离的分量码的组合)使译码的复杂度本质上由分量码决定, 从而使译码成为可能。

14.6.2 编码器

正如上面所说的, Turbo 码使用几个码的组合。一种组合方法是串行级联, 14.2 节已经讨论过。本节讨论码的并行级联, 其原理如图 14.15 所示。源数据流直接输出, 同时送到几个编码器分支。每个分支首先包括一个交织器且各分支都不同, 经过交织器后, 每个并行分支上的数据流通过卷积编码器, 将信息矢量 $\mathbf{u}$ 映射到输出。原则上, 并行编码器的数量是任意的, 考虑到数据速率, 最常用的是两个编码器(其码率为 $1/3$)。

用递归系统卷积码(RSC)对每个分支编码是有优势的, 其结构大致如图 14.16 所示。一个带反馈的移位寄存器用于计算编码比特。由于这种码是系统码, 所以有一个输出是直接把源比特映射到输出的。这些比特实际上不进行传输, 因为 Turbo 码总是能够传输原始数据序列的(见图 14.15 中最上边的分支), 只发送编码后的比特。因此, 这样构成的编码器的速率为 1, 即对每个源比特输出 1 比特^①(见图 14.17)。

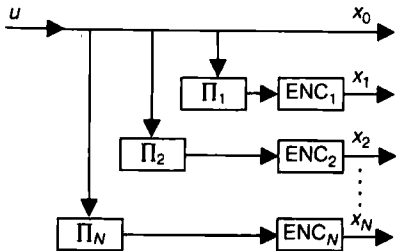


图 14.15 Turbo 编码器的结构, 其中 Π 表示交织器。引自 Valenti [1999] © M. Valenti

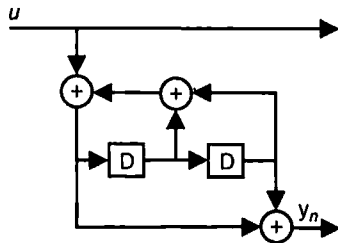


图 14.16 递归系统卷积编码器的结构, 其中 D 表示时延单元。引自 Valenti [1995] © M. Valenti

^① 如果只是把该编码器用来作为正常的卷积编码器, 码率则为 $1/2$, 因为编码器对每个输入比特都将输出信源比特和冗余比特。

综上所述,对于其中一个编码器(直通分支),使用原始数据序列作为输入,而对其他编码器,序列首先进行交织。序列(原始序列加上来自其他编码器的冗余比特)经过多路复用,最后得到的码是系统码,因为它仍然包括原始序列。由于系统部分只传输一次,整个系统的码率为 $1/(N + 1)$ 。

除了这个基本结构, Turbo 码编码器还有很多其他结构。由于这个领域的研究还很活跃,这里不再给出编码器结构的分类。

上述编码器的码率为 $R_c = 1/3$,但是可以通过删余增加码率。例如,输出 x_1 和 x_2 可以作为多路复用器的输入,既可以使用来自第一个编码器的比特 x_1 ,丢弃编码器 2 的比特 x_2 ,也可以使用 x_2 而丢弃 x_1 ,这样编码器的码率变为 $R_c = 1/2$ 。

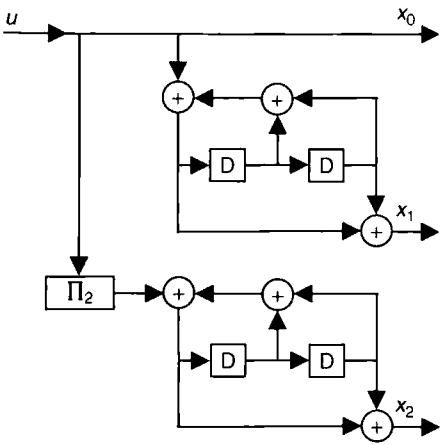


图 14.17 基于递归系统卷积编码器的 Turbo 码结构。引自 Valenti [1999] © M. Valenti

14.6.3 Turbo 译码器

编码往往是个简单的过程,而译码过程中却经常会出现一些问题。译码器中的强力实现,将组合的码看成一个单一的复杂的码,就会需要极高的计算量。而 Turbo 码有着明显的优点,可以对两个分量码分开解码,然后将来自两个译码器的信息结合起来。

Turbo 码迭代译码,通过交换分量译码器的软信息完成。图 14.18 是 Turbo 译码器的框图,它包括两个软输入软输出 (SISO) 译码器、一个交织器和一个解交织器。SISO 译码器不仅输出不同的比特,同时还有判决的置信信息。这种置信度可由对数似然比描述,由 Bahl et al. [1974] 提出的类维特比的算法计算,即众所周知的 BCJR 算法。对数似然比定义为

$$\log \left[\frac{\Pr(b_i = +1|r)}{\Pr(b_i = -1|r)} \right] \tag{14.51}$$

即接收到 r 后比特 b_i 为 $+1$ 和 -1 的条件概率之比的对数值,可见对数似然比包括三部分:

- 比特先验概率的对数似然比(由于所有比特等概率出现,对数似然比一般为 0);
- 来自接收到的原始数据(即直接观测值)的信息,该对数似然比与信噪比有关;
- 外来对数似然度,包括来自译码的信息,第二个译码器协助第一个译码器的外部信息,第一个译码器协助第二个译码器的外部信息。

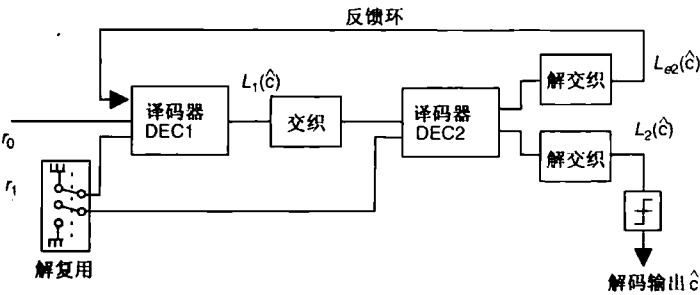


图 14.18 Turbo 译码器结构。引自 Sklar [1997] © IEEE

前两部分在未编码系统中也存在,最后一个来自于其他分量码译码的贡献。重要的是,一个分量译码器的外部信息与译码器1的输入无关,否则外部信息与直接观测信息之间会有很强的相关性,译码器1只确定自己的意见。

假设迭代开始时,对数似然比外信息为0,译码器1像“普通”的卷积译码器一样(虽然是软输出)对接收信号译码,接收机根据软输出计算译码器2处的对数似然比外信息。由译码器2的输出计算得到的外部信息用于译码器1的下次迭代。同理,译码器1在下次迭代的对数似然比外信息用于译码器2的后续迭代。重复该过程直到它收敛。

增加迭代次数可以改进误比特率性能。在加性高斯白噪声信道中,迭代次数 $N_{it} = 5$ 一般就足以达到收敛,而在衰落信道中一般需要 $N_{it} = 10$ 次迭代。每个信息比特的运算次数可以估计出来:

$$N_{op} \leq N_{it}(62 + 8/R_c) \quad (14.52)$$

每信息比特需要的典型运算次数为 400 ~ 800。这就是为什么直到进入本世纪以后 Turbo 码才被广泛采用的原因,因为 400 kbps 的传输速率的译码要求高达 300 MIPS(百万指令每秒),如此大量的信号处理能力只有那个时间点之后才能达到。

Turbo 码可以近似接近信道容量。作为一个例子,图 14.19 给出了码率为 1/2 的 Turbo 编码器在加性高斯白噪声信道中的误比特率。显然,6 次迭代就足以接近收敛值,18 次迭代后,在信噪比为 0.7 dB 时,误比特率达到 10^{-5} 。

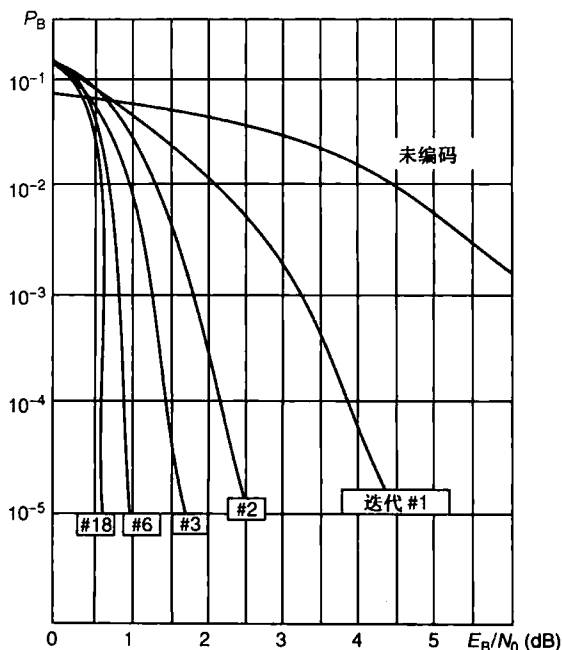


图 14.19 $R_c = 1/2$ 的 Turbo 码在不同迭代次数下的误比特率,交织器长度为 64 000,信道为加性高斯白噪声信道。引自 Sklar[1997] © IEEE

14.7 低密度奇偶校验码

1993 年出现 Turbo 码时,立刻引起了大量的关注,似乎人们第一次见到可以用实际的码接近香农极限。但事实上早在 20 世纪 60 年代就已经解决了这个问题! Gallager 于 1961 年在他

的博士论文中设计了可以接近香农极限的线性分组码,称为低密度奇偶校验(LDPC)码,还提出了有效的迭代译码机制。但是,由于在当时迭代译码超出了计算机的能力,这篇文章被完全忽视了。20 世纪 90 年代中期的几篇文章又重新发现了这些码,直到那时,曾经认为不可能的译码复杂度看起来已经很合理了。从那时开始,发表了大量论文,也提出了很多的改进。

14.7.1 低密度奇偶校验码的定义

低密度奇偶校验码是线性分组码(如 14.2 节所讨论的),要注意的是它们不是通过生成矩阵 $\mathbf{G}$ 来定义的,而是通过奇偶校验矩阵 $\mathbf{H}$ 来定义的。这是个关键技巧,因为一般引起最大问题的是译码而不是编码。因此,定义一个译码简单的结构有重要的意义! 分组大小及相应的校验矩阵的维数很大,但是矩阵中的非零元素个数保持很低。准确地说,非零元素个数与元素总个数的比值很小,这就是为什么称为“低密度”码的原因。和 Gallager 一样,我们定义一个码长为 N 的 (N, p, q) 低密度奇偶校验码,其奇偶校验矩阵中每列有 p 个 1, 每行有 q 个 1。为了使码具有好的性质,一般需要满足 $p \geq 3$ 。如果所有行线性独立,则码率为 $(q - p)/q$ 。

可以根据一些简单规则构造出好的奇偶校验矩阵。首先,将矩阵水平地分成 p 个相同大小的子阵,然后将一个“1”插入子阵的每一列。例如,将第一个子阵定义为 q 个级联单位阵,或者如下的结构:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

(14.53)

然后,其他子阵由第一个子阵进行随机的列置换得到。例如,可以得到如下所示的 $(20, 3, 4)$ 码 [Davey 1999] :

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(14.54)

显然这个结构满足每列有 p 个 1, 每行有 q 个 1, 通过观察可以发现该结构是随机的。

14.7.2 低密度奇偶校验码的编码

由于低密度奇偶校验码是由奇偶校验矩阵定义的,其编码过程比“正常”的分组码复杂一些。在一般的分组码中,只需将码字矢量与生成矩阵相乘。但是,对于低密度奇偶校验码,生成矩阵是未知的。幸运的是其计算不是很难:通过高斯消元和列的重排,可以把奇偶校验矩阵排成以下形式:

$$\tilde{\mathbf{H}} = (-\mathbf{P}^T \quad \mathbf{I}) \quad (14.55)$$

则相应的生成矩阵为

$$\mathbf{G} = (\mathbf{I} \quad \mathbf{P}) \quad (14.56)$$

注意, 由于高斯消元过程, 生成矩阵一般不是稀疏的, 这也就表示编码过程需要更多的运算。不过, 幸运的是, 这些都是离散的已知比特的运算, 因此都很简单, 所以编码的复杂性一般不成问题。

14.7.3 低密度奇偶校验码的译码

正如前面提到过的, 奇偶校验矩阵的稀疏结构是能够进行合理复杂度的译码的关键, 但是这还远远不够! 完成一个最大似然译码是个 N-P 难题(也就是说, 需要检验所有可能的码字, 并与接收信号进行比较)。因此常用到称为置信传播的迭代算法。下面具体描述这个算法^①。

设接收信号矢量为 $\mathbf{r}$ 。我们将奇偶校验方程转换成图示形式。在所谓的“Tanner 图”中(见图 14.20)^②, 区分下面两种节点。

1. 变量(比特)节点。每个变量节点对应一个比特, 其状态可以是 0 或 1。变量节点对应奇偶校验矩阵中的列, 图中用圆圈来表示这些节点。
2. 约束节点。约束节点(校验节点)描述奇偶校验方程。如果没有错误, 则约束节点的输入合计应为 0。这是由伴随式的定义得到的, 没有错误时伴随式为 0。约束节点与奇偶校验矩阵的行对应。图中用正方形来表示这些节点。

因为有两种不同类型的节点, 且相同类型的节点之间没有连接, 所以这种图称为“二分图”。除了约束节点和变量节点之外, 还有通过接收信号的观测值得到的外部信息, 它们会影响判决。

如果奇偶校验矩阵中的某个元素为 1, 则表示有相应的约束节点与变量节点相连, 即如果 $H_{ij} = 1$, 则约束节点 i 与变量节点 j 相连。来自观测信号的软信息(即外部信息)与变量节点连接, 还需要知道变量幅度的概率密度函数, 即变量节点在已知外部信息条件下有某种状态的概率。

在这种图上的译码过程通过消息传递或置信传播实现。每个节点收集进来的信息, 按照所谓的“局部规则”进行计算, 并将计算结果传递给其他节点。本质上, 给定外部信息 r_j 和来自其他约束节点的信息, 第 j 个变量节点告诉自己所连接的约束节点, 它认为的(变量节点)的值是什么, 这个消息用 λ_{ij} 表示。反过来, 给定约束节点所有来自其他变量节点的信息, 第 i 个约束节点告诉第 j 个变量节点它认为该变量节点应该是什么, 该消息为 μ_{ij} 。这个过程如图 14.21 所示。

下面从数学角度描述加性高斯白噪声信道中的译码方法。

1. 首先, 仅给定外部信息 $\mathbf{r}$ 时, 数据比特决定自己认为是什么值。已知噪声的统计特性后, 变量节点很容易分别计算出它们为 0 和 1 的概率, 把这些信息传递给约束节点。但是反过来, 约束节点暂时还没有传递有意义的信息给变量节点, 因而有

① 注意, 译码过程可以基于伴随矢量(这里将选择这种方法)和数据矢量。

② 有两种图形表示法, 这里用的是“Tanner 图”和“Forney 因子图”[Loeliger 2004]。

$$\mu_{i,j}^{(0)} = 0, \quad \text{对所有 } i \quad (14.57)$$

$$\lambda_{i,j}^{(0)} = (2/\sigma_n^2)r_j, \quad \text{对所有 } j \quad (14.58)$$

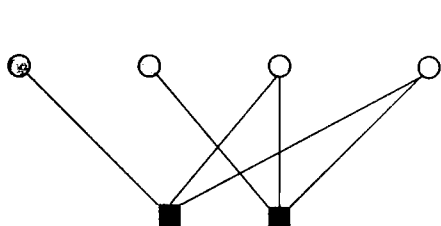


图 14.20 奇偶校验矩阵 $\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$ 的 Tanner 图

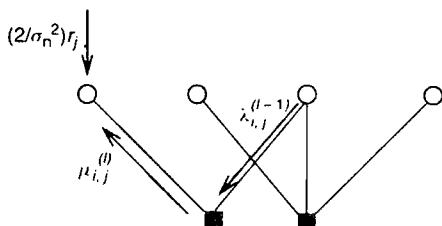


图 14.21 因子图中的消息传递

2. 然后, 约束节点给每个变量节点传递一个不同的信息。为了详细说明上述原理, 来看约束节点 i : 假设一组连线结束于节点 i , 它们来自于一个变量节点的集合 $A(i)$ 。现在, 每个校验节点都包含两部分重要的信息: (i) 它知道与该节点连接的所有数据比特的值(或概率); (ii) 它还知道进入该节点的所有比特之和模 2 为 0 (这是奇偶校验矩阵的基本点)。根据这些信息可以计算出它认为第 j 个数据比特应该有的值的概率。由于是加性高斯白噪声信道, 有连续的输出, 且不是二进制信道, 所以必须使用对数似然比代替简单的比特被逆转的概率, 从而消息变为

$$\mu_{i,j}^{(l)} = 2 \operatorname{atanh} \left(\prod_{k \in A(i)-j} \tanh \left(\frac{\lambda_{i,k}^{(l-1)}}{2} \right) \right) \quad (14.59)$$

其中 $A(i) - j$ 表示“集合 $A(i)$ 中除 j 之外的所有元素”, 即与第 i 个变量节点连接的除第 j 个节点之外的所有其他约束节点。上标 $(l-1)$ 表示第 $(l-1)$ 次迭代, 即利用前一次迭代的结果。

3. 接下来, 根据约束节点传递的信息及外部信息更新对比特节点的看法。这个规则很简单:

$$\lambda_{i,j}^{(l)} = (2/\sigma_n^2)r_j + \sum_{k \in B(j)-i} \mu_{k,j}^{(l)} \quad (14.60)$$

其中, $B(j) - i$ 表示与第 j 个约束节点连接的除第 i 个节点以外的所有其他变量节点。

4. 从以上结果, 可以计算出比特为 0 或 1 的伪后验概率:

$$L_j = (2/\sigma_n^2)r_j + \sum_i \mu_{i,j}^{(l)} \quad (14.61)$$

在此基础上试着判决码字。如果码字是一致的, 即伴随式为 0, 则停止译码。

例 14.8 一个低密度奇偶校验码的译码。

解:

考虑该算法的一个非常简单的例子。设奇偶校验矩阵为

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \quad (14.62)$$

令码字为

$$\bar{y} = [0 \ 1 \ 1 \ 0 \ 1 \ 0] \quad (14.63)$$

经过一个加性高斯白噪声信道，其 $\sigma_n^2 = 0.237$ 对应于 $\gamma = 6.25$ dB。接收码字为

$$\bar{r} = [-0.71 \ 0.71 \ 0.99 \ -1.03 \ -0.61 \ -0.93] \quad (14.64)$$

然后按照步骤 1 (如上所述)，由外部信息计算出似然度为

$$\bar{\lambda}^{(0)} = [-6.0 \ 6.0 \ 8.3 \ -8.7 \ -5.2 \ -7.9] \quad (14.65)$$

对接收似然值采用硬门限会引起码字中第 5 位出现错误：

$$[0 \ 1 \ 1 \ 0 \ 0 \ 0] \quad (14.66)$$

图 14.22 给出了消息传递算法是如何根据上述方法进行迭代而得到正确解的。

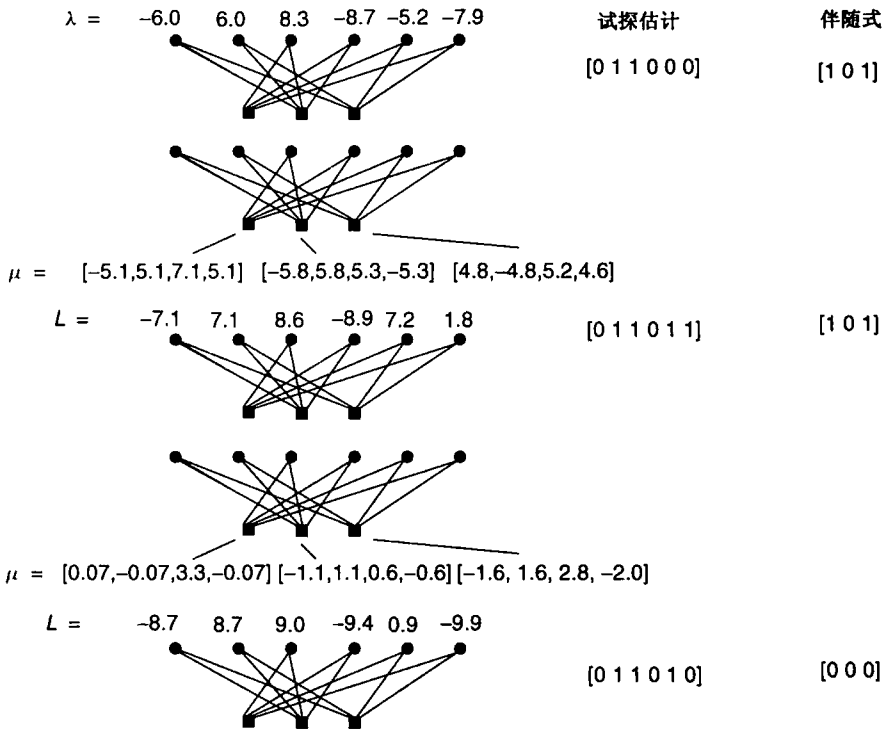


图 14.22 低密度奇偶校验消息传递的迭代举例

14.7.4 性能改进

可以看出，如果 Tanner 图能卷成树结构，即每个节点都是另一个节点的“父”节点或“子”节点，但不能同时都成立，则置信传播算法总是收敛于最大似然解。也就是说，在 Tanner 图中不会出现循环。短循环(见图 14.23)会引起收敛的问题：如果节点从错误的置信出发，它们只会自己内部加强它，而不能被来自其他变量节点的信息改变对自己的估计。在低密度奇偶校验码的设计中，构造没有短循环的码是最重要也最有挑战性的任务。但是，能化简为纯树结构的码往往不是好码，尽管它们可以通过置信传播算法准确地译码。

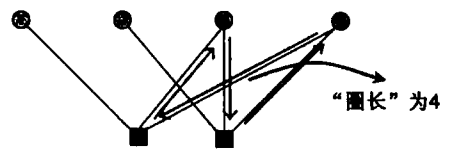


图 14.23 来自图 14.19 的 Tanner 图，示出了一个短循环(“圈长”为4)

改进收敛性和性能的另一个重要领域是使用非规则码。这意味着行和列权重是不固定的，

这与我们到现在为止所做的假设不同,而是我们只规定平均权重,允许有些列有更高的权重。与这些权重更高的列关联的节点经常收敛得更快,且可以将其“可靠”信息传播给其他节点,进而也改进了其他节点的收敛性。

低密度奇偶校验码可以达到的性能是很惊人的,(长分组)能够以零点几分贝的差距逼近香农容量。在低密度奇偶校验译码过程中还有一个有趣的计算复杂度的规则:当速率趋于香农容量的 $(1-\delta)$ 时,每比特的复杂度趋于 $(1/\delta)\log_2(1/\delta)$ [Richardson and Urbanke 2008]。对于其他大多数码,复杂度以指数 $\exp(1/\delta)$ 增长。

14.8 衰落信道的编码

衰落信道中的错误结构和加性高斯白噪声信道中的是不同的。突发错误的存在不可忽略:当信道由于多径分量的相消干涉而呈现(瞬时)大衰落时,其错误概率会比在相长干涉时大得多。这些突发错误的纠正可以通过使用专门用于纠正突发错误的码来实现,或者通过交织“分散”错误的突发结构来实现。

14.8.1 交织

与符号持续时间相比,无线信道的变化是很慢的。典型地,一个移动台在一个衰落深陷区(空间范围约为 $\lambda/4$)内会持续 $10 \sim 100$ ms,因而有大量的比特被严重衰减(也因此更容易受噪声引起错误的影响)。一般的码通常无法纠正大量的错误。为了理解这个基本原理,考虑一个简单的速率为 $1/3$ 的重复码^①。一个编码比特错误判决的概率 $P_{\text{single}} \sim Q(\sqrt{\gamma})$,其中 γ 为瞬时信噪比。由多数准则可得最终判决错误的概率近似为 $P_{\text{single}}^2 \sim Q^2(\sqrt{\gamma})$,其中利用了两个连续比特的信噪比相同的事实。在不同信道状态下求平均得到误比特率的近似表达式为

$$\text{BER} \sim \int_0^\infty \text{pdf}_\gamma(\gamma) Q^2(\sqrt{\gamma}) d\gamma \quad (14.67)$$

使用交织后,与一个源比特相关的3比特的存在较大的传输间隔,使得它们的信噪比都不同(见图14.24)。因此,只有当其中两个独立传输处于相独立的衰落深陷区时才会出现错误,而这种情况是不太可能发生的(见图14.25)。从数学上讲,这可以表示为

$$\begin{aligned} \text{BER} &\sim \int_0^\infty \int_0^\infty \text{pdf}_{\gamma_1}(\gamma_1) Q(\sqrt{\gamma_1}) \text{pdf}_{\gamma_2}(\gamma_2) Q(\sqrt{\gamma_2}) d\gamma_1 d\gamma_2 \\ &= \left(\int_0^\infty \text{pdf}_\gamma(\gamma) Q(\sqrt{\gamma}) d\gamma \right)^2 \end{aligned} \quad (14.68)$$

要强调的是,交织只有与编码结合时才会减少平均误比特率。对于一个没有编码的系统,交织器虽然仍能分散突发错误(有时是必要的),但并不会减少平均误比特率。交织的另一个问题在于其增加了传输等待时间。在话音通信中,需要保持等待时间低于 50 ms,在这种情况下,可能会出现最大等待时间小于衰落深陷持续时间的情况,这会大大降低交织器的有效性。

在这些基本考虑之后,现在来看交织器的结构。一般有两种基本结构:分组交织和卷积交织。在分组交织中,一次交织大小为 $N_{\text{interleav}} D_{\text{interleav}}$ 的数据分组。交织结构是一个矩阵:比特按行顺序读入,按列顺序读出(见图14.26)。原来相邻的比特现在相距 $D_{\text{interleav}}$,由交织引起的等

① 每个比特传送3次。接收端首先进行硬判决,然后进行多数判决。如果它确定有2或者3次译码的比特为+1,就断定源数据比特为+1。

待时间为 $N_{\text{interleav}} D_{\text{interleav}} T_B$ 。因为必须等到矩阵填满之后才能将比特读出，所以等待时间比比特分开的间隔大得多。

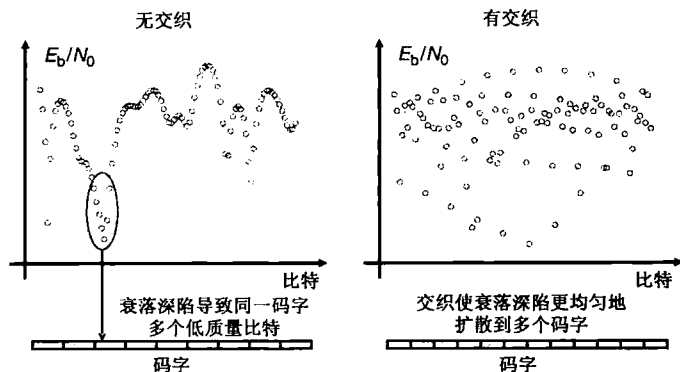


图 14.24 交织器的效果

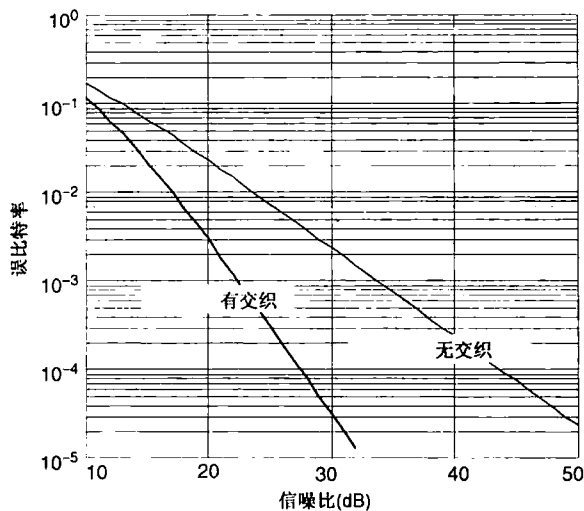


图 14.25 一个 1/3 码率的重复码有无交织的误比特率，采用多数准则硬判决，信道为平坦瑞利衰落信道

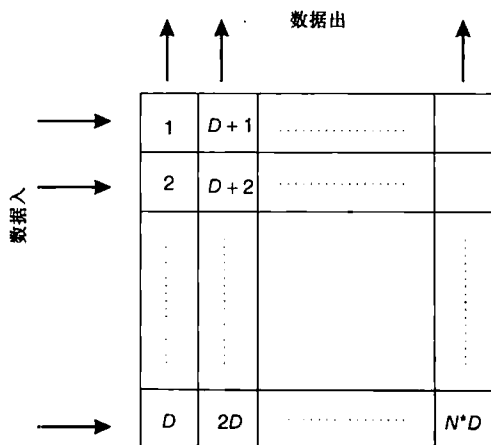


图 14.26 分组交织器的结构

对于卷积交织,数据以连续流进行交织,与卷积编码类似。与分组交织相比,减少了等待时间(相同的比特间隔)。一般分组交织与分组编码一起使用,卷积交织与卷积编码结合使用。

14.8.2 分组码

现在将上面提到的原则应用于分组码。因为有了交织,码字中的每个比特(符号)有独立的衰落。这样,即使有些比特由于在衰落深陷区而有较差的信噪比,码内的冗余也能够恢复出信息。如果没有交织,码的有效性就会下降很多。

作为例子,考虑一个硬译码的分组码:令 $K = 12$, $N = 23$, 最小距离 $d_{\min} = 7$, 编码增益为 2 dB。由于是慢衰落,如果没有交织,所有 23 比特经过的就是相同的信道,即每个码字的增益都是 2 dB。未编码系统要达到 10^{-3} 的误比特率需要 26 dB 的信噪比。有编码时,信噪比减小到 24 dB,但是仍然很差。也就是说,即使有编码器,误比特率减小只与 γ^{-1} 成正比,只不过比例常数变了。

使用交织会明显地改变这种情况。只有当码字中至少有 4 个位置出现错误时才会发生误码(注意,可以纠正 $\left\lceil \frac{d_{\min}-1}{2} \right\rceil = 3$ 个错误),误比特率近似与下式成正比 [Wilson 1996] :

$$\sum_{i=4}^{23} K_i \left(\frac{1}{2+2\gamma_B} \right)^i \left(1 - \frac{1}{2+2\gamma_B} \right)^{23-i} \approx \frac{1}{\gamma_B^4} \quad (14.69)$$

其中, K_i 为常数。更一般地,可发现一个汉明距离为 $d_{\min}$ 的采用“硬”译码的分组码的分集增益可以达到 $\left\lceil \frac{d_{\min}-1}{2} \right\rceil + 1$ 。

当交织与软译码一起使用时,其分集阶数几乎是 $d_{\min}$ 的两倍。转换成数学术语描述,需要求最小化的量度记为

$$\min_s \sum_i |r_i - \alpha_i s_i|^2 \quad (14.70)$$

则成对差错概率为 [Benedetto and Biglieri 1999]

$$P(s \rightarrow s_E) = \prod_{i \in A} \frac{1}{1 + |s_i - s_{E,i}|^2 / (4N_0)} \quad (14.71)$$

其中, A 为使 $s_i \neq s_{E,i}$ 的下标的集合。对于线性码,差错概率变为

$$P \leq \sum_{w \in W} \left(\frac{1}{1 + R_c \gamma} \right)^w \quad (14.72)$$

其中, R_c 为码率, W 是码的非零汉明权重的集合。显然可以看出,最小距离决定分集阶数(误比特率-信噪比曲线的斜率)。

14.8.3 卷积码

卷积码的性能度量和设计规则与分组码非常相似。事实上,上面给出的关于软译码分组码的数学描述可以直接应用到卷积码。这样,码的欧氏距离就不再是个重要的量度,而是应努力寻找最小汉明距离。如果码序列在很多位置都不同,则所有可区别的比特同时处于一个衰落深陷区的概率很小。换句话说,希望彼此可能混淆(即开始和结束于网格中的同一状态)的两个序列尽量长。这种序列的最小长度常称为码的“有效长度”。

基于这些简单的考虑,可以得出一些在平坦衰落信道中的码设计规则:

- 在误比特率的方程中没有最小欧氏距离,所以加性高斯白噪声信道和衰落信道中的设计准则是完全不同的。由于瑞利衰落信道和加性高斯白噪声信道都只是实际中非常重要的莱斯信道的极端情况,所以找到同时适合这两种信道的码极具重要意义。
- 衰落信道中最重要的参数是有效长度,出现在误比特率的指数部分。

14.8.4 级联码

衰落信道中最重要的一类码是卷积码和 RS 码的级联。在 1995 年之前,当需要很高的编码增益时,这种码确实是最受欢迎的。即使在今天,在因高数据速率或复杂度限制而不能用 Turbo 码和低密度奇偶校验码时,这种方法仍然是受欢迎的。

在 14.2.7 节中讨论过,RS 码擅长纠正突发的比特错误,实际上比纠正分布式错误的能力强得多,这使它看起来很适合衰落信道。但是,在特定的相关时间和比特速率下,无线信道的衰落长度(单位为比特)比 RS 码能纠正的长度大得多。所以,RS 码以如下方式应用:数据首先进行卷积编码,得到的数据流进行交织,以分散任何可能的错误突发(长的交织要比为长错误突发设计编码简单得多)。卷积码不会遇到一长串严重受扰的数据(如没有交织的情况),卷积码的译码会导致错误突发,因为这是卷积码的一个基本性质。这些错误突发通过外编码,即专门适合解决这些突发的 RS 码来纠正。

14.8.5 衰落信道下的网格编码调制

对于网格编码调制,同样的设计准则也可用于卷积码。图 14.27 表明码的有效性在不同信道中可以非常不同。我们给出了 Ungerboeck[1982]设计的用于加性高斯白噪声信道的一种码和 Jamali and Le-Ngoc[1991]中设计的用于衰落信道的一种码,很显然两种码在“自然”环境下有更好的性能。

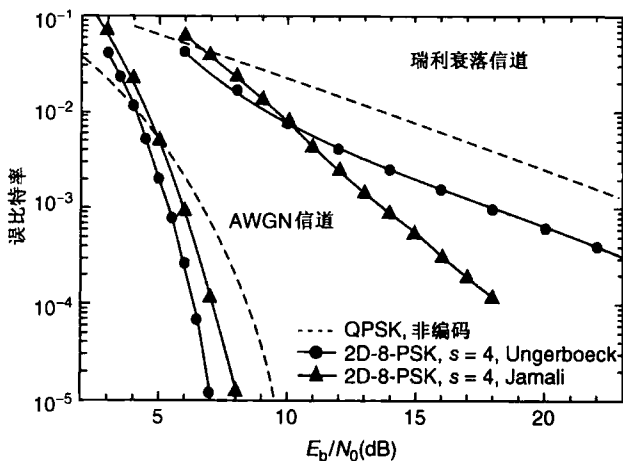


图 14.27 衰落信道中的误比特率: 非编码正交相移键控和 8-PSK 网格编码调制。引自 Mayr[1996]© B. Mayr

网格编码调制对于抵抗 12.3 节讨论的时延扩展所引起的错误也很有帮助。Chen and Chuang[1998]表明这些错误本质上是由码字之间的欧氏距离决定的,该量度的最小化给出了可以抵抗这类错误的码(见图 14.28)。平坦衰落信道下的码的搜索相对比较成熟,但是频率选择性信道下的码的设计是最近才发展起来的。

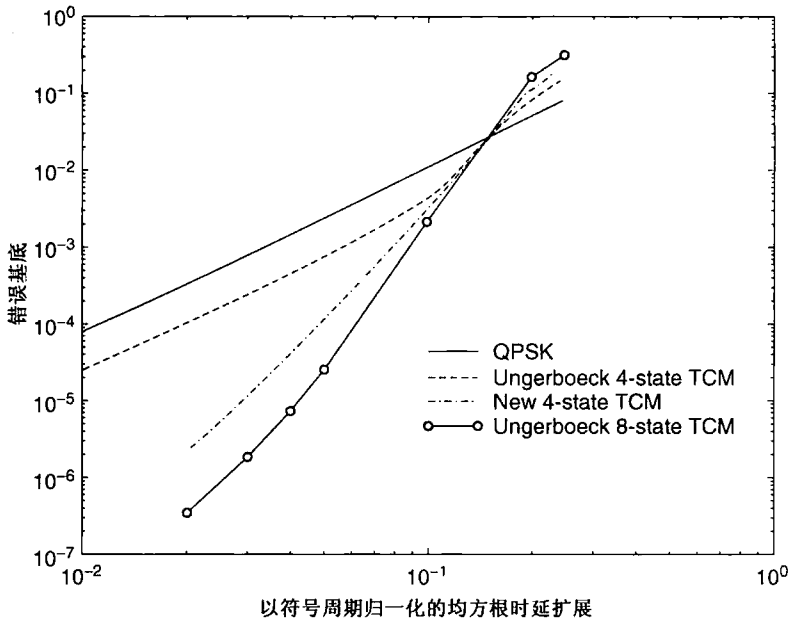


图 14.28 在频率选择性衰落信道下网格编码调制的错误平层。引自Chen and Chuang[1998] © IEEE

14.9 衰落信道的信息论性能限

讨论了衰落信道中实际码的性能之后，现在分析衰落信道下的信息论限。下文主要考虑频率平坦衰落信道，它可以用时变的信噪比 γ 描述。仅在本节最后讨论了频率选择性信道。

14.9.1 各态历经容量与中断容量

首先重申加性高斯白噪声信道的两个关键事实：(复)信道的归一化容量表示为

$$C(\gamma) = \log_2 [1 + \gamma] \text{ bits/s/Hz} \tag{14.73}$$

而且这个容量是在假设编码具有无限长度码字的条件下推导出的。

因为码字长度无限，并遍历各种不同的衰落状态，于是衰落信道的各态历经(香农)容量为信道的所有状况下容量的期望值。该量假设一个无限长的码涵盖所有的信道实现。

然而，在实际情况下经常面对的是，数据用一个接近香农限但持续时间远小于信道相干时间的码进行编码。具有合理块长度(如 10 000 比特)的低密度奇偶校验码接近香农限(见 14.7 节)。对于 10 Mbps 的数据速率，这样一个块可以在 1 ms 内传输完。这远远短于无线信道(见第 5 章)的典型相干时间 10 ms。这样，每种信道状态都对应一个容量值。这个容量(有时称为瞬时容量)是一个具有一定累积分布函数的随机变量。今后将观察这个分布函数，或者等价地这个容量可以保证在所有信道状态中占有 $x\%$ 。后一个量又称为中断容量。

因为信道是时变的，令发射功率随时间变化也是有意义的。它能否被应用及应用到何种程度，取决于发送端的信道状态信息量(CSIT)，即发送端知道关于接收端信噪比的信息。

14.9.2 仅接收端已知信道状态信息的容量

如果发送端不知道信道状态信息,那么发送端无法根据信道状态调整发射功率,所以它只能发射固定功率。那么香农容量(每单位带宽)是加性高斯白噪声信道容量基于信噪比分布的平均。

$$E\{C(\gamma)\} = \int_0^{\infty} \log_2[1 + \gamma] \text{pdf}_{\gamma}(\gamma) d\gamma \text{ bps/Hz} \quad (14.74)$$

由 Jensen 不等式可立即得到 $E\{C(\gamma)\} < C(E\{\gamma\})$;也就是说,衰落信道中的香农容量小于具有同样平均信噪比的加性高斯白噪声信道下的容量。

现在来看中断容量或者容量累积分布函数的计算。映射 $\gamma \rightarrow \log(1 + \gamma)$ 建立了两个随机变量之间的转换,于是可以运用雅可比行列式进行变量之间的转换,从而推导出 C 的概率密度函数:

$$\text{pdf}_C(C) = \text{pdf}_{\gamma}(2^C - 1) \ln(2) 2^C \quad (14.75)$$

例如,在瑞利衰落信道中

$$\text{pdf}_C(C) = \frac{1}{\bar{\gamma}} \exp\left[-\frac{2^C - 1}{\bar{\gamma}}\right] \ln(2) 2^C \quad (14.76)$$

可得到累积分布函数

$$\text{cdf}_C(C) = 1 - \exp\left[-\frac{2^C - 1}{\bar{\gamma}}\right] \quad (14.77)$$

在可接受的中断给定的情况下,可由 $\text{cdf}_{\gamma}(\gamma)$ 计算相对应的 γ ,将该值引入容量公式,从而得到中断容量。

通过这个例子可以看到,在有些信道星座上容量为零。因此,想保证一个零中断传输是不可能的。

14.9.3 CSIT 和 CSIR 的容量——注水法

当发送端已知信道状态时,它可以调节功率(服从一个长时功率约束)来使信道容量最大化。因此,可得到各态历经容量

$$C = \max_{P(\gamma)} \int_0^{\infty} \log_2\left[1 + \gamma \frac{P(\gamma)}{\bar{P}}\right] \text{pdf}_{\gamma}(\gamma) d\gamma \quad (14.78)$$

有如下约束条件:

$$\int_0^{\infty} P(\gamma) \text{pdf}_{\gamma}(\gamma) d\gamma \leq \bar{P} \quad (14.79)$$

这个最优化问题可以由下式解决:

$$\frac{P(\gamma)}{\bar{P}} = \begin{cases} \frac{1}{\gamma_0} - \frac{1}{\gamma}, & \gamma > \gamma_0 \\ 0, & \text{其他} \end{cases} \quad (14.80)$$

其中,将式(14.79)的限制条件代入式(14.80)来确定 γ_0 。这种功率分配策略称为“注水法”,将在 19.8 节加以详述。本质上它意味着,如果信道状况非常差(信噪比低于门限值 γ_0),则发射端不发送信息。如果信噪比高于门限值,则信噪比越好,发射端发射功率越大。

当考虑中断容量时,可能试图通过让发射功率补偿接收端损耗的方法来提高中断容量,即 $P(\gamma)/\bar{P} = K_{\text{inv}}/\gamma$,其中 K_{inv} 是由式(14.79)决定的一个比例常量,结果为 $K_{\text{inv}} = 1/E\{1/\gamma\}$ 。

这种方法导致瞬时容量 $C_{\text{inv}} = \log_2(1 + K_{\text{inv}})$ 与 γ 无关,也就不会产生中断,即容量永远不会低于 C_{inv} ; 因此 C_{inv} 又称为“零中断容量”。注意,在瑞利衰落中, $E\{1/\gamma\} \rightarrow \infty$, 所以 $C_{\text{inv}} \rightarrow 0$ 。

当允许中断占有情况的 $x\%$ 时,最好的策略是在 $x\%$ 的最坏信道情况下根本不发送,而在剩下的信道情况下控制发射功率,从而使信噪比保持为一个常量。

最后也是非常重要的一点,在分集接收的情况下,当 γ 表示分集合并器(见 13.4 节)输出的信噪比时,上述公式仍然成立。

14.10 附录

请访问 www.wiley.com/go/molisch。

深入阅读

在很多优秀教材中都可以看到信道编码方面的具体内容,如 Lin and Costello[2004], MacKay[2002], McEliece[2004], Moon[2005], Sweeney[2002] 和 Wilson[1996]。Steele and Hanzo[1999] 的第 4 章给出了分组编码的具体描述,包括 RS 码的译码算法, Sklar and Harris[2004] 中给出了直观的介绍,读者也可以在 Lin and Costello[2004] 中找到更详细的内容。Johannesson and Zigangirov[1999] 中具体介绍了卷积码。网格编码调制码最早是在 20 世纪 80 年代由 Ungerboeck[1982] 发明的, Biglieri et al.[1991] 中给出了极好的教程说明。比特交织编码调制在教程手册 Fabregas et al.[2008] 中进行了描述。Turbo 码的详细内容可以在 Schlegel and Perez[2003] 中找到,也可以在优秀的教材 Sklar[1997] 中找到。低密度奇偶校验码最早是在 Gallager[1961] 中提出的,由 MacKay and Neal[1997] 重新发现,并在 MacKay[2002] 中以一种非常容易理解的方式进行了描述。非规则低密度奇偶校验码的重要改进由 Richardson et al.[2001] 提出。Richardson and Urbanke[2008] 对低密度奇偶校验码进行了详细描述。Bahl 等人在 20 世纪 70 年代[Bahl et al. 1974], Hagenauer 在 20 世纪 80 年代在软输入软输出译码的研究中建立了很重要的理论基础。信息论的基础知识的描述可在 Cover and Thomas[2006] 和 Yeung[2006] 中见到。

第15章 语音编码

Gernot Kubin

格拉茨技术大学信号处理与语音通信实验室

奥地利格拉茨市

15.1 概述

15.1.1 语音电话作为对话多媒体业务

1904年,当 O. Nußbaumer 在格拉茨大学物理实验室第一次成功传输语音与音乐时,没有人能预见到无线多媒体通信这一历史性成就在 100 年后获得的巨大发展。现在已经出现了许多新的媒体类型,如文本、图像和视频,而且现代的业务范围从基本的单向媒体下载、浏览、消息传送、应用软件共享、广播及实时流媒体,发展到双向的交互式实时文本会话(聊天)、语音和视频电话等。尽管如此,语音电话依然是所有通信业务的支柱,并且继续作为任何移动通信终端都不可或缺的一部分。基于上述原因,下文将集中讨论语音信号的信源编码方法,即为了传输(或者存储)而对其进行有效的数字表示。

数字语音编码的成功起始于在公共交换电话网络(PSTN)中引入编码速率为 64 kbps 的脉冲编码调制语音进行数字交换,并继续结合一系列先进的压缩标准,从 20 世纪 80 年代早期的 32 kbps 到 16 kbps,一直到 90 年代末期的 8 kbps,全部集中于长途电路倍增技术而同时又要保持传统有线电话的高质量水平(收费话音品质)。对无线电话来讲,在一开始的 20 世纪 80 年代中期,数字编码对比特率和复杂度的要求非常苛刻,在语音质量上的一些折中看上去还可以接受,因为当时的用户或者从未有过使用移动电话的经历,或者期望值受到了相对低质量的模拟移动无线电系统的影响。然而这种状况变化很快,20 世纪 90 年代初提出的第一批标准在 5 年内就被彻底推翻,新的标准由于采用新的编码算法,语音质量显著提高,并且比特率保持在 12 kbps,而由此带来的复杂度的增加却由于微电子器件和算法实现的相对进步而减轻了。

15.1.2 信源编码基础

信源编码的基础是 1959 年由香农[Shannon 1959]奠定的,他不仅提出了非理想信道下的信道编码理论,而且提出了信号压缩的率失真理论。后一理论基于如下两个部分:

- 随机信源模型允许在信源信息中表示其冗余特性;
- 失真测度用来表示用户信源信息的相关度。

在渐近无限的时延和复杂度,以及某些简单的信源模型和失真测度条件下,可以证明存在一个达到给定失真水平所必需的比特率下界;反之亦然,对于给定的比特率,存在一个可以达到的容许失真度下界。在电话系统中,复杂度的影响越来越小,而时延则成为决定通话质量的

实质性问题,因为当时延超过 100 ms 时,就会严重影响交互通话的质量。因此,从率失真理论得到的主要启示可认为是存在一种基本参数之间的三方面折中:速率、失真和时延。传统的电话网是基于电路交换模式运行的,其传输时延实质上是由电磁波传播的时间决定的,因此只有在处理卫星链路时,时延才会明显。然而,分组交换网络越来越多地应用于电话中,如 IP 电话(VoIP),其时延会在路由队列中积累。在这些系统中,时延成为最基本的参数,它将决定着可达到的速率-失真之间的折中。

失真较小并且尚在可接受范围内的信源编码称为有损编码,而有限几个零失真的情况称为无损编码。大多数情况下,只有幅度离散的信号考虑对其脉冲编码调制语音转换编码时,有限速率才能达到无损编码,即对已采用传统脉冲编码调制编码的数字化语音信号进行数字压缩。然而,对于采用电路交换的无线语音电话来说,这样的无损编码器有两个缺点:首先,它浪费了宝贵的无线频带资源(使用了比能够达到典型用户期望的质量多得多的比特);其次,它常常导致变速率的比特流,比如使用哈夫曼编码器时,而这一点不能与电路交换传输提供的固定速率有效地匹配。

然而,变速率编码是与分组交换网及某些电路交换网的信源信道联合编码的某些应用高度相关的话题(见 15.4.5 节)。香农定理证明,在理想条件下,信源编码与信道编码过程可以完全分离,两个编码步骤可以独立设计和优化。这在具有有限时延或者时变条件的实际限制下不成立,此时只有信源和信道编码器联合设计才能达到最优。在这种情况下,由网络提供的固定速率可以方便地分为一个可变的信源编码速率和一个可变的信道编码速率。

15.1.3 语音编码器设计

信源编码理论表明了设计语音编码系统时如何利用信源冗余模型和自定义的相关性模型。感知相关性将在后面章节中讨论,此处根据信源模型给出了语音编码器设计的一般分类。

1. 波形编码器。波形编码器设计一个自适应动态系统时不直接使用信源模型,它将原始的语音波形映射到一个经处理的波形,该波形可以在给定数字信道上以更少的比特传输。译码器实质上是编码器的逆过程,用来还原原始波形的一个可靠估计。所有的波形编码器都有一个共性,即通过增加比特率可以渐进逼近原始脉冲编码调制波形的无损变换编码。对于这些系统,将误差信号定义为原始信号与解码信号的差是合理的(尽管还没有引入失真的感知相关性的直接测度)。
2. 基于模型的编码器,又称为声码器。声码器取决于明确的信源模型,用一个小的参数集来表示语音信号。编码器需要对这些参数进行估计、量化,然后在数字信道中传输。解码器利用接收到的参数控制一个实时实现的信源模型,以产生解码的语音信号。增加比特率将导致语音质量在非零失真处饱和,该非零失真的水平受信源模型的系统误差的限制。直到最近,基于模型的编码器的性能才达到使这些误差不至于产生明显的感知影响的水平,进而可以在具有轻微质量下降限制的极低速率(2.4 kbps 及以下)的应用中采用。此外,根据解码器中的信号产生过程,解码波形与原始波形不同步,因此波形误差的定义在描述基于模型的编码器失真时是无用的。
3. 混合编码器。混合编码器目的在于前两种设计的优化组合。它首先采用一个基于模型的方法提取语音信号参数,但仍然直接在波形层面上计算模型误差。这种模型误差或残留波形由波形编码器传输,而模型参数以边信息的形式量化和传输。这两种信息流

在解码器中结合在一起,用来重构波形的可靠近似,这样混合编码器就具备了波形编码器渐近无损编码的特性。它们的优势在于明确的语音模型参数化,允许在设计时利用比纯波形编码器中的单一可逆动态系统更先进的模型。

基于模型观点的语音编码设计认为,一个语音编码系统的描述应该从解码器开始,该解码器潜在包含一个语音模型的实现。编码器作为信号分析系统用来提取相关模型参数和残留波形。因此,编码器比解码器复杂度更大,更难理解和实现。在这种意义上,语音编码标准可能仅指定解码器及传输数据流的格式,而将最佳匹配的编码器设计留给工业界去竞争。

进一步的设计问题

语音编码器对信源模型的依赖性很自然地导致其性能依赖于此模型与被编码信号的匹配程度。任何信号只要它不是由靠近麦克风的唯一讲话者产生的纯净信号,都可能经受附加的失真,这就要求额外的性能测试,而且可能需调整设计方案。这些额外问题的例子适用于音乐编码器(例如,如果停下来等待特定的一方时),严重的声学背景噪声下的编码器(例如在车内说话,可能开着车窗时),存在其他人喋喋不休的谈话时的编码器(例如在咖啡店里),回声编码器(例如在免提模式下讲话),等等。

进一步的系统设计还包括窄带语音与宽带语音的选择。窄带语音在传统的有线电话中使用(例如,模拟带宽从 300 Hz 到 3.4 kHz,采样频率为 8 kHz),宽带语音质量近似于调频广播(例如,模拟带宽从 50 Hz 到 7 kHz,采样频率为 16 kHz),它比普通常规电话业务(POTS)有显著的语音质量提高。其次,即使采用非常复杂的信道编码,信道误码的鲁棒性,如单个比特错误、突发错误、整个帧或包的丢失之类的问题,同样是信源编码的设计问题(将在 15.4.5 节中讨论)。最后,在从说话人到听话人的网络路径中,可能有多个声码器级联步骤,每一步都需要从一种编码标准转换到另一种编码标准,每一次都潜在造成了语音质量的进一步损伤。

15.2 语音

音乐的声音可以由大量信号产生装置来产生,比如管弦乐队的乐器。而语音的产生装置,即使在声学层面或感知层面,都相当独特,而这构成了语音模型的物理基础。

15.2.1 语音的产生

简而言之,语音通信由以自然语言为代码,以人类的声音为其载体的信息交换组成。语音是由复杂的振荡器,即声带,受肺部的气流压力激励而产生的。从无线工程的角度看,这个振荡器产生的是准周期信号,即离散多音(DMT)信号。对男声而言,其基音频率 f_0 在 100 ~ 150 Hz 之间,女声则为 190 ~ 250 Hz,童声是 350 ~ 500 Hz。其频谱从低频到高频缓慢衰减,跨越达几千赫兹的频率范围。从相对带宽看,可以把它看成超宽带信号^①,这就是这些信号在很多不同的自然环境中都能很稳健且功率效率高(比如歌剧演员不用扩音器)的原因之一。

声音信号沿声道进行传播,通过喉咙、口腔、鼻腔等,这些发声器官可以认为是一个声学波导管,用其共振频率修整信号的频谱包络,这些共振频率称为共振峰频率。最后,信号从一

^① 如果考虑光和声音的相位速度相差约 10^6 的比例因子,那么这种类比将更加显著,1 kHz 的声波和 1 GHz 的电磁波波长大致相同。

个较大的空间中(人的头部)的相对窄小的出口(嘴或者鼻孔)辐射出来,引起高通辐射特性,使得远场语音信号呈现典型的 20 dB 每十倍频程的频谱提升。

声音的产生

除了振荡的声音信号(发声),其他的声源也能成为信号载体,比如在声道较窄条件下产生的紊乱噪声(比如摩擦音,“lesson”中的[s])或者由于声道完全闭合,然后突然放开气流产生的冲击压力产生的声音(比如爆破音,“attempt”中的[t])。如果在声音产生的过程中声带不振动,这样的语音称为清音,否则为浊音。在后一情况中,准周期信号和具有噪声特性的信号可以共存(在“浊擦音”中,像“puzzle”中的[z]),这样的声音经常称为混合激励语音^①。

除了以上提到的 3 种发声方法(发声、摩擦音、爆破音)外,还有经常被忽略的第 4 种发声方式:无声。例如,比较单词“mets”和“mess”。它们的主要区别在于:在“mets”中元音[ε]和[s]之间[t]的闭合期有一个无声的阶段,而在“mess”中无此阶段。否则,这两个发音就相同了。因此[t]的信息是在无声期间传达的。

发音

声音产生机制为语音提供了载体,发音机制以语言代码对其进行调制。这种代码是有次序和层次的,即一系列的语音(用语言符号的话讲就是音素)组成音节,一系列的音节组成单词,再由单词组成短语,进而构成了语言。然而,发音的过程不是按纯粹的顺序方式组织的。对于每一个语音,发声器官波导的形状不是由一个稳定状态姿势转化到另一个稳定状态姿势。并且,发音的姿态(唇、舌、软颚、喉咙的运动)是相互重叠和交错的(常常跨越整个音节),它大多是异步的和连续变化的,从而使调制方式是连续而非离散键控的。这种联合发声的现象在自动语音识别与合成系统中是一个核心问题,即把连续变化的声音变为离散的符号串(反之亦然)。

15.2.2 语言声学

信源滤波器模型

对于语音声学的科学研究可以追溯到 1791 年, W. von Kempelen, 时任玛利亚·特丽萨皇后宫廷里的高官^②, 出版了他对作为人类语音产生的物理模型的第一个机械说话装置的描述。现代语音声学的基础是 20 世纪 50 年代由 G. Fant[Fant 1970]奠定的, 基于此产生了语音信号的信源滤波器模型(见图 15.1)。

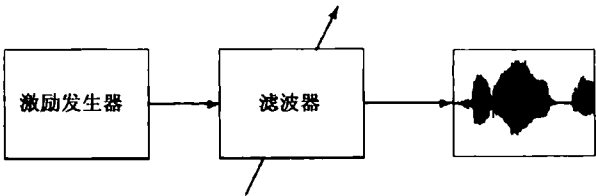


图 15.1 语音的信源滤波器模型: 激励信号发生器提供信源驱动, 激励一个随时间慢变化的滤波器, 滤波器根据共振峰频率修整频率包络来产生语音波形

① 注意, 根据定义, 混合激励语音显然是浊音的子集。
② 玛利亚·特丽萨皇后是 18 世纪初叶奥地利的公主和匈牙利的皇后。

尽管这个模型仍然源于对自然语音产生过程的理解,但它还是严重偏离了它的物理基础。特别地,所有的自然声源(它可以位于沿着声道的许多位置,并且常受到该处气流的控制)都被合成了一个单一的信源,并由它独立地激励滤波器。而且,这种模型只有一个输出,而自然发音系统可能在声道的口腔和鼻腔之间转换,甚至还会合并使用。因此,这个模型的真正价值不在于准确描绘人类的生理结构,而在于其在语音声学建模方面的灵活性。特别是,这种结构很好地表达了语音信号的典型性质,这在时间和频谱分析中已被证明。

语谱图

图 15.2 给出一个典型的语音时频分析的例子。

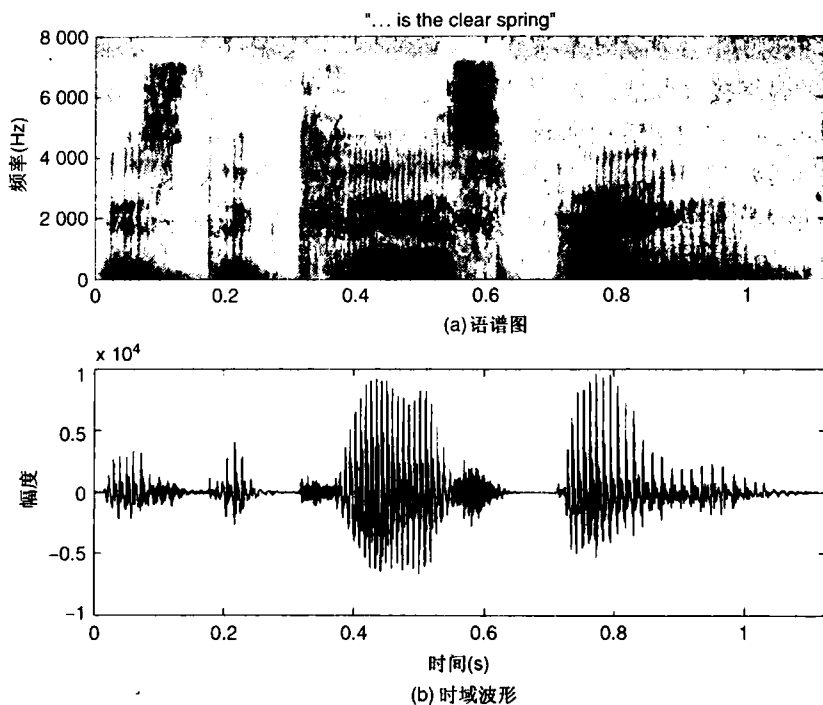


图 15.2 一段男声语音:“is the clear spring”的语谱图,其模拟带宽在 7 kHz 内,采样频率 $f_s = 16$ kHz,横轴为时间,纵轴为频率,深色区域表示高能量密度

图 15.2(b)是短语“is the clear spring”的时域波形,讲话者为一名男性,限制在 7 kHz 的模拟带宽内,采样频率为 16 kHz。这幅图清晰地显示了 4 种激励信源机制的变换。可以注意到,浊音的“准周期”信源可以解释为 3 个基本“周期”的间隔,而基本“周期”在对应于“the”的第二个浊音段的情况中则表现出高度的不规则性。并且,语音包络有很强的起伏,通常在 20~30 dB 之间。

图 15.2(a)是同一信号的语谱图,它表示信号能量的时间-频率分布,其中深一些的区域对应高能量密度。这种表示形式可以通过滤波器组或者短时傅里叶分析得到。它阐明了信号的整体性质(像 7 kHz 的防混叠低通滤波器),通过观察可以看到在 3.4 kHz 以上还有很多语音能量(尤其是摩擦音和爆破音),这说明传统电话的频带过窄导致自然话音质量严重下降^①。由此得到如下局部性质:

① 普通常规电话业务带宽限制对可懂度的影响,可以从用英语打电话时需要采用“alpha/bravo/charlie...”的方式拼写字母来交流一个未知的专有名称得到佐证(有些类似中国人打电话时用“洞”,“么”,“捌”分别表示 0, 1, 7——译者注)。

1. 类噪声信号分量表现出宽带不规则的能量分布;
2. 类脉冲信号分量表现为宽带毛刺;
3. 准周期分量表现为有间隔(男声在 10 ms 左右)的狭窄纵向条纹,对应时域上的脉冲序列;
4. 共振峰频率表现为由于声道共振引起的能量浓度的缓慢漂移,而这显然与前后内容密切相关,就像这句话中三个[i]的音都不同。

前 3 条性质与信源或者信源滤波器的激励信号有关。第 4 条性质与滤波器和它的共振频率或极点相关。这些极点的时域慢变化模拟了共振峰频率的时域变化。

15.2.3 语音感知

人类语音的最终接收者是人类的听觉系统,它是一个不同寻常的接收机,包括两个宽带的、能适合空时滤波的定向天线(外耳),根据单独的、单声道头部相关传输函数 HRTF(方位和频率的函数)滤波,通过耳间的时延估计,产生听觉的立体声效果。其次,有一个高度自适应的机械的阻抗匹配网络(中耳),其动态范围超过 100 dB,还有一个具有 3000 多条信道的、锁相的、基于门限的接收机(内耳耳蜗的毛细胞),其自身噪声很低(刚刚超过人类能听到自己的血液循环引起噪声的水平),它将信号转换成一个高度并行的、低速率的同步表达形式,这种形式特别适合低功耗、非精确电路的(人类神经系统)分布式处理。

听觉语音模型

心理声学方面的听觉模型,即感知的行为模型在音频编码器中(像普遍应用的 MP3 标准)很流行,因为它使我们可以从信息的不相关部分中分离出相关部分。例如,某些信号分量可能被其他信号掩蔽以至于使它完全不被听到。自然地,一个高质量的音频编码器,对信源的性质和其内在的冗余不能做先验假定,它要求通过控制有损编码的失真,以便失真被相关信号分量掩蔽。

在语音编码中,情况则恰恰相反:我们具有大量的关于信源的先验知识,而且可以基于信源模型和它的冗余进行编码器设计,与音频编码器相比,感知的质量要求在某种程度上有所放宽,即大多数语音编码器中的一些没有被掩蔽的,可听到的失真是可接受的(注意,长时间以来,人们已经习惯于 3.4 kHz 的带宽限制引起了失真,这在听音乐时就不可以)。因此,感知模型在语音编码器设计中的作用更小,虽然最初由 Schroeder et al. [1979]提出的简单感知加权滤波器被大多数语音编码标准接纳用来进行感知上有益的量的噪声成形。关于对语音和音频进行透明编码的可逆听觉模型的最新研究成果可以参考 Feldbauer et al. [2005]的综述。

感知质量测量

检测语音编码器的好坏要靠测听。直到现在,测量语音编码器质量的最好方法仍是由一定数量的测听人员(20 多人或者更多)做可控的听力测试。相关实验过程已经在国际电信联盟建议 P.800 中进行了标准化,它包括绝对分类打分和相对分类打分测试。平均意见得分值(MOS)就是属于第一类的一个重要例子,它要求测听人员对听到的语音质量进行打分,分值从 1 到 5,其中 1 分为劣,5 分为优。传统的窄带语音在采用 64 kbps 对数脉冲编码调制进行编码时,其 MOS 得分在 4 分附近。经过恰当设计的试验设置,可以做到高的可重复性和分辨能

力。通过国际电信联盟的调制噪声参考单元(MNRU),可以产生人为控制的失真来建立测试条件,以校准这种测试方法。

除了单向测听实验,在语音传输存在时延(包括信源编码时延)的情况下,双向的会话测试也很重要。这种测试导致了语音质量评价的全面规划工具的建立,这就是被 ITU-T 建议 G.107 标准化的 E 模型,它涵盖了单向编码、传输损失到会话延时的影响,或者从移动服务接入获得的主观用户优势等各种效应。

作为一种可以省去烦琐的测听和(或)会话测试的手段,客观语音质量评价已经变得越来越重要。这种测量技术往往基于感知模型来评估编码失真的影响,而编码失真可以通过比较原始信号与解码信号得到。此类方法也称为插入法,因为它要求在设备(网络)中传输一种特定的测试语音信号,以便在接收端能够使用原始信号的未编码形式。一个标准化的例子为语音质量感知评价(PESQ)(可与 ITU-T 建议 P.862 进行比较)。

非插入式或者单端语音质量评价仍在研究中。它不需要原始语音信号就能够评价语音质量,正如打电话时不需要直接接入电话的另一端,就能判断一个电话连接的质量很差。第一个标准化尝试(从 2004 年 5 月始)是单边语音质量测量(Single Sided Speech Quality Measure) 3SQM(ITU-T 建议 P.563)。

15.3 语音信号的随机模型

15.3.1 短时平稳模型

与语音信号的产生和感知相关的所有物理知识中,语音编码只关心实际的声学波形,并且由于它是一个携带信息的信号,因此需要一个随机模型的框架。首先,需要确定怎样利用物理信号生成机制的时变特性,这需要采用非平稳的随机过程模型^①。根据声波传输信道的时变性,这个非平稳过程有如下两个“外部”的源。

- 发音器官的运动,它把声道的边界条件以每秒 10 ~ 20 个音的速率成型。由于发音器官冲击响应的典型时延扩展少于 50 ms,对于这种非平稳性来讲,采用短时平稳表示是足够的。
- 声带的运动,它以基础频率 100 ~ 500 Hz 振动,使声道最低端(即声带)的边界条件以准周期的速率变化。这种情况下,通过与时延扩展比较可知这种变化太快,不适合采用短时平稳模型,多普勒频移引起的调制效应变得很重要,只有循环平稳表示能处理这种影响(可与 15.3.5 节比较)。

许多经典的语音模型都忽略了循环平稳性,所以我们先讨论短时平稳模型。对于这类模型,随机特性是随时间缓慢变化的,因此采用子采样估计,即只需对每一帧信号重新估计一次模型参数,而在大多数系统中,一帧为 20 ms(在采样频率为 8 kHz 时对应 $N = 160$ 个采样点)。对于有些应用,比如在解码端采用参数模型产生信号,参数需要更新得更快(如每 5 ms 的一个子帧),这可以通过对已有基于帧的估计进行插值得到。

^① 另一种观点是引入超模式来控制时变语音生成模型参数的变化。一个平衡的超模式能够反映随机挑选言语的平衡过程,以便总的双重语音信号模型是平衡的。

Wold 分解定理

在一帧中, 信号可以看成平稳随机过程的样本函数。这样就可以应用 Wold 分解定理 [Papoulis 1985], 该定理保证任何平稳随机过程都可以分解为两部分之和: 常规分量 $x_n^{(r)}$ 和奇异分量 $x_n^{(s)}$, 前者可以理解为经过滤波的噪声, 在线性系统中很难预测; 后者是一系列正弦信号的组合, 在线性系统中可以很好地预测:

$$x_n = x_n^{(r)} + x_n^{(s)}$$

(15.1)

注意, 这些正弦信号不必谐波相关, 可以具有随机的相位和幅度。接下来的三节将证明这个结果(已经在 1938 年的随机过程理论中建立起来)是当今一系列语音模型的理论基础, 这些模型仅仅在实现的细节上略有不同。这些模型是: 线性预测声码器、正弦模型和谐波 + 噪声模型(HNM)。

15.3.2 线性预测声码器(LPC)

Wold 分解定理的第一种实现方式强调它与线性预测(LP)理论的关系, 线性预测理论最早由 Itakura and Saito [1968] 提出。该模型的结构由将常规分量表示成白噪声通过一个线性滤波器, 这个滤波器是因果稳定的, 而且有一个因果稳定的逆, 即最小相位滤波器。同样的滤波器也用来产生浊音中的奇异分量, 它包含许多谐波相关的正弦波, 这些正弦波可以模型化为周期性脉冲序列, 通过线性滤波器。这样就能灵活调整所有谐波的频谱包络, 但不保留原始相位信息。因为现在相位是由最小相位滤波器的相位响应决定的, 它与对数幅度响应(通过希尔伯特变换)严格对应。进一步讲, 两种激励类型(白噪声和脉冲序列)只有一个滤波器, 该模型将叠加简化为时域的硬切换, 这就要求白噪声和准周期信号在时域严格分隔(比较图 15.3 的解码器模块)。这种简化使得系统模型与混合激励信号的情况不相符。

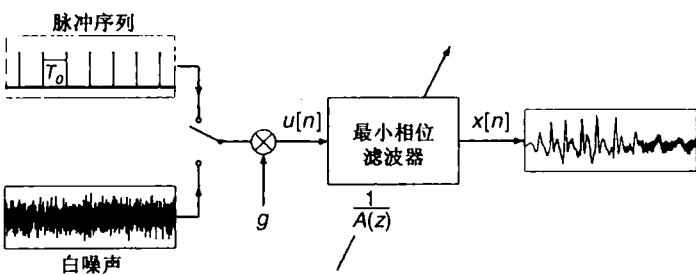


图 15.3 解码器中采用的线性预测声码器信号发生器

线性预测分析

对于每一帧, 线性预测编码器要估计模型参数, 并进行量化和编码以便进行传输。这里仅讨论线性预测滤波器的参数估计问题, 基音周期 $T_0 = 1/f_0$ 的估计在下文讨论(见 15.4.3 节)。第一步, 把最小相位滤波器限制为一个全极点滤波器, 即滤波器传输函数的所有零点都集中在 z 平面的原点, 而只有极点反映频率响应。这是基于如下两点考虑的。

- 声道主要由其谐振频率来刻画, 对波峰的感知要比对波谷的感知敏感得多。
- 对于一个最小相位滤波器, 所有的零极点都在单位圆内。这样, 当一个零点 z_0 处于 $|z_0| < 1$ 的位置时, 可以由几何级数的极点来代替, 这些极点在 $|z| = 1$ 处收敛:

$$(1 - z_0 z^{-1}) = \frac{1}{1 + z_0 z^{-1} + z_0^2 z^{-2} + \dots} \quad (15.2)$$

将信号模型转化为 z 变换域表示:

$$X(z) = \frac{U(z)}{A(z)} \quad (15.3)$$

其中, $X(z)$ 表示语音信号, $U(z)$ 为激励信号, 滤波器传输函数为 $1/A(z)$ 。其时域表示为

$$x_n = - \sum_{i=1}^m a_i x_{n-i} + u_n \quad (15.4)$$

其中, 对采样频率为 8 kHz 的语音信号, 滤波器或者预测器的阶数选为 $m = 10$, a_i 是预测器系数, 通常 a_0 归一化为 1^①。

对于给定的语音信号帧, 通过估计模型参数 $\hat{a}_i$, 使模型失配达到最小。它的误差通过预测误差信号 e_n 表示:

$$e_n = x_n - \hat{x}_n = x_n - \left(- \sum_{i=1}^m \hat{a}_i x_{n-i} \right) = \sum_{i=1}^m (\hat{a}_i - a_i) x_{n-i} + u_n \quad (15.5)$$

当且仅当 $\hat{a}_i = a_i (i = 1, \dots, m)$ 时, 对于非相关激励 u_n , 预测误差功率达到最小。此时, 预测误差信号 e_n 与激励信号 u_n 一致。注意, 采用预测框架只是为了模型拟合, 而不是预测或者外插未来的信号。为了将这些估计器应用于短时平稳语音数据, 对每一帧更新 N 个样本的情况, 必须选择一个 $L \geq N$ 的窗, 其中大于号意味着窗要有一些交叠或者超前。典型地, 应用一个特殊的窗函数 w_n 是为了减轻由加窗机制带来的数据非连续性。如果非对称窗的峰值靠近最新的样本, 它将在估计精确度和时延之间获得一个较好的折中。窗函数可以采用两种不同的方式应用于数据, 产生了如下两种线性预测分析方法。

1. 自相关法。该方法定义基于如下加窗语音信号:

$$\hat{x}_n = \begin{cases} w_n \cdot x_n, & n = 0, 1, \dots, L-1 \\ 0, & \text{窗外的 } n \end{cases} \quad (15.6)$$

的预测误差功率估计为

$$\sum_{n=-\infty}^{\infty} e_n^2 = \sum_{n=-\infty}^{+\infty} \left(\tilde{x}_n + \sum_{i=1}^m \hat{a}_i \tilde{x}_{n-i} \right)^2 \quad (15.7)$$

这里数据加窗导致了这种方法需要无限求和的局限性。

2. 协方差法。该方法不对语音直接加窗, 而是通过对误差信号加窗, 进而定义预测误差功率估计为

$$\sum_{n=m}^{L-1} (w_n \cdot e_n)^2 = \sum_{n=m}^{L-1} w_n \left(x_n + \sum_{i=1}^m \hat{a}_i x_{n-i} \right)^2 \quad (15.8)$$

这里必须小心地选择求和限, 避免用到窗以外的信号样本。

在这两种方法中, 最小化二次代价函数得到一个未知参数 $\hat{a}_i$ 的线性方程系统。在自相关法中, 系统矩阵变成严格意义上 Toeplitz 结构的相关矩阵, 它可以由 Levinson-Durbin 递推算法

① 增益 g 可以包含在激励信号的幅度中。

高效求解。这种算法将估计器的运算次数由 $O(m^3)$ 降到 $O(m^2)$, 并保证多项式 $\hat{A}(z)$ 的根在单位圆内。自协方差法中, 没有这种结构可以利用, 需要有更高的复杂度和保证 $\hat{A}(z)$ 的稳定性的附加条件。自协方差法的优点是系数估计值的精确度大大提高了, 因为它没有了自相关法中的直接加窗处理, 避免了一些系统误差。最后, 为了提高线性预测分析的数学性能, 通常加入一些(预)处理过程, 包括输入语音信号的高通预滤波(以滤除多余的低频分量), 在相关函数估计中应用带宽扩展, 在滞后为零处的自相关的修正, 这对应着加上一个微弱的白色基底噪声(在数据的 -40 dB 处)。

15.3.3 正弦模型

Wold 分解定理的第二种实现方式强调由 MacAuley and Quatieri[1986]提出的采用正弦波进行频谱建模的思想。在这个模型中, 两部分信号分量由同一系列正弦波的和组成。如果将正弦波的相对相位频繁地随机化, 至少做到每帧一次, 则类噪声的规则分量就可以由这样和式(根据谱表示理论)近似表示。这种方法的优点是: 奇异分量中的原始相位信息能够保留下来(当然, 如果可用比特率允许对它进行编码), 而且它的奇异分量与常规分量的组合上更加灵活。通常, 在高频段, 即使浊音也包含有类噪声分量, 这样就可以用一个低次谐波的固定相位模型和一个高次谐波的随机相位模型来表示。这种频域的硬切换假设了类噪声分量和准周期信号分量的频率是隔离的。这样就能进行混合激励语音信号的建模, 激励信号之间的转换被预先限定在帧的边界(而线性预测声码器在原理上允许清浊转换有更高的时间分辨率)。

正弦模型的进一步发展是多带激励(MBE)编码器, 它允许不同频段话音分开处理, 因此能更精确地描述混合激励现象。这也建立了与不依赖特定信源模型的、在通常的音频编码中应用最广泛的子带编码或变换编码原理的联系。

15.3.4 谐波 + 噪声模型

Wold 分解定理的第三种实现方法力争在时域和频域都不采用硬切换的情况下, 达到奇异分量和常规分量叠加的完全实现。这样, 谐波 + 噪声模型的解码器结构可能是最简单的, 只需要遵循式(15.1)即可, 然而它的编码器难度最大, 因为要同时估计叠加的连续谱(常规分量)和离散谱(奇异分量, 假设是谐波相关的)。现在, 该模型仅限于在语音和音频合成中应用, 而在语音编码中, 人们已经意识到谐波 + 噪声模型频谱所达到的复杂度需要更详尽的时域变化分析作为补充, 而这是传统的短时平稳模型所望尘莫及的。

15.3.5 循环平稳模型

语音的短时平稳模型忽略了由声带振动引起的快速时变。对于浊音来讲, 这种准周期性振荡, 不仅作为发声器官的主要激励而且是所有信号统计的循环调制。具有周期统计特性的随机过程称为循环平稳过程。对于我们的应用, 基音周期和相关联的振荡模式会随时间缓慢变化, 因此需要引出短时循环平稳过程来研究这一特性。对于一个给定的语音帧, 它的波形可以分解为一个循环均值信号(对应于 Wold 分解中的奇异分量)和一个具有零均值的类噪声过程, 该过程具有周期时变性的自相关函数。更重要的是, 类噪声分量的方差是时间的周期函

数,表示浊音的类噪声分量的周期包络调制。这种效果在具有较低的基音周期的男声中可以清楚地听到,并且它也是循环平稳模型相对于谐波+噪声模型的主要改进。这种模型最早作为原型波形内插(PWI)编码器由 Kleijn and Granzow[1991]引入,当时循环均值解释为“慢变化的波形”,周期时变类噪声分量称为“快变化的波形”。更多最近的研究进展是结合滤波器组进行的,通过在时域预弯折信号,信道可以适应 f_0 的多倍的情况。循环平稳信号建模的关键是提取可靠的“基音标记”。提取标记就能直接进行后续基音周期的时域同步。正如所有语音特性随时间变化一样,同步问题是一个主要的挑战,并且引出了以自激振荡非线性动态系统来描述语音信号的特点,这种方法的好处是能够解释语音信号特定的非加性不规则性,比如抖动或者倍周期现象。

15.4 量化和编码

一旦选定了信号模型和参数估计算法,编码器的任务就是将参数量化,在混合编码中,还要量化模型残余波形的样点值^①。为了提高信源编码效率,很多模数转换器都已经发展出多种优于简单均匀量化方案的技术。

15.4.1 标量量化

标量量化泛指对一个随机变量的量化,它可以是一个波形样本或单个模型参数。如果需要量化的变量超过一个,则独立处理每一个变量。作为一个一般化的例子,假如随机变量 X 的连续值 x 满足 $t_i \leq x < t_{i+1}$,则 x 量化为区间 $[t_i, t_{i+1}]$,其索引 $i = Q(x)$ 用(二进制)码字传输。接收端把接收到的码字解码为量化器索引,使用该索引查找一个存储高分辨率重构值的表,得到 $x^{(q)} = Q^{-1}(i)$ 。后者的操作有时称为“逆量化”,尽管量化器由于静态非线性是一个多对一映射,没有反函数。标量量化器的典型性能指标为量化信噪比(SQNR),其值基于量化误差 $q = x^{(q)} - x$ 的均方值:

$$\text{SQNR}_{\text{dB}} = 10 \log_{10} \frac{E(x^2)}{E(q^2)} = 10 \log_{10} \frac{E(x^2)}{E((x^{(q)} - x)^2)} \quad (15.9)$$

其中,符号 E 表示数学期望算子,即对变量 X 求加权平均,权值为概率值。该定义包含两种失真,一种是溢出失真,即输入值超过量化器的边界的情况,另一种为与单个量化步长相关的颗粒失真。在大多数设计中,采用合适的量化比例可以避免溢出失真,因此颗粒失真将是确定质量效果的关键因素。当处理波形量化时,将误差功率用输入功率归一化,这样可以使该量度和人类的感知更相关一些。

在均匀量化中,所有的区间有着相同的宽度,即 $t_{i+1} - t_i = q = \text{常量}$ 。最著名的非均匀量化定律是脉冲编码调制语音信号的8比特对数量化,它有两个版本,一个是美国和日本采用的 μ 律,另一个是欧洲采用的A律。对数量化有一个显著的特点,即如果可以避免溢出且当信号

① 注意,这是前向自适应语音编码的情况。对于极低延迟后向自适应编码而言,为了避免参数和边信息一起传输,则只利用已经解码的语音样本来估计模型参数。这要求在传输端同时实现编码器和解码器,并假定工作在低信道误码率的状态。从解码数据中估计出来的参数仅仅隐含某些相对于在前向自适应方案中当前帧中得到的参数的延迟。然而,由于短时平衡假设,这种延迟造成的模型性能损失是可接受的,这种损失由抑制边信息带来的比特率降低近似得到了补偿。

未靠近 0 时(此时,对数函数可用线性函数近似,对应于 12 比特的均匀量化器),其量化信噪比几乎独立于^①信号功率 $E(x^2)$ 。这样,对数量化器可以处理语音信号包络的大幅度时域变化,包括语音信号自身固有的变化(高度不同的语音的短时平稳序列,温和的和大声的语音)和外在的变化(与麦克风之间的距离变化)。

自适应量化使量化器的动态范围比对数量化更大,可以使输入功率波动达 40 dB 或者更高,但看不出对量化信噪比性能有明显影响。自适应量化器实现时采用一种自适应增益控制机制,针对最常用的一个变量,通过逐个样本反馈自适应的方式进行。

如果随机变量 X 的幅度服从非均匀分布,对于给定数量的量化区间 I ,均匀量化器达不到最小的量化信噪比性能。最优量化器门限值的集合 $t_i(i = 1, 2, \cdots, I)$ 和重构值 $x^{(q)}(i) (i = 1, 2, \cdots, I)$ 可以采用迭代的 Lloyd-Max 算法得到,该算法交替使用下面的两个隐含条件,对于固定输入功率而言,这两个条件使均方量化误差最小。

- 1. 最佳量化区间的端点为相邻重构值的中值:

$$t_i = \frac{x^{(q)}(i) + x^{(q)}(i - 1)}{2}$$

(15.10)

- 2. 最佳重构值为量化区间的质心,该值为通过输入幅度概率密度 $f_X(x)$ 计算得到的条件数学期望值:

$$x^{(q)}(i) = E(x|t_i \leq x < t_{i+1}) = \int_{t_i}^{t_{i+1}} f_X(x)x \, dx$$

(15.11)

实际上,概率分布往往是未知的,质心的计算是通过估计实际试验中采集的大量数据样本集合的类条件均值获得的。

15.4.2 矢量量化

矢量量化(VQ)将 d 个随机变量标量 X_k 组成一个 d 维的随机矢量 $\mathbf{X}$ 。在 d 维空间中,标量量化区间变成了多维的凸胞腔 $V_i(i = 1, \cdots, I)$ 。如果一个输入矢量 $\mathbf{x} \in V_i$,那么 $Q(\mathbf{x}) = i$,值 i 就映射成一个二进制码字用来传输。Shannon[1949]表明,采用矢量量化并渐进地增加维数 d ,就能达到信源编码的率失真限,就像信道编码中,当同时考虑很大的信号样本分组时,可以得到最好的编码性能。与标量量化相比,矢量量化有 3 个共知的优点。为了进行这种对比,下面研究达到特定的量化失真时每维需要的比特数 $\log_2(I)/d$ 。

- 1. 存储优势。即对统计相关性的处理。一般地,一个随机矢量表现出各个分量之间的统计相关性,可能超出线性相关。所以,即使采用线性变换,如离散余弦变换或主分量分析,对随机矢量的分量化解相关,由于包含在剩余(非线性)相关中的冗余,所有分量的联合量化仍然更有效。
- 2. 空间填充优势。在标量量化中,量化区间的形状没有那么多选择,在多维情况下,胞腔可以变换形状以达到最佳空间填充度。例如,根据密集球装问题(dense sphere-packing)准则。这允许对一个给定容积的空间覆盖,在满足相同最大失真(通过指定胞腔的线性维的大小)的情况下采用较少的胞腔,即更少的码字比特数 $\log_2(I)$ 。对于大

① 实际上,对于对数量化,量化信噪比基本独立于输入的幅度概率分布,这使它对未知的或剧烈变化的信号特性表现出稳健性。

维数 d 的情况, 这种优势可以达到每维 0.25 比特。对于在 4 kbps 或以下的低比特语音编码(即每个样本小于 0.5 比特), 这是一个非常有意义的贡献。

3. 形状优势。这和标量变量的最佳非均匀量化得到的增益相似, 量化器胞腔的形状(和大小)与随机矢量的概率分布相匹配。这种优势仅和预先指定胞腔数目 I 的限定分辨率的量化器有关。对于限定熵的量化器(这里仅关心所有码字的平均熵, 既不限定胞腔或码字的个数, 也不限定码字的长度), 就没有形状优势, 这一点可以通过码字本身的无损熵编码得到。

最佳矢量量化器的设计与标量非均匀量化器的设计相似。Lloyd-Max 算法推广到多维, 称为 Linde-Buzo-Gray 或者 LBG 算法。

次最佳矢量量化

矢量量化维数 d 越大, 矢量量化性能就越好。然而, 如果设计一个矢量量化器, 给定每维的比特数, 则矢量量化的复杂度将随维数 d 呈指数增长, 因为胞腔数 I 随 d 呈指数增长。由于多维的胞腔可能形状很复杂, 一般用它们的重构矢量 $\hat{x}_q(i)$ 来定义并通过选择使量化失真 $D(x_q(i), x)$ ($i = 1, 2, \dots, I$) 最小的索引号 $i = Q(x)$ 进行量化。一种经典的失真测度是平方失真或者欧氏距离 $\|x_q(i) - x\|^2$ 。这样就需要保存一个表格或者码书, 该码书包含 I 个 d 维的矢量, 需要搜索整个码书以获得最小失真, 这需要 I 次失真测度的计算。在欧氏距离情况下, 每次的计算代价与 d 成正比。因此, 即使对每一维, 存储代价和操作次数也直接正比于 I , 所以存在所谓的“维数灾难”问题(复杂度与维数呈指数增长)。

为了解决这个问题, 发展出很多矢量量化器设计和实现的简化策略。出乎意料的是, 这些在次最佳码书集合中搜索一个好矢量的次最佳算法, 其性能仍好于低维数的最优矢量量化系统。在早期基于矢量量化的波形编码中, 当维数为 10 时, 采用在 $I = 2^{10} = 1024$ 个条目中全搜索, 编码器可达每维 1 比特, 新的次优系统在维数 $d = 40$ 时, 采用贪心算法在 $I = 2^{35} = 34.4 \times 10^9$ 个条目中搜索可达每维 0.875 比特, 而且性能更佳^①。这种简化的次最佳设计的例子如下:

- 增益形状量化, 码书矢量为一个标量增益和一个单元矢量的乘积(模为 1);
- 多阶矢量量化, 第一个码书提供粗略估计, 后续 $K-1$ 个码书依次细化量化;
- 分裂矢量量化, 高维矢量分成 K 段 d/K 维的矢量, 每段独立量化。

在后两种情况下, 复杂度约减少到 d/K 维矢量量化的复杂度。对于这种比较而言, 非指数因子的关系不大。

下面讨论代数码书, 这是最成功的简化方案之一, 已经被几个国际标准采纳。在代数码书中, 码书幅度限制在一个三元表 $\{-1, 0, +1\}$ 中, 而且只允许少量的非零幅度(例如, 在 40 维矢量中允许 10 个)。这种设计受到早期的“多脉冲激励”和“规则脉冲激励”技术的启发, 在“多脉冲激励”和“规则脉冲激励”技术中, 源滤波器模型的源激励被裁剪成少数非零幅度。因为, 在求取内积的计算中包含欧氏距离计算, 代数结构将每次计算从 40 个乘加运算减少到只有 10 个加运算。而且, 非零脉冲的选取在贪心策略之后进行, 不需要尝试所有可能的组合。最后, 采用几个子采样的脉冲序列或者多相的交织梳进行高效编码(交织单脉冲排列, ISPP)。

^① 注意, 与先前的 10 维码书相比, 40 维码书的全搜索将带来复杂度的增加, 每维增加的复杂度相当于乘以一个 $2^{23} = 8.4 \times 10^6$ 的因子。

注意,这种码书采用系统结构建立,可以较均匀地覆盖整个矢量空间,因此不需要训练或者优化码字矢量。

线谱对

当应用矢量量化器来量化语音模型的参数(非波形样本)时,需要特别考虑参数量化误差引起的失真。一种常用的应用是同时量化 $m = 10$ 维的线性预测系数 a_i ,这样会引起如下问题。

1. 量化参数矢量必须对应一个最小相位滤波器,该滤波器允许稳定的滤波及逆滤波操作。
2. 量化误差最好对应引入的谱失真,即原始谱包络和由量化参数描述的谱包络之间的均方误差,二者均用 dB 表示。经验表明,为了达到透明编码质量,对大多数参数矢量而言,这个失真值必须小于 1 dB。
3. 在短时平稳模型中,从一帧到下一帧的转换必须经由两个参数矢量采用简单的内插机制,使它们能一直保证平稳性,并且避免内插的谱包络出现意外谱偏移。

所有这些问题最好在矢量量化之前,即进行预测器系数到线谱频率非线性映射时解决,这种线谱频率明显成对出现,因此也称为线谱对(LSP)。通过引入两个系数分别为偶对称和奇对称的 $m+1$ 阶多项式 $F_1(z)$ 和 $F_2(z)$,可以导出线谱对系数:

$$F_1(z) = A(z) + z^{-1} \cdot z^{-m} A(z^{-1}) \quad (15.12)$$

$$F_2(z) = A(z) - z^{-1} \cdot z^{-m} A(z^{-1}) \quad (15.13)$$

其中, $z^{-m} A(z^{-1}) = \sum_{i=1}^m a_{m-i} z^{-i}$ 为时间反转预测器多项式。由于 $A(z) = (F_1(z) + F_2(z))/2$, 这第一步是可逆的。然而,对最小相位滤波器 $A(z)$, 镜像多项式 $F_1(z)$ 和 $F_2(z)$ 的根不在单位圆内,而是正好在单位圆上,因此它们的位置可以由一个实值频率完全确定。而且,两个多项式的根位置集合遵循一个联合排序关系:随着频率的增加,这两个根集合在单位圆上严格交替出现。这些性质大大简化了多项式根的搜索并指明了相邻线谱频率对的组合(各来自一个镜像多项式)。并且,这种排序关系有助于估计线谱对在时间上的插值。最后,这种矢量由 10 个实数组成,它的矢量量化经常采用分裂矢量量化来实现。分裂矢量量化包含按频率增加而排序的三部分,而这三部分对矢量的贡献相对独立并将次最优量化效应降到最低。

15.4.3 预编码中的噪声成形

混合语音编码不完全取决于源信号模型,另外还要观察和传输建模误差或残留波形。如果用预测信号模型,残留波形就是预测误差 $e_n = x_n - \hat{x}_n$ 。如果量化预测残差 e_n 用来传输,则得到 $e_n^{(q)} = e_n + q_n$, 据此解码器可以重构 $x_n^{(d)} = \hat{x}_n + e_n^{(q)}$ 。译码器从延迟的重构信号样本 $x_{n-i}^{(d)}$ 计算预测值 $\hat{x}_n$:

$$\hat{x}_n = - \sum_{i=1}^m a_i x_{n-i}^{(d)} \quad (15.14)$$

则导致一个递归的或者闭环重构滤波器:

$$x_n^{(d)} = - \sum_{i=1}^m a_i x_{n-i}^{(d)} + e_n^{(q)} \quad (15.15)$$

$$A(z)X^{(d)}(z) = E^{(q)}(z) = E(z) + Q(z) \quad (15.16)$$

然而,在编码器端计算预测信号样本 $\hat{x}_n$ 时,可以采用如下两种不同的方法。

1. 开环预测, 利用延迟的原始信号样本, 这些样本只有在传输端能够得到:

$$\hat{x}_n = - \sum_{i=1}^m a_i x_{n-i} \quad (15.17)$$

$$\hat{x}_n = - \sum_{i=1}^m a_i x_{n-i} \quad (15.18)$$

从上式及式(15.16), 可得到解码器输出:

$$X^{(d)}(z) = X(z) + \frac{Q(z)}{A(z)} \quad (15.19)$$

2. 闭环预测, 在计算编码器中预测值时, 利用与解码器[见式(15.14)]一样的重构样本, 得到

$$e_n = x_n - \hat{x}_n = x_n + \sum_{i=1}^m a_i x_{n-i}^{(d)} \quad (15.20)$$

因此, 解码器输出为

$$x_n^{(d)} = - \sum_{i=1}^m a_i x_{n-i}^{(d)} + e_n^{(q)} \quad (15.21)$$

$$= - \sum_{i=1}^m a_i x_{n-i}^{(d)} + x_n + \sum_{i=1}^m a_i x_{n-i}^{(d)} + q_n \quad (15.22)$$

$$X^{(d)}(z) = X(z) + Q(z) \quad (15.23)$$

在这种设置中, 编码器预测器完全采用了解码器的闭环结构, 故得名。

量化噪声 $Q(z)$ 用白色频谱建模。根据式(15.19), 开环预测导致总体失真 $X^{(d)}(z) - X(z)$, 其频谱形状与语音模型的频谱 $1/A(z)$ 相似, 这将降低该噪声的可闻度, 因为人类听觉存在掩蔽效应特性。然而, 把噪声用 $1/A(z)$ 滤波也会得到一个总体增益, 该增益使总的噪声功率和不采用预测直接进行语音波形量化所观察到的功率完全相同。根据式(15.23), 闭环预测保持白噪声频谱且噪声不会放大。然而, 由于量化噪声频谱形状与语音频谱不同, 后者在频谱的谷底处可能听得到。上述两种极端情况的一种折中可以参见采用感知噪声成形滤波器的方法 [Schroeder et al. 1979]。至此, 量化噪声可看成量化器输出和输入之间的差 $q_n = e_n^{(q)} - e_n$, 该噪声经过滤波反馈到量化器的输出 $E(z) \rightarrow E(z) + (W(z) - 1)Q(z)$ (其中零延迟增益为 $w_0 = 1$)。因此, 开环预测系统的解码器输出等于 $X^{(d)}(z) = X(z) + \frac{W(z)}{A(z)}Q(z)$ 。如果选择加权滤波器 $W(z) = A(z/\gamma)$, 其中 $0 < \gamma \leq 1$, 就可以控制解码器失真的频谱形状在白色谱 ($\gamma = 1$) 和语音模型频谱 ($\gamma = 0$) 之间的任意位置。注意, 这种感知加权滤波器的频谱扩展可以简单地通过设置 $w_i = a_i \gamma^i$ 得到。

长时预测

到现在为止讨论的短时预测器(STP)仅仅针对相邻信号样本之间的统计依赖(短时相关), 这和频谱包络的非平坦性(即语音的共振峰结构)相关。浊音的谐波结构可以导致另一种对应于语音信号长时相关的精细结构。这些相关性在准周期信号波形的重复结构中可以看到。最简单的长时预测器(LTP) $B(z)$ 从位于过去一个周期 T_0 前的样本预测当前样本 x_n :

$$\hat{x}_n = b \cdot x_{n-T_0} \quad (15.24)$$

由于短时预测器和长时预测器均为稳定的线性系统,可按任意顺序级联,但经常是先执行短时预测,后执行长时预测更好一些。对于长时预测器,一般优先选择闭环预测。最佳参数 b 和 T_0 可以通过分析从最短的基音周期(高音女声)到最长的周期(低音男声)范围的信号自相关得到。即使在某些情况下得不到真正的基音周期(比如,周期倍增),长时预测器仍然可以通过提取与自相关函数最大值相关的冗余来提高编码器的性能。

只要真正的基音频率 T_0 不是采样间隔的 T_s 的整数倍(“分数基音”),长时预测器周期模型就会遇到一个特殊问题。作为一种典型处理,采用 3~6 倍的因子进行插值可以提高有效采样分辨率。另一种处理方法是提高长时预测滤波器 $B(z)$ 的阶数,这样可以同时解决信号插值和预测的问题。

15.4.4 合成分析

合成分析过程是在多脉冲激励线性预测编码中引入的[Atal and Remde 1982],稍后在文献 Atal and Schroeder[1984]和 Schroeder and Atal[1985]中将其与矢量量化结合在一起,形成码激励线性预测(CELP)。本质上,带有感知加权的开环预测/闭环预测准则的重新表示中,采用线性合成步骤代替编码器的线性预测分析步骤即体现了合成分析的原理。这样,潜在的信号模型的基本结构变得更直接和明显,当与矢量量化结合时有利于计算。注意,这种合成分析形式仍然严格等价于前面提到的噪声成形预测编码。

合成分析法首先从一个允许值的集合(码书)中选取一个量化的残留信号样本 $e_n^{(q)}$ 来产生或者合成一个(或多个)信号样本。为了达到这个目的,码书中的所有样本都要通过线性预测合成滤波器 $1/A(z)$ 来产生一个所有候选编码信号 $x_n^{(d)}$ 的集合。编码失真可以通过差值 $x_n^{(d)} - x_n$ 计算。从上面可知,这种失真很明显与量化误差 q_n 通过滤波器 $W(z)/A(z)$ 成正比,编码失真通过逆加权滤波器 $A(z)/W(z)$ 滤波,就可以得到量化噪声 $q_n = e_n^{(q)} - e_n$ 。如果令该变量的平方幅度取最小值,就可以在码书的所有可能候选残差 $e_n^{(q)}$ 中有效进行最邻近判决,以寻找最接近 e_n 的条目。

这种合成分析解释的能力在于它本身能够对任意残留波形条目进行描述,只要这些条目能够列出来。这特别适合残留波形进行矢量量化的情况。作为一种典型情况,不是采用一个维数 $d=40$ 的单一样本,而是把样本组成子帧。现在感知加权合成误差的平方幅度估计是通过将 d 个样本的平方相加计算的。在矢量量化应用中,建议采用如下两个重新组织的步骤来降低计算复杂度且不牺牲性能。

1. 正如一次处理 d 个样本的块或者子帧,当所有滤波器从一个子帧转移到下一个子帧时,它们必须记忆其状态。这意味着滤波器对一个子帧的输出将是一个零输入响应(衰变滤波器状态)和零状态响应(当前输入驱动的分量)之和。只有后者分量提供从矢量量化码书中选择最佳码字的信息。因此,应从语音信号 x_n 中首先减去零输入响应,这样码书码矢量选取只基于零状态响应。
2. 对于每个码矢量的零状态响应,必须使该矢量通过线性预测合成滤波器 $1/A(z)$ 和(逆)加权滤波器 $A(z)/W(z)$ 。这种滤波器级联可以简化为总的传输函数 $1/W(z)$,从而得到图 15.4 所示的原理框图。

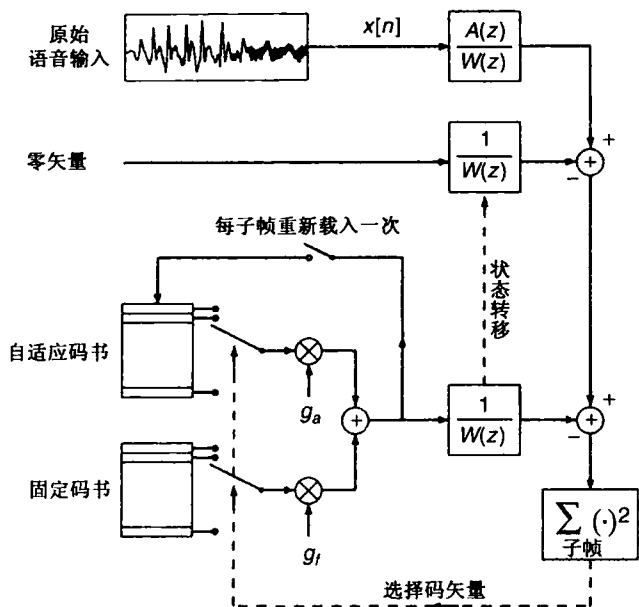


图 15.4 语音的码激励线性预测编码，将零输入和带有自适应与固定码书激励的零状态响应分开处理的框图。加权滤波器为 $W(z) = A(z/\gamma)$

自适应码书

合成分析规则还可以扩展到试图采用从时延约一个基音周期 T_0 的（规整）样本预测来解释当前子帧的长时预测器。由于基音周期可以在一定范围内变化，一种有效的处理是在缓冲区内存储残留信号自身在该范围内取得的子帧长度的矢量。该缓冲区称为自适应码书。这样，一个残留子帧可以表示为两个矢量的加权和，一个选自自适应码书，一个选自固定码书。在大多数系统中，这两种选取是顺序进行的，以使首先选出的自适应码书中的条目能够刻画信号的周期性，随后选出的固定码书条目能够表示残差部分的剩余的“随机性”（见图 15.4）。这种方法本质上与采用长时预测器等同，对于高音调的语音存在较小的差别，因为基音周期比子帧长度要短。与长时预测器相同的还包括，需要采用信号插值来提高自适应码书的时间分辨率才能保证高质量，这种插值运算至少要在长时相关的峰值附近进行。

15.4.5 信源信道联合编码

不等差错保护

一旦一个完整语音帧的所有信源参数都计算出来了，这些参数的码字需要组装起来进行传输，并且要加入冗余比特进行信道编码。信道编码通常假定码字序列独立同分布（iid），并且所有码字对接收机同等重要。对混合编码而言，后者显然不成立，因为有些码字表示模型参数，如基音频率、增益或线谱对（LSP），有些码字则表示采用固定矢量量化码书的码字表示后的残存波形。重要参数的信道误码（ T_0 可能是最重要的一个）将导致译码信号严重的波形误差，例如重建信号可能听起来比原信号低得多或者高得多。意义不太重要的码字上的信道误差可能几乎觉察不到，因为一个参数正确的语音模型能够产生出相当好的语音，即使残留波形的细节是错误的。总之，“所有信源比特是相等的，但有些比其他担负更重要的任务。”这导致

码字可以根据重要程度或者对信道敏感程度不同进行分类,每一类采用不同级别的信道保护,从最复杂的卷积码到根本不采用差错保护。

自适应多速率编码

信道和信源的统计特性都是时变的。因此,优化编码策略不得不根据时变信道状态和信源状态进行自适应调整。对电路交换通信而言,信道需要总的速率是固定的,因此自适应机制的任务是在信源编码速率和信道差错保护之间确定一个最佳折中,以使总的速率保持不变,即自适应多速率(AMR)编码。因为信源统计特性要比信道状态信息更新快得多(基本上每帧都不同,比如每 20 ms),当前的系统大多仅仅根据信道状态信息选择最佳的信源/信道速率划分。依靠这种信息,信源编码器对每一帧首先产生不同的信源比特,信道编码器利用剩余的比特进行恰当的不等差错保护,方法如上一段所述。

信源编码优化

按照信源信道联合编码方法,可采用如下两种方法对信源编码进行优化。

- 编码器优化。例如优化矢量量化的索引分配,使用代表矢量量化索引的码字中的比特误差产生最小平均感知影响,这和标量索引分配中的格雷编码策略非常相似。多描述编码是一种试图解决所有码字等相关性的更进一步的编码器技术。
- 解码器优化。接收到码字的残留冗余用来协助信道译码器增加整个解码过程的可靠性。

15.5 从语音传输到声音的远程呈现

语音传输或者电话技术的终极目标一直是在一个地理上分离的地点再现说话人的声音。从这个意义上讲,电话仅仅是远程呈现的一种特殊形式,而远程呈现是要虚拟呈现一个或者多个远程人物,最好从一定程度上再现远程的环境来增强本地的现实逼真度。下文的简要讨论首先从语音传输的简单附加功能开始,逐渐发展到最先进的三维虚拟音频业务。

15.5.1 语音激活检测

在两个人之间对称的谈话中,每个人约有 50% 的时间是不讲话的。人们早就注意到这个事实,并且在同一个传输信道中进行电话对话复用利用了利用了这个特性,这种技术称为数字语音插空(DSI)。对无线通信而言,多址接入到同一个共享无线频谱时需要降低用户之间的干扰。这可以通过仅在讲话人正在说话时才在空中接口传输语音帧来实现,而讲话状态由语音激活检测器(VAD)检测。对一个不活跃的说话人,非连续传输(DTX)将导致:

1. 更低的移动终端的功耗(节约了基带处理和无线电传输功率),从而延长了电池寿命;
2. 更小的空中接口多用户干扰,从而提高了移动接入网络性能;
3. 在采用包交换的情况下更低的网络负载,从而提高了骨干网的容量。

在一个典型的实现中,语音传输在 7 个非活动帧后被切断,然后发送一个静寂描述(SID)帧,每 20 ms 只有 35 比特,作为声音背景噪声的模型,以便在接收端再生舒适噪声生成(CNG)。这种噪声模型至少每 24 帧更新一次,可看成听觉场景的虚拟现实呈现的第一步,而通过使对方隐晦地知道这是开车时的通话等保持了通信双方的一致性。

这种非连续传输方法有时看成信源控制的速率自适应,它工作在非常高的信源描述层(即仅是信源开和信源关),比如不能根据语音内容(如浊音对清音)来快速改变信源速率。这和前面描述的自适应多速率编码(结合话音激活检测器,信源速率为零或者当前允许的最大信源速率,还要根据信道状态变化)也不相同。

15.5.2 接收端增强

如果由于空中接口的严重衰落导致整个语音帧丢失,那么这种错误不可能通过信道编码来纠正,因为信道编码只是针对个别的或者突发错误。这种情况下,需要错误隐藏技术,利用接收到的邻近的帧通过插值得到丢失的帧。对第一个丢失的帧而言,这种方法通常工作得很好,但如果有多连续丢失的帧,用这种方法可能产生发音不自然的语音。对后面的这种情况,后续的替代帧将逐渐衰减,而达到信号幅度慢慢减弱的效果(通常需要6帧=120 ms)。

一种更进一步的接收端信号增强为自适应后滤波,根据频谱包络构建成形滤波器和长时预测滤波器,以使由于有失真编码引入的噪声不易被收听者听见。自适应播放缓冲区或者抖动缓冲区有益于抵消包交换网络中的丢失帧。如果一个终端设备具有宽带语音输入/输出功能(模拟带宽为50 Hz~7 kHz),从窄带网络接收到的语音可能通过一种称为带宽扩展的合成机制,人工创建3.4 kHz以上和300 Hz以下的频带而进行增强。

15.5.3 回声和噪声

说话人的声学环境可能不仅包括有利于表现声学场景的有用环境信息,还会包含烦人的噪声,或者由于多个说话人采用免提电话造成事实上共用一个空间而产生的回响和声音反射。在这种情况下,不仅需要控制噪声以让人感到舒适,而且语音编码系统(如果非常强地依赖一个语音信号模型)将无法恰当地处理噪声和回声。回声和噪声联合控制的解决方案通常能够达到降低这两种损伤的最佳折中。未经控制的回声可能会对整个通信系统的稳定性产生严重的影响。严格的性能要求已经由ITU-T制定为标准,如建议采用G.167和G.168。

15.5.4 远程呈现的业务增强

语音使能业务

在远程呈现的框架下,用户将能够接入比传统电话范围宽得多的业务。对于超小尺寸的移动终端,如果机器有语音对话功能,将会大大方便业务的接入。其范围包括简单的通过语音进行人名拨号(可能需要一个话音密语来激活),通过文本语音合成(TTS)系统来阅读电子邮件,或者通过分布式语音识别(DSR)在移动终端提取语音识别用的特征矢量,并进行精度足以进行模式识别的编码(但没必要能够在中心服务器端再生语音信号),然后作为数据在无线链路上传输。这样就能传输基于超过电话中用的高质量信号的识别特征,并且采用特定的信源信道编码机制达到最有利于远端语音识别服务器的目的(而不仅仅面向收听的人)。

分布式语音识别采用数据信道携带语音信息,而蜂窝文本电话(CTT)的应用允许在语音信道携带文本数据,以便提供增强通信手段。这样,有听觉或发音障碍的人就能利用工作在数字语音信道上的调制解调器,采用聊天方式交换交互文本(不仅仅是短信息)。

个性化

对于个性化服务,说话人认证是必须的,而这有别于移动站或者基础设施本身的认证。说

话人的身份可以通过语音辨识或确认技术确定,而且,如果需要,语音编码前可以在语音信号中加一个个性化的水印。

对于说话人隐私,传统的公用电话亭可能会被虚拟讲话区所替代。在该区域外,活动语音抵消(使用可穿着扬声器阵列)将使旁边的人几乎听不到电话交谈的声音。

三维音频

虚拟讲话区已经涉及到语音电话学的高级音频处理的概念。自从采用了麦克风和扬声器阵列及虚拟增强音频技术之后,高级音频处理获得极大的推动,它允许三维声场的空间呈现(例如,立体混响)潜在地转换成双声道耳机收听。如果将个性化的头部相关传输函数(HRTF)和头部实时跟踪相结合,便能够在真实环境的背景中放置虚拟的声源,创造出浸入式的远程呈现,从而把远程虚拟的参加者和本地参加者融合在一起,成功地召开远程会议,这当然是期望的最佳效果。

从根本上来讲,这需要进行重大的转变,从现在的语音编码标准转换到允许高质量和包含元数据信息的多通路音频。在这个方向上的第一步已经通过自适应多速率宽带语音编码器标准化实现,这是第一个几乎同时被有线通信(ITU)、无线通信(欧洲电信标准协会/第三代合作伙伴计划 ETSI/3GPP)和网络电话(互联网工程任务组 IETF)接受的语音编码标准。

深入阅读

语音编码的经典教材为 Jayant and Noll[1984]和 Kleijn and Paliwal[1995]。Vary et al.[1998]是一本用德语撰写的精彩教材,其改进的英文版于 2006 年已出版。最近, Kondo[2004]的第二版已经出版。另一个语音信号处理的各个方面的精彩综述为 Rabiner[1994]。Berger[1971], Gray[1989]和 Kleijn[2005]论述了信源编码理论,可能是研究语音编码最配套的材料。对于一般的语音信号处理,我们推荐经典的 Rabiner and Schafer[1978]和更近的教材 Deller et al.[2000], O'Shaughnessy[2000]和 Quatieri[2002]。对于非线性语音模型的扩展,参见 Chollet et al.[2005]和 Kubin[1995]。Gersho and Gray[1992]对矢量量化进行了广泛的讨论。Hanzo et al.[2001]主要论述无线语音传输的信源信道联合编码。通信中的声学信号处理的更宽泛的讨论可以参阅 Gay and Benesty[2000], Haensler and Schmidt[2004]和 Vaseghi[2000]。带宽扩展是 Larsen and Aarts[2004]的主要论题。口语处理的极好的参考书是 Huang et al.[2001]和 Jurafsky and Martin[2000]。Gibson et al.[1998]则处理一般多媒体数据的压缩。

第16章 均衡器

16.1 引言

16.1.1 时域和频域均衡

无线信道会产生时延色散,也就是从发射机到接收机,各多径分量会有不同的传输时间(见第6章和第7章)。时延色散会导致码间干扰,这将严重影响数字信号的传输。在第12章已经看到,即使时延扩展小于码元持续时间,误比特率性能也将明显地降低。正如第二代和第三代蜂窝系统中经常出现的,时延扩展与码元持续时间相当,或大于码元持续时间,若不采取对抗措施,则误比特率将高得令人无法接受。虽然编码和分集技术可以减小由码间干扰带来的误码,但不能完全消除(见第13章和第14章)。另一方面,时延色散也可起到正面作用。由于不同多径分量的衰落是统计独立的,所以分解后的各多径分量就可以作为分集支路。接收机若能分离和利用分解后的多径分量,则时延色散就提供了延迟分集的可能性。由于传输函数是信道冲激响应关于傅里叶变换对 $t \rightarrow f$ 的傅里叶变换,所以延迟分集也被视为频率分集(见第13章)。

均衡器是工作于两种方式的接收机结构,它们能减小或消除码间干扰,同时又利用信道固有的延迟分集。均衡器的工作原理既可以在时域描述,也可以在频域描述。

在频域,要记住时延色散与频率选择性相对应。换句话说,就是码间干扰的产生是由于传输函数在所考虑的系统带宽内不为常数,因此均衡的目的就是校正由信道引入的失真,也就是使信道和均衡器的传输函数的乘积为常数^①。这可以用数学形式表示:令信源信号为 $s(t)$;它被送入一个信道冲激响应为 $h(t)$ 的无线信道(准静态),接收后再通过一个冲激响应为 $e(t)$ 的均衡器。进一步假设发送信号采用脉冲幅度调制(见第11章),则有

$$s(t) = \sum_i c_i g(t - iT) \quad (16.1)$$

其中, c_i 为复发送符号。要求

$$H(\omega)G(\omega)E(\omega) = \text{常量} \quad (16.2)$$

其中, $E(\omega)$, $H(\omega)$ 和 $G(\omega)$ 分别为 $e(t)$, $h(t)$ 和 $g(t)$ 的傅里叶变换。

在时域的等效公式要求接收信号在采样时刻无码间干扰。定义 $\eta(t)$ 为信道冲激响应 $h(t)$ 和基带脉冲 $g(t)$ 的卷积:

$$\eta(t) = h(t) * g(t) \quad (16.3)$$

因而要求

$$[\eta(t) * e(t)]_{t=iT_s} = \begin{cases} 1, & i=0 \\ 0, & \text{其他} \end{cases} \quad (16.4)$$

如果信道特性已知,并且不随时间变化,就可以构造一个滤波器(用硬件),使之完成传输函数所要求的均衡。然而在无线通信中,信道是未知的和时变的。前者可以通过发送训练序列,即一个已知的比特序列来解决。根据接收信号 $r(t)$ 和基带脉冲 $g(t)$ 的形状信息,接收机可

^① 实际上,这是均衡器的一种特殊形式,即所谓的“迫零”线性均衡器,将在下面详细讨论。

以估计信道的冲激响应 $h(t)$ 。时变问题可以通过在“足够短”的时间间隔内重复发送训练序列来解决,因此在一定时间间隔内,均衡器能够随着信道状态的变化自行调整,这就是通常所说的“自适应均衡”。

这些年已研究出很多种不同的均衡器。其中最简单的是线性均衡器,它通常是一个带有抽头的延迟线滤波器,其系数与信道状态相适应(见 16.2 节)。判决反馈滤波器从接收信号中计算出(并减去)过去符号引起的码间干扰(见 16.3 节)。时延色散信道中的最佳检测方法是最大似然序列估计(见 16.4 节)。不需要训练序列的盲均衡,虽然已被许多研究人员进行了深入研究,但在实际系统中并不常用。

16.1.2 信道和均衡器的模型

在下面的章节中,为了描述不同的均衡器结构,需要建立信道和均衡器的离散时间模型。现在给出一个这样的模型,并且给出一个对于最佳接收机来说非常重要的噪声白化滤波器的概念。

接收机的第一级是用来抑制噪声功率的滤波器。该滤波器应确保所有信息都包含在 $t_i + iT_s$ 时刻的采样值里。这可以通过一个与 $\eta(t)$ 相匹配的滤波器来实现, $\eta(t)$ 是信道冲激响应和基带脉冲的卷积。因此,匹配滤波器输出端的样值序列可表示为

$$\psi_i = c_i \zeta_0 + \sum_{n \neq i} c_n \zeta_{i-n} + \hat{n}_i$$

(16.5)

其中, ζ_i 是 $\eta(t)$ 的自相关函数的采样值, $\hat{n}_i$ 是一个自相关函数为 $N_0 \zeta_i$ 的复高斯随机变量序列,也就是匹配滤波器的输出噪声。

自相关函数的采样值 ζ_i 的 z 变换可以分解为

$$\Xi(z) = F(z)F^*(z^{-1})$$

(16.6)

这个分解不是唯一的。但是,选择所有根均在单位圆内的 $F^*(z^{-1})$ 是很有利的。在这种情况下, $1/F^*(z^{-1})$ 是一个稳定且可实现的滤波器。

现在,如果把匹配滤波器和 $1/F^*(z^{-1})$ 级联,这个级联输出端的噪声就又变成白的,并且自相关函数为 $N_0 \delta_k$,因此它也称为噪声白化滤波器。基带脉冲、无线信道、匹配滤波器和噪声白化滤波器级联的冲激响应的采样值则称为离散时间信道。它的冲激响应记为 f_k , z 变换记为 $F(z)$ 。注意,该信道的冲激响应是因果的,因此输出信号可以写为

$$u_i = \sum_{n=0}^{L_c} f_n c_{i-n} + n_i$$

(16.7)

其中, L_c 是离散时间信道冲激响应的长度。因此噪声白化滤波器也经常称为前置均衡器。

这一章将使用到以下概念:

c_i	第 i 个复发送符号
$\eta(t)$	基带脉冲和信道冲激响应的卷积
e_i	均衡器的第 i 个系数
ζ_i	$\eta(t)$ 的自相关函数的第 i 个采样值
$\hat{n}_i$	自相关函数为 $N_0 \zeta_i$ 的复高斯随机变量序列的第 i 个采样值
n_i	非相关复高斯随机变量序列的第 i 个采样值
f_i	时间离散信道冲激响应的第 i 个采样值
u_i	时间离散信道输出信号的第 i 个采样值
$\hat{c}_i$	发送符号 c_i 的估计值
ε_i	发送符号的估计值与真实值 $c_i - \hat{c}_i$ 的偏差

16.1.3 信道估计

对于带有均衡器的数据检测,一个常用的方法是将 $\mathbf{f}$ 和 $\mathbf{c}$ 的估计分离。第一步,用训练序列(也就是已知的 $\mathbf{c}$)来估计 $\mathbf{f}$ 。在随后未知的有效数据传输期间,假设估计的冲激响应就是真实值,然后求解以上关于 $\mathbf{c}$ 的方程。

这一节讨论通过训练序列来估计信道冲激响应的问题。信道估计和第8章介绍的“信道探测”技术有很大的相似性。通过发送周期为 N_{per} 的伪随机序列,就可以得到一个简单的信道估计。伪随机序列的自相关函数近似于一个狄拉克函数。更确切地讲,序列 $\{b_i\}$ 的周期延拓与其自身的时间翻转序列的卷积^①,就是间隔为 N_{per} 个符号的狄拉克脉冲之和:

$$\{b_{-i}\} *_{\text{per}} \{b_i\} \approx \sum_{n=-\infty}^{\infty} \delta_{i-nN_{\text{per}}} \quad (16.8)$$

其中, $*_{\text{per}}$ 表示循环卷积。如果这个序列通过冲激响应为 $\mathbf{f}$ 的信道,那么这个相关器的输出为

$$\{\hat{f}_i\} = (\{b_{-i}\} *_{\text{per}} \{b_i\}) * \{f_i\} \quad (16.9)$$

此时,如果信道冲激响应的持续时间比 N_{per} 短,那么就仅仅是 $\mathbf{f}$ 的一个周期重复(见图 16.1)。

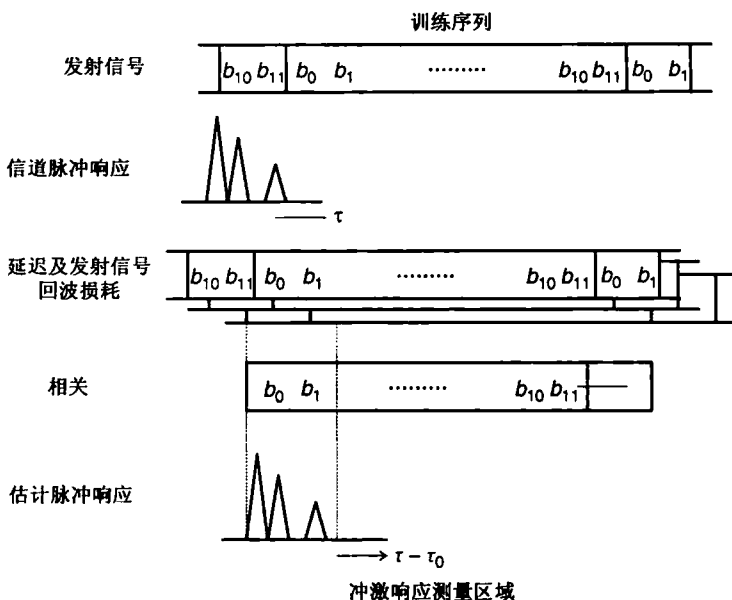


图 16.1 采用相关处理的信道估计器原理

实际上,我们并不发送伪随机序列的周期延拓序列,而仅是一个简单的实现。为了避免有效比特(未知)对相关器输出可能造成的影响,在该序列前后必须发送已知的缓冲比特。

采用训练序列的信道估计技术有以下几个缺点。

1. 频谱效率下降。因为训练序列不传输任何有用信息。例如,全球移动通信系统在每帧 148 比特中使用 26 比特作为训练序列(见第 21 章)。
2. 对噪声敏感。为了保证合理的频谱效率,训练序列必须很短。然而,这意味着训练序

^① 严格地讲,需要取时间翻转序列的复共扼,但是因为发送序列通常是实的($+1/-1$),所以比特和复符号可看成等效的。

列对噪声敏感(较长的训练序列可以对噪声进行平均),并且对探测序列的非理想性也敏感(因为伪随机序列的通断比随序列长度的增加会增大)。若通过迭代算法实现信道估计,则仅能采用快速收敛的迭代算法,但这种算法会导致高的剩余差错率。

3. 过时估计。如果在训练序列发送之后信道发生了改变,则接收机检测不出这种变化。使用一个过时的信道估计会导致判决错误。

虽然存在这些问题,但在每一个实际的系统中均使用这种基于训练序列的信道估计。原因是其他信道估计方法(如盲技术,见 16.7 节)需要很大的运算量,并且性能会受到其他大量数值因素的影响。

16.2 线性均衡器

线性均衡器是一种试图模拟信道传输函数反函数的简单线性滤波器,其作用是尽量使信道和均衡器的传输函数的乘积满足一定的准则。这个准则可以是使信道-滤波器级联的传输函数完全平坦,或者是使滤波器输出端的均方误差最小。

线性均衡器的基本结构如图 16.2 所示。根据 16.1.2 节的系统模型,发送序列 $\{c_i\}$ 经过一个存在色散和噪声的信道传输,那么在均衡器的输入端可得到序列 $\{u_i\}$ 。现在需要确定具有 $2K+1$ 个抽头的有限冲激响应滤波器的系数(横向滤波器,见图 16.3)。该滤波器应将序列 $\{u_i\}$ 转换成序列 $\{\hat{c}_i\}$:

$$\hat{c}_i = \sum_{n=-K}^K e_n u_{i-n} \tag{16.10}$$

它应该尽可能地接近序列 $\{c_i\}$ 。定义误差 ε_i 为

$$\varepsilon_i = c_i - \hat{c}_i \tag{16.11}$$

我们的目的是找到一个滤波器,使得

$$\varepsilon_i = 0, \quad N_0 = 0 \tag{16.12}$$

这样可以得到迫零(ZF)均衡器,或者

$$E\{|\varepsilon_i|^2\} \rightarrow \min, \quad \text{对于 } N_0 \text{ 取有限值} \tag{16.13}$$

这样可以得到最小均方误差(MMSE)均衡器。

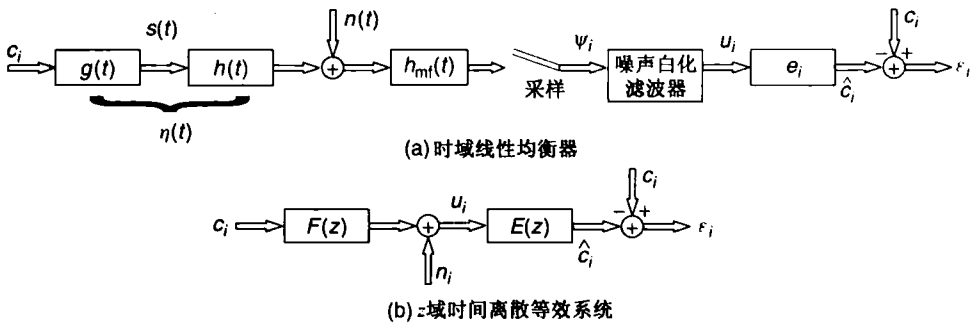
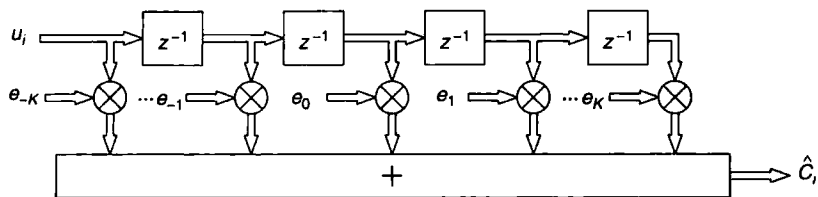


图 16.2 线性均衡器基本结构

图 16.3 线性横向滤波器结构。注意 z^{-1} 表示延迟一个采样间隔

16.2.1 迫零均衡器

从频域上看，迫零均衡器通过选择均衡器的传输函数为 $E(z) = 1/F(z)$ ，使得信道和均衡器组合的传输函数完全平坦（常数）。从时域上看，它使最大码间干扰取最小值（峰值畸变准则）。附录 16. A (www.wiley.com/go/molisch) 证明了这两个准则是相同的。

迫零均衡器对于消除码间干扰来说是最佳的。然而，信道也引入噪声，噪声将被均衡器放大。在信道传输函数取值很小的频率处，均衡器具有很强的放大作用，因而也将相应频率的噪声进行了放大。结果，在检测器输入端的噪声功率比无均衡器时更大（见图 16.4）。

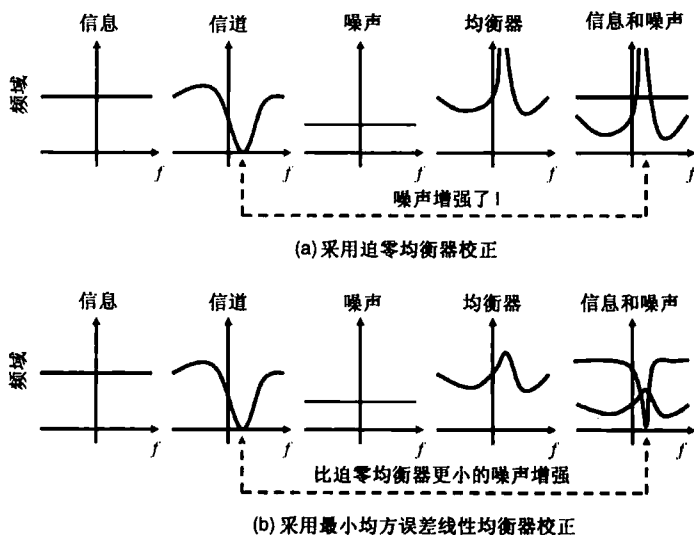


图 16.4 噪声增强的图示

自相关函数的采样值 ζ_i 的傅里叶变换 $\Xi(e^{j\omega T_s})$ 与 $\eta(t)$ 的傅里叶变换 $\hat{\eta}(e^{j\omega T})$ 的关系为

$$\Xi(e^{j\omega T_s}) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \left| \hat{\eta}\left(\omega + \frac{2\pi n}{T_s}\right) \right|^2, \quad |\omega| \leq \frac{\pi}{T_s} \quad (16.14)$$

检测器中的噪声功率为

$$\sigma_{n-\text{LE-ZF}}^2 = N_0 \frac{T_s}{2\pi} \int_{-\pi/T_s}^{\pi/T_s} \frac{1}{\Xi(e^{j\omega T_s})} d\omega \quad (16.15)$$

只有当频谱密度没有奇点 Ξ (或只是可积的) 时，它才是有限的。

16.2.2 最小均方误差准则

均衡器的最终目标不是使码间干扰最小，而是使误比特率最小。噪声增强使得迫零均衡

器达不到这个目的。一个更好的准则就是使发送信号和均衡器的输出信号之间的均方误差达到最小。

因此我们寻找一个滤波器使得下式最小:

$$\text{MSE} = E\{|\varepsilon_i|^2\} = E\{\varepsilon_i \varepsilon_i^*\} \quad (16.16)$$

如附录 16. B (www.wiley.com/go/molisch) 所示, 它可以由系数为 $\mathbf{e}_{\text{opt}}$ 的滤波器来实现:

$$\mathbf{e}_{\text{opt}} = \mathbf{R}^{-1} \mathbf{p} \quad (16.17)$$

其中, $\mathbf{R} = E\{\mathbf{u}_i^* \mathbf{u}_i^T\}$ 是接收信号的相关矩阵, $\mathbf{p} = E\{\mathbf{u}_i^* c_i\}$ 为接收信号和发送信号的互相关。考虑频域、噪声白化滤波器和均衡器 $E(z)$ 的级联的传输函数为

$$\tilde{E}(z) = \frac{1}{\Xi(z) + \frac{N_0}{\alpha_s^2}} \quad (16.18)$$

上式就是维纳滤波器的传输函数。因此, 均方误差为

$$\sigma_{\text{n-LE-MSE}}^2 = N_0 \frac{T_S}{2\pi} \int_{-\pi/T_S}^{\pi/T_S} \frac{1}{\Xi(e^{j\omega T_S}) + \frac{N_0}{\alpha_s^2}} d\omega \quad (16.19)$$

与式(16.15)比较表明: 最小均方误差均衡器的噪声功率小于迫零均衡器的噪声功率(见图 16.4)。

例 16.1 均衡器的系数和线性均衡器的噪声增强。考虑信道冲激响应为 $h(\tau) = 0.4\delta(\tau) - 0.7\delta(\tau - T_S) + 0.6\delta(\tau - 2T_S)$ 的信道, $N_0 = 0.3$, 并且 $g(t) = g_R(t, T_S)$ 。计算迫零均衡器和最小均方误差均衡器输出端的噪声方差。

解:

首先注意 $\eta(t) = h(t) * g(t) = 0.4g(t) - 0.7g(t - T_S) + 0.6g(t - 2T_S)$ 。对于给定的矩形脉冲有 $\sigma_s^2 = 1$ 。 $\eta(t)$ 自相关函数的 z 变换可表示为

$$\Xi(z) = 0.24z^2 - 0.7z + 1.01 - 0.7z^{-1} + 0.24z^{-2} = F(z)F^*(z^{-1})$$

然后选择 $F(z) = 0.4 - 0.7z^{-1} + 0.6z^{-2}$ 。因此, $F^*(z^{-1}) = 0.4 - 0.7z' + 0.6z'^2$ 的根为 $0.58 \pm 0.57j$, 它们均在单位圆之内。

因此, 迫零均衡器的传输函数为

$$E_1(z) = \frac{1}{F(z)} = \frac{1}{0.4 - 0.7z^{-1} + 0.6z^{-2}} \quad (16.20)$$

根据式(16.15), 噪声方差为

$$\begin{aligned} \sigma_{\text{n-LE-ZF}}^2 &= N_0 \frac{T_S}{2\pi} \int_{-\pi/T_S}^{\pi/T_S} \frac{1}{0.24e^{j2\omega T_S} - 0.7e^{j\omega T_S} + 1.01 - 0.7e^{-j\omega T_S} + 0.24e^{-j2\omega T_S}} d\omega \\ &= N_0 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{0.24e^{j2\omega} + 0.24e^{-j2\omega} - 0.7e^{j\omega} - 0.7e^{-j\omega} + 1.01} d\omega \\ &= N_0 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{0.48 \cos 2\omega - 1.4 \cos \omega + 1.01} d\omega \\ &\approx 2.94 \end{aligned} \quad (16.21)$$

因此, 有效信噪比为 $1/\sigma_{\text{n-LE-ZF}}^2 = 0.34$ 。

最小均方误差均衡器表示为

$$E_2(z) = \frac{F^*(1/z)}{F(z)F^*(1/z) + N_0} = \frac{0.4 - 0.7z + 0.6z^2}{0.24z^2 - 0.7z + 1.31 - 0.7z^{-1} + 0.24z^{-2}} \quad (16.22)$$

式(16.19)给出了其噪声方差。注意,与迫零均衡器的唯一不同是积分项中的分母增加了一个 N_0/σ_s^2 :

$$\sigma_{n\text{-LE-MSE}}^2 = N_0 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{0.48 \cos 2\omega - 1.4 \cos \omega + 1.31} d\omega \quad (16.23)$$

$$\approx 0.46$$

正如所期望的,最小均方误差均衡器的噪声方差小于迫零均衡器的噪声方差。有效信噪比[Proakis 2005]为

$$\gamma_{\infty} = \frac{1 - \sigma_{n\text{-LE-MSE}}^2}{\sigma_{n\text{-LE-MSE}}^2} = 1.17 \quad (16.24)$$

与无码间干扰(因此不需要均衡器)的情况 $1/N_0 = 3.33$ 相比,可得出:最小均方误差均衡器和迫零均衡器分别使有效信噪比(在均衡器输出端的信噪比)下降了 4.5 dB 和 10 dB。

16.2.3 均方误差均衡器的自适应算法

为了确定均衡器的最佳权重,可以直接解方程(16.17)。然而,这需要 $(2K+1)^3$ 阶的复操作。为了简化计算复杂度,通常采用迭代算法。迭代算法的性能用以下标准衡量。

- 收敛速度。接近最后的结果需要多少步迭代?通常假设在迭代期间信道保持不变,然而如果一个算法收敛太慢,由于在算法收敛以前信道就已经改变了,所以它将不可能达到一个稳定状态。
- 失调。迭代算法的收敛值和精确的均方误差解之间偏差的大小。
- 每一步迭代的计算量。

下面讨论两种广泛采用的算法:最小均方算法和递归最小二乘算法。

最小均方算法

最小均方(LMS)算法,也称为随机梯度算法,由以下几步构成。

1. 用 $\mathbf{e}_0$ 对权值进行初始化。
2. 利用这个值,计算均方误差梯度的一个近似值。真实的梯度值不可能得到,因为它是一个期望值。相反地,我们使用 $\mathbf{R}$ 和 $\mathbf{p}$ 的估计值(它们的瞬时实现):

$$\hat{\mathbf{R}}_n = \mathbf{u}_n^* \mathbf{u}_n^T \quad (16.25)$$

$$\hat{\mathbf{p}}_n = \mathbf{u}_n^* c_n \quad (16.26)$$

其中,下标 n 表示迭代次数。梯度被估计为

$$\hat{\mathbf{V}}_n = -2\hat{\mathbf{p}}_n + 2\hat{\mathbf{R}}_n \mathbf{e}_n \quad (16.27)$$

3. 接下来,通过在负梯度方向调整权值的方式来更新权值向量的估计:

$$\mathbf{e}_{n+1} = \mathbf{e}_n - \mu \hat{\mathbf{V}}_n \quad (16.28)$$

其中, μ 是一个用户自定义参数,它决定了收敛速度和剩余误差。

4. 如果满足了中止条件,例如权值向量的相对变化低于一个预先设定的阈值,那么这个算法就已经收敛。否则,返回第二步。

如果满足：

$$0 < \mu < \frac{2}{\lambda_{\max}} \tag{16.29}$$

则可以证明最小均方算法收敛，其中 $\lambda_{\max}$ 是相关矩阵 $\mathbf{R}$ 的最大特征值。问题是我们不知道该特征值(计算它比求相关矩阵的逆需要更大的计算量)。因此，必须猜测 μ 的数值。如果 μ 太大，收敛速度会比较快，但是算法有时会发散。如果选择的 μ 太小，那么收敛的概率将很大，但是非常慢。一般来说，收敛速度取决于相关矩阵的条件数(例如最大特征值和最小特征值的比)：条件数越大，最小均方算法的收敛速度就越慢。

递归最小二乘算法

在很多情况下，最小均方算法收敛非常慢。此外，该算法仅当接收信号的统计特性满足一定的条件时才有效。另一方面，一般的最小二乘准则并不需要这样的假设。它仅分析产生的 N 个误差 ε_i ，并选择权值使均方误差之和最小。这种一般的最小二乘问题也可以由递归算法来完成，称为递归最小二乘(RLS)算法。详细内容见附录 16. C(www.wiley.com/go/molisch)。

算法比较

确定均衡器的系数有两类算法：直接实现法(维纳滤波器，最小二乘准则)和迭代法(最小均方算法，递归最小二乘算法)。

对于维纳滤波器，必须首先确定相关矩阵；这是数值计算的主要部分，特别是当权值数很小时。实际矩阵的逆运算需要 $(2K + 1)^3$ 步操作。还可以直接使用数据矩阵，并且求它的逆(最小二乘算法)。数据矩阵的构造使其需要的数值计算比相关矩阵少；另一方面，求逆的运算量大得多，并且取决于所用的比特数。

比较迭代算法可以发现，最小均方算法通常收敛太慢。递归最小二乘算法收敛比较快，但是剩余误差却较大。图 16.5 给出了在增强型数字无绳电话(DECT) (www.wiley.com/go/molisch)中的均方误差的典型实例。可以看到，递归最小二乘算法在 10 比特之后就收敛，而最小均方算法需要将近 300 比特。考虑到信道的时变性和频谱效率，快速收敛比一个极小的剩余误差率更重要。

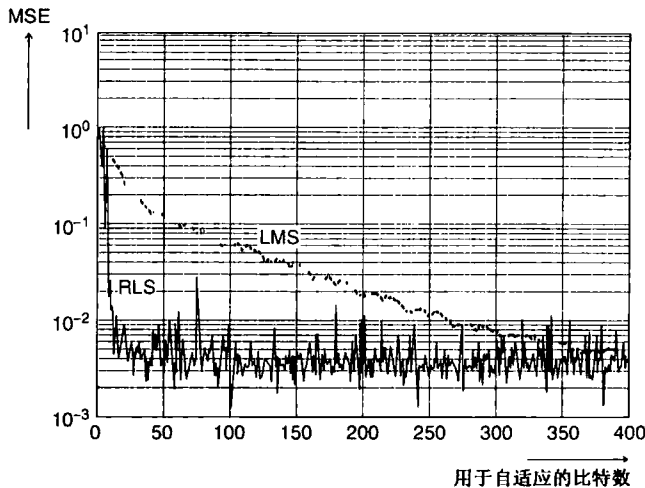


图 16.5 判决反馈均衡器的均方误差与迭代次数的函数关系(见下文)。对于最小均方, $\mu = 0.03$; 对于迭代最小均方, $\lambda = 0.99, \delta = 10^{-9}$

另一方面,最小均方算法需要的(复)运算少得多(见图 16.6;术语“DFE”和“梯度格”将在下面解释)。图中的一个重要结论是:当权值数比较少时,算法的复杂度不会有明显的不同。对于 5~8 个权值的情况,这个差别小于 50% (最小均方除外,因为它由于收敛速度太慢而不适合)。对于均衡而言,收敛、稳定和容易实现是其选择的主要标准。

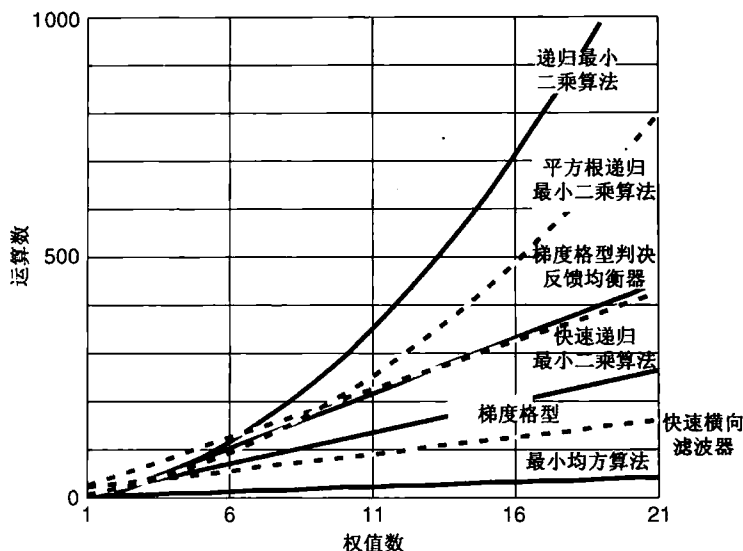


图 16.6 采用不同算法的判决反馈均衡器每一步迭代的运算数(见下文)

16.2.4 其他线性结构

到目前为止,我们已经研究了横向有限冲激响应滤波器。然而,线性滤波器也可以由其他结构来实现,一种可以利用的结构是递归滤波器(无限冲激响应滤波器, IIR)。这些滤波器的优点是实现均衡所需的抽头较少,其主要缺点是传输函数不仅会有零点,而且会有极点,因此将不稳定。基于这个原因,无限冲激响应滤波器在实际中很少被采用。

另一种可能的结构是格形滤波器。其均衡算法的方程与横向滤波器的不同(详细内容见 Proakis[2005])。

16.3 判决反馈均衡器

判决反馈均衡器(DFE)有一个简单的根本前提:一旦正确地检测出一个比特,就可以利用该信息,并结合信道冲激响应的信息来计算由该比特引起的码间干扰。换句话说,确定该比特将给接收信号的后续采样值带来什么样的影响。然后,每个比特引起的码间干扰可以从这些后面的样值中减去。

判决反馈均衡器的方框图如图 16.7 所示。判决反馈均衡器由一个传输函数为 $E(z)$ 的前向滤波器和一个传输函数为 $D(z)$ 的反馈滤波器组成,其中前向滤波器是一个一般的线性均衡器。一旦接收机确定了接收符号,就可以计算出它对所有未来样值的影响(后向码间干扰),并且(通过反馈)从接收信号中减去。关键的一点是,码间干扰是基于硬判决之后的信号计算出来的,这样就从反馈信号中消除了加性噪声。因此,判决反馈均衡器的错误概率比线性均衡器要小。

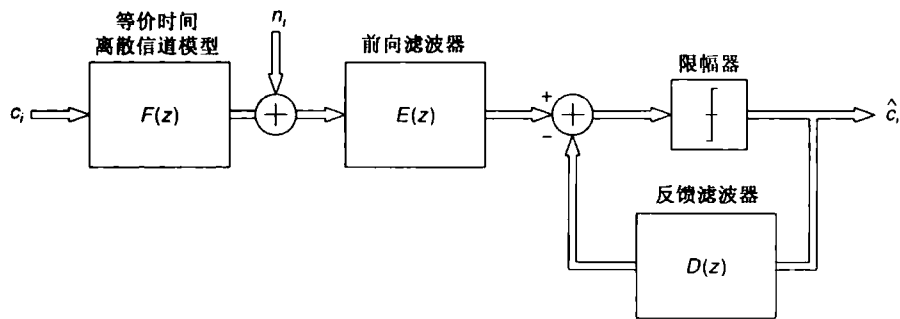


图 16.7 判决反馈均衡器的结构

判决反馈可能存在的一个问题是错误传播。如果接收机有一个比特判决不正确，那么计算出来的后向码间干扰也是错误的，因此到达判决装置的后面信号样值中的码间干扰甚至会比未均衡时还要严重。这导致错误判决和后向码间干扰错误相减的恶性循环。

当误比特率比较低时，错误传播通常不起作用。但是，应注意小的差错率通常通过编码来获得。因此，必须先对这些比特进行译码，再将它们重新编码（这样信号就成为接收信号的非噪声版本），并将这个新的信号用于判决反馈均衡器的反馈中^①。下面，仅考虑无错误传播的未编码系统。

16.3.1 最小均方误差判决反馈均衡器

通过在噪声增强和剩余码间干扰之间寻求平衡，最小均方误差判决反馈均衡器的目标仍然是使均方误差最小化。判决反馈均衡器中的噪声增强不同于线性均衡器，因为前向滤波器的系数不同；而后向码间干扰不会影响噪声增强，现在的目标是使噪声和（平均）前向码间干扰之和最小化。显然，两者的性能也不同。

前向滤波器的系数可以由下式计算：

$$\sum_{n=-K_{ff}}^0 e_n \left(\sum_{m=0}^{-l} f_m^* f_{m+l-n} + N_0 \delta_{nl} \right) = -f_{-l}^*, \quad l, n = -K_{ff}, \dots, 0 \tag{16.30}$$

其中， K_{ff} 是前馈滤波器的抽头数。那么反馈滤波器的系数为

$$d_n = - \sum_{m=-K_{ff}}^0 e_m f_{n-m}, \quad n = 1, \dots, K_{fb} \tag{16.31}$$

其中， K_{fb} 是反馈滤波器的抽头数目。

假设一些理想情况，如反馈滤波器必须至少与后向码间干扰一样长，它必须有足够多的抽头来满足式(16.30)，它不存在错误传播，那么均衡器输出端的均方误差为

$$\sigma_n^2(\text{DFE} - \text{MMSE}) = N_0 \exp \left(\frac{T_s}{2\pi} \int_{-\pi/T_s}^{\pi/T_s} \ln \left[\frac{1}{\Xi(e^{j\omega T_s}) + N_0} \right] d\omega \right) \tag{16.32}$$

① 因为译码器自身存在时延，所以这个任务极具挑战性。可能的解决方案是将均衡、译码及软信息的交换联合设计（见第14章）。

16.3.2 迫零判决反馈均衡器

迫零判决反馈均衡器在概念上甚至更简单。正如 16.1.2 节所提到的, 噪声白化滤波器消除了所有的前向码间干扰, 这样得到的有效信道是纯因果的。后向码间干扰被反馈支路减去。检测器的有效噪声功率为

$$\sigma_n^2(\text{DFE} - \text{ZF}) = N_0 \exp \left(\frac{T_s}{2\pi} \int_{-\pi/T_s}^{\pi/T_s} \ln \left[\frac{1}{\Xi(e^{j\omega T})} \right] d\omega \right) \quad (16.33)$$

该式表明噪声功率比未均衡时大, 但是比线性迫零均衡器小。

例 16.2 使用例题 16.1 的信道, 计算最小均方误差判决反馈均衡器和迫零判决反馈均衡器的噪声增强。

解:

从例题 16.1, 有 $\Xi(e^{j\omega T}) = 0.48 \cos 2\omega T_s - 1.4 \cos \omega T_s + 1.01$ 。把它代入式(16.32)可得

$$\sigma_n^2(\text{DFE} - \text{MMSE}) = N_0 \exp \left(\frac{T_s}{2\pi} \int_{-\pi/T_s}^{\pi/T_s} \ln \left[\frac{1}{\Xi(e^{j\omega T}) + N_0} \right] d\omega \right) \quad (16.34)$$

$$= N_0 \exp \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \left[\frac{1}{0.48 \cos 2\omega - 1.4 \cos \omega + 1.31} \right] d\omega \right) \quad (16.35)$$

$$\approx 0.33 \quad (16.36)$$

所以输出信噪比为 2。因此, 与加性高斯白噪声情况相比, 信噪比恶化了 2 dB。对于迫零判决反馈均衡器, 输出端的噪声方差是 0.83, 因此信噪比下降了 4.4 dB。

16.4 最大似然序列估计: 维特比检测器

上面研究的均衡器结构会影响发送符号的判决。对于最大似然序列估计, 从另一个角度出发, 试图确定最有可能发送的符号序列。这种情况与卷积码译码非常相似。事实上, 通过一个时间色散信道的传输可以看成是一个码率为 $R_c = 1/1$ 的卷积编码。最大似然序列估计器是性能最好的均衡器。

时间离散信道的输出信号可以写成

$$u_i = \sum_{n=0}^{L_c} f_n c_{i-n} + n_i \quad (16.37)$$

其中, n 是方差为 σ_n^2 的高斯白噪声。对于一个长为 N 的接收序列, 接收信号向量 $\mathbf{u}$ 的联合概率密度函数(取决于数据向量 $\mathbf{c}$ 和冲激响应向量 $\mathbf{f}$)为^①

$$\text{pdf}(\mathbf{u}|\mathbf{c}; \mathbf{f}) = \frac{1}{(2\pi\sigma_n^2)^{N/2}} \exp \left(-\frac{1}{2\sigma_n^2} \sum_{i=1}^N \left| u_i - \sum_{n=0}^{L_c} f_n c_{i-n} \right|^2 \right) \quad (16.38)$$

$\mathbf{c}$ 的最大似然序列估计(对于一个给定的 $\mathbf{f}$)是确定使联合概率密度函数 $\text{pdf}(\mathbf{u}|\mathbf{c}, \mathbf{f})$ 最大化的

① 在这里及后面的内容中, 均假设所有的符号是等概率发送, 因此最大似然序列估计与最大后验估计是一样的。

向量值。因为仅在指数项上有变量,所以只需使下式最小化:

$$\sum_{i=1}^N \left| u_i - \sum_{n=0}^{L_c} f_n c_{i-n} \right|^2$$

(16.39)

与卷积码译码类似,确定最佳发送序列有许多不同的算法。接收机首先产生有效发送序列与信道冲激响应卷积后可能形成的所有序列,然后试图找到与接收信号有最小距离(最佳度量)的序列。最直接(但是计算量仍然非常大)的方法就是穷搜索。实际上,可以用维特比算法代替穷搜索算法。

如上所述,只有当最大似然序列估计器输入端的加性噪声是白噪声时,最大似然序列估计器才是最佳的。因此,在检测器中使用的样值必须是噪声白化滤波器的输出。该滤波器必须适应于当前的信道状态,并且每个信道实现都需要进行频谱分解。鉴于这些困难,总的输入滤波器(匹配滤波器和噪声白化滤波器)常常被一个带宽近似为符号周期倒数的矩形滤波器代替。但要注意在这种情况下,如果每个符号仅一个采样值,就不再能提供充分的统计特性。

例 16.3 维特比均衡。

解:

这个例子给出了一个符号流的维特比检测工作原理,符号流经过的信道时间离散冲激响应如下:

$$\mathbf{f} = \begin{pmatrix} 1 \\ -0.5 \\ 0.3 \end{pmatrix}$$

(16.40)

信道可以看成权重分别为 1, -0.5, 0.3 的抽头延迟线;如图 16.8 的上半部分所示,左侧给出了抽头延迟线模型,右侧给出了类似于第 14 章讨论的卷积码的单元模型。为了便于解释,选择一个具有实冲激响应的信道,调制方式为二进制相移键控。图 16.8 的下半部分显示了网格图中可能的转移。由于 $L_c=2$,必须考虑网格中的 4 种状态,在等效的移位寄存器单元中,可能的状态数等于调制字符 $M=2$ 的大小。进一步假设已知网格图的初始状态,即 -1 -1 (例如,因为已知比特在译码开始之前已经发送)。在图 16.9 的底部,显示了网格图的演变过程。靠近转移图附近的数字为所考虑序列。

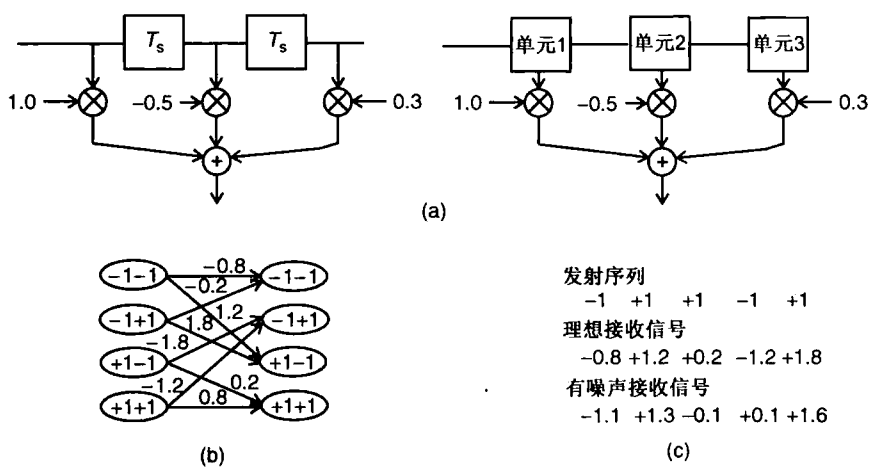


图 16.8 (a) 抽头延迟线信道; (b) 转换概率, (c) 发送和接收信号

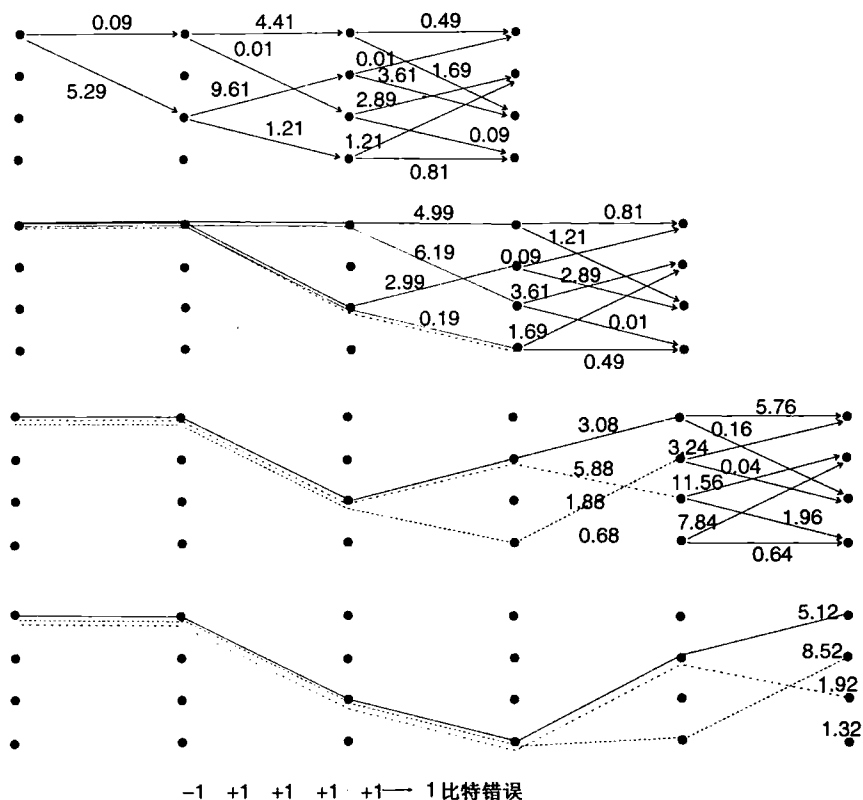


图 16.9 时延色散信道中发送序列检测的维特比算法

16.5 均衡器结构的比较

图 16.10 给出了均衡器结构的分类。为一个实际系统选择均衡器时,必须考虑如下准则。

- 误比特率最小化。在这里,最大似然序列估计器优于其他所有结构的均衡器。判决反馈均衡器虽然次于最大似然序列估计器,但优于线性均衡器。各种结构之间定量的差别取决于信道冲激响应。
- 信道能否处理信道传输函数中的零点。迫零均衡器存在此问题,因为它们取信道传输函数的倒数,因此在均衡器传输函数中存在极点。最小均方误差和最大似然序列估计均衡器都不存在这个问题。
- 计算量。线性均衡器和判决反馈均衡器的计算量没有明显的差别。根据自适应算法的不同,运算量随均衡器的长度(权值数)呈线性、二次方、三次方增加。对于最大似然序列估计器,计算量随信道冲激响应的长度呈指数增长。对于短的信道冲激响应(例如,在 GSM 系统中,信道冲激响应的长度最多持续 4 个符号周期),最大似然序列估计器的计算复杂度和其他均衡器结构可比拟。
- 对信道错误估计的敏感性。由于差错传播效应,判决反馈均衡器对信道估计错误的敏感性大于线性均衡器。另外,迫零均衡器也比最大似然序列估计均衡器敏感。
- 功耗和成本。可以根据计算量得出。

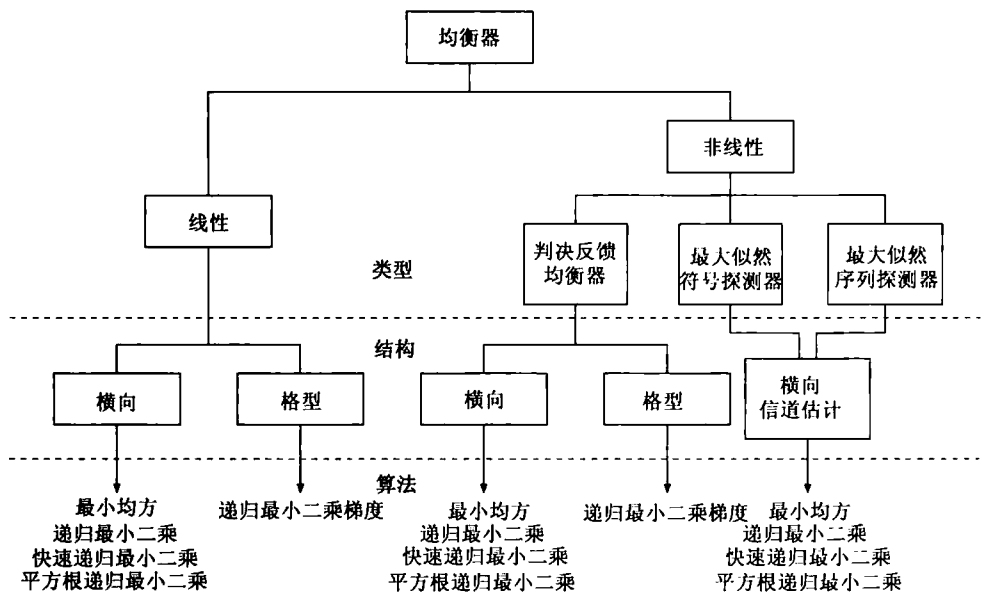


图 16.10 均衡器结构的分类

16.6 分数间隔均衡器

在很多情况下，接收机以符号频率 $(1/T_s)$ 对信号进行采样和处理。如果采样速率低于奈奎斯特速率，并且匹配滤波器仅仅与发送脉冲（不是接收的失真了的脉冲）相匹配，它就是次最佳的。通过升余弦滤波器（滚降系数为 α ）滤波后的信号，频谱扩展到 $(1 + \alpha)/2T_s$ ，因此奈奎斯特速率为 $(1 + \alpha)/T_s$ 。分数间隔均衡器是基于采样速率不低于奈奎斯特速率的均衡器。均衡器的抽头间隔为 $T_s a/b$ ，其中 $a < b$ ，并且 a 和 b 为整数。

分数间隔均衡器也可以解释为用一步实现均衡和匹配滤波。它们对于采样时刻的误差也不敏感。其缺点是所需的抽头数大于符号间隔均衡器，因此计算量更大。

16.7 盲均衡

16.7.1 简介

“标准”的均衡器的工作分两个阶段：训练阶段和检测阶段。在训练阶段，通过信道发送一个已知的比特序列，并且在接收机中将失真信号和（本地产生的）未失真信号进行比较，这样就能提供用于均衡的信道冲激响应信息（见 16.1.3 节对 pros 和 cons 的讨论）。相反，盲均衡利用发送信号的已知统计特性来估计信道和数据。通过使均衡器输出的某些统计特性与发送信号的已知统计特性相匹配，来调整均衡器的系数。

盲均衡的优点包括：

- 整个时间段都用来确定信道的冲激响应，而不采用短的训练序列。
- 因为不用浪费时间来发送训练序列，所以频谱效率提高了。

以下的信号特性可用于盲均衡。

- 恒包络。对于很多信号(频移键控、最小频移键控和高斯最小频移键控),包络(振幅)都是恒定的。
- 统计特性。例如循环平稳性。
- 字符有限。在信号星座图中只有某些离散值是有效点。
- 频谱相关。信号的频谱和它们自身移位后的信号频谱相关。
- 这些特性的联合。

比较成熟的盲算法包括:(i)恒模算法;(ii)盲最大似然序列估计;(iii)基于高阶统计特性的算法。下面将讨论这些算法。

16.7.2 恒模算法

恒模算法(CMA)是盲均衡中最早的算法。其最简单的形式采用最小均方自适应算法。16.2 节介绍了数据辅助的最小均方算法,它是基于期望信号(在接收机中已知)和均衡器输出的误差最小化的算法。在盲最小均方算法中,必须首先由均衡器的输出来产生期望信号。把均衡器的输出通过一个非线性函数就可以得到期望信号。那么,误差信号就是这个非线性函数的输出和均衡器的输出之间的差值。传统均衡器和恒模算法均衡器之间的不同如图 16.11 所示。

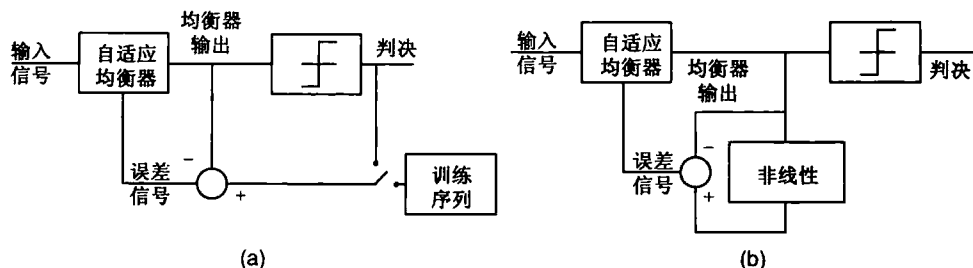


图 16.11 (a) 传统均衡器结构; (b) 基于常模算法的均衡器结构

这个非线性函数可能是无记忆的,也可能是 m 阶记忆的。各种类型的算法(通常用其发明者的名字来命名)通过不同的非线性来区分。最著名的是 Sato 算法(1975 年)和 Godard 算法(1980 年)。

用于盲自适应的最小均方算法与传统的最小均方算法有相同的缺点:收敛速率太慢(增大步长 μ 可加快收敛速度,但是收敛之后的误差会变差)。该算法也可能收敛于一个局部最小值。这些问题可以通过解析恒模算法[van der Veen and Paulraj 1996]来解决。这是一个在无噪声情况下提供精确解的非迭代算法,并且它对噪声也具有稳健性。

16.7.3 盲最大似然估计

对于传统的最大似然序列估计,训练序列用来确定信道的冲激响应 $\mathbf{f}$ 。然后在实际数据传输中使用这个估计,仅需采用最大似然估计解出 $\mathbf{c}$ 。对于盲估计, $\mathbf{c}$ 和 $\mathbf{f}$ 必须同时进行估计。这可以由如下任意一种方法实现。

1. 通过对所有可能的数据序列求平均得到的 $\text{pdf}(\mathbf{u}|\mathbf{f}, \mathbf{c})$, 估计出信道的冲激响应。该方法有两个缺点:它需要一段相当长的计算时间来进行平均,并且它是次最佳的。
2. 可以对所有可能的数据序列确定 $\mathbf{f}$ 的最大似然估计,然后选择具有最好总测度的 $\mathbf{f}, \mathbf{c}$

对。该方法比平均方法的计算量更大,但也更精确。此外,Proakis[2005]已经提出了减少计算量的方法。

16.7.4 使用二阶或高阶统计特性的算法

一般来说,二阶统计特性(自相关函数)不能提供信道冲激响应的相位信息,除非当接收信号的自相关函数具有周期性。因此循环平稳特性是盲估计方法使用二阶统计特性的基础。与第6章中的讨论类似,应区分严循环平稳和宽循环平稳。如果随机过程以采样周期 T_{per} 的整数倍平移时,其所有的统计特性都保持不变,则称之为严平稳。对于宽平稳,只需要均值和自相关函数满足以下条件则可:

$$E\{x(t + iT_{\text{per}})\} = E\{x(t)\} \quad (16.41)$$

$$E\{x^*(t_1 + iT_{\text{per}})x(t_2 + iT_{\text{per}})\} = E\{x^*(t_1)x(t_2)\} \quad (16.42)$$

如果信号被过采样,得到的样值序列就一定是循环平稳的。实际的信道估计是基于接收信号的不同的相关矩阵,具体可参考 Tong et al. [1994, 1995]。使用高阶统计特性不要求信号必须是循环平稳的,然而它的精确度通常低得多。

16.7.5 评价

从发展历史上来说,盲均衡在20世纪70年代年首先提出用于多端计算机网络(一个中心站与多个端口相连),在80年代和90年代,关于这个主题开展了大量的理论研究工作,并提出了一些具有挑战性的数学问题。然而直到现在,在实际系统中,真正的盲均衡还没能取代基于训练序列的均衡。主要原因似乎是两者在计算量和可靠性上有着巨大的差别。实际上,盲均衡需要一段很长的时间去收敛,因此在快速时变的无线信道中的性能并不好。最近几年提出了另一种方法,它首先利用训练序列来得到信道的初始估计,然后使用盲(判决辅助的)均衡来更新这些估计值[Loncar et al. 2002]。

16.8 附录

请访问 www.wiley.com/go/molisch。

深入阅读

Lucky et al. [1968]对自适应均衡器首次进行了全面的介绍,至今仍值得一读。本章关于线性均衡器和判决反馈均衡器内容的介绍引自 Proakis[2005],该书还有许多更令人关注的详细介绍。Haykin[1991]描述了自适应滤波器,采用了另一种方法来解释均衡器。Belfiore and Park[1977]也介绍了判决反馈均衡器。分数间隔均衡器在 Gitlin and Weinstein[1981]和 Ungerböck[1976]中均进行了分析。维特比均衡器是维特比算法[Viterbi 1976]的一种应用。Gorokhov[1998]讨论了信道估计误差带来的影响。Vitetta et al. [2000]详细介绍了各种均衡器的结构,包括信道状态信息(理想的、估计的或平均的)的影响。Sayed and Kailath[2001]研究了谱分解技术。盲均衡的算法太多,在此不一一列出。作为例子,我们仅列出 Giannakis and Halford[1997], Liu et al. [1996], Sato[1975]和 van der Veen and Paulraj[1996]等文献。迭代均衡器的描述在 Wymeersch[2007]中。

第四部分 多址和高级收发信机方案

第 17 章	多址和蜂窝原理
第 18 章	扩展频谱系统
第 19 章	正交频分复用
第 20 章	多天线系统
第 21 章	认知无线电
第 22 章	中继、多跳和协作通信
第 23 章	视频编码

本书第三部分描述了单个发射机如何与单个接收机进行通信。然而，就大多数无线系统而言，多部设备应当能够在同一地理区域内同时进行通信，因此系统必须提供多址 (multiple access) 手段。就大多数第一代和第二代的蜂窝系统、无绳电话及无线局域网而言，不同设备或者在不同频率上，或者在不同时间上进行通信。第 17 章将讨论基于这一原理 (称为“频分多址”、“时分多址”和“分组无线电”) 的多址方案。另一种不同的多址类型是基于信号扩展原理的，将用户信号扩展到很宽的带宽上，并确保对每个用户信号的扩展都是独一无二的。这种多址方式允许多个用户同时发送信号，而接收机可以从这种扩展中确定出“空中 (on-air)”信号里哪一部分来自于哪个用户。这种扩展频谱 (spread spectrum) 方式将在第 18 章描述，采用码分多址形式的第三代蜂窝系统就是这种方式的应用实例。第 18 章还将描述 Rake 接收机，一种使码分多址系统能够有效对抗信道时延色散影响的设备，以及在多址环境中，如何利用多用户检测来较大幅度地提升系统性能。

随着无线系统的数据速率变得越来越高，对实际应用而言，均衡器和 Rake 接收机都会变得过于复杂。正交频分复用是解决这一问题的一种方式。在这种方式中，数据流被分解成大量的子数据流，每个子数据流被调制在一个不同的载波上；因此，通常又将正交频分复用称为一种多载波调制 (multicarrier modulation) 方法。第 19 章描述了正交频分复用原理，以及将码分多址和正交频分复用结合起来的某些变化形式。

在将时域、频域和码域用于信号传输之后，“空域就成为了有待开发的最后领域”。多天线单元支持利用空域来提高数据速率，也可以支持更多的同时使用系统的用户，或者使传输质量得到改善。利用多部天线的形式之一就是天线分集，已在第 13 章专门讨论过；天线分集主要用于对抗深衰落在传输质量方面造成的负面影响。第 20 章描述了智能天线理论，智能天线能够提高一个传输系统中可获得服务的用户数目。此外，这一章还将讨论多输入多输出 (MIMO) 系

统,这种系统在发射机和接收机处均具有多部天线,从而可以利用它们并行地传送若干个数据流。这样,在不需要增加额外的频谱的情况下,就可以提高数据速率。

大多数情况下,指定部分的频谱将被分配给特定业务或运营商,而不是被用到所有的场合中。第 21 章描述的认知无线电,旨在通过使无线电设备能够监测频谱占用情况,并通过对(当地)未使用的部分频谱的占用来消除这些频谱使用的低效率状况。

第 22 章描述“用户协作”,这种方式下不同的设备并非彼此竞争(因而就会产生干扰),而恰恰相反,它们彼此相助来将消息转发至信宿。这种协作的最简单形式就是中继(relay),其唯一目的就是转发消息。在更大规模的网络中,诸中继节点可以相互协作。而且,在许多网络中,节点可能根据需要将其属性在信源、中继节点和信宿之间进行变更。

随着视频成为无线通信系统中越来越重要的应用,这些应用的信源编码特性将会影响到系统性能。第 23 章将给出视频编码技术的概述。

第 17 章 多址和蜂窝原理

17.1 引言

一个无线通信系统使用分配给其特定业务的指定频段。频谱是一种稀缺资源且已分配的频段不易被扩展。鉴于此,无线系统必须为在指定频段内允许尽可能多的用户同时通信做出一些规定。

使多个用户同时进行通信的问题可以被划分为两个部分:

1. 只有一个基站的情况下,怎样与多个移动台同时通信?
2. 在多个基站的情况下,如何将频谱资源分配给它们,以使可能用户的总数最大化?另外,这些基站在给定地理区域内如何布置?

对于第一个问题,有多种不同的方式可以实现允许多个用户同时与一个基站通话的目的,我们称之为多址(MA)方式。本章将讨论以下 3 种方式:

- 频分多址,将不同的频率分配给不同的用户。
- 时分多址,将不同的时隙分配给不同的用户。
- 分组无线电,可以被看成时分多址的一种形式,它自适应地将时隙分配给用户。

频分多址和时分多址在 17.2 节和 17.3 节讨论。实际上,这些小节相当简要,这似乎贬低了时分多址和频分多址的重要性。然而,这两种多址方式在概念上很容易理解,并且与它们的具体实施有关的许多概念已在第 11 章、第 12 章和第 16 章讨论过。分组无线电作为时分多址的一种变形,是无线数据通信的重要形式,将在 17.4 节讨论。除了以上提到的方案之外,码分多址为每个用户分配不同的码字,近些年来得到了持续增长的普及。第 18 章将讨论这一方案及其他扩频方式。最后,空分多址是多天线系统的一种多址格式,它可以和其他所有多址方式结合使用,将在 20.1 节讨论。双工(duplexing)的概念,即在收发信机处将发送与接收分离开,将在 17.5 节进行分析。

所有这些方式的目标都是使频谱效率最大化,即最大化每单位带宽里可支持的用户数。如上所述,还有一个不同的(尽管也是相互关联的)问题:如何设计系统以使每单位带宽和单位面积可支持的用户数最大化?完成这一目标显然需要多个基站并涉及对其进行频谱资源的分配。这就需要在不同基站处重复使用相同频谱,从而引出蜂窝原理。蜂窝原理将在 17.6 节讨论。

17.2 频分多址

17.2.1 频率划分实现多址

频分多址(FDMA)是最老的概念最简单的多址方式。系统给每个用户分配一个频率(子)带,即可用频谱的一部分(通常是连贯的),如图 17.1 所示。频带的分配常常在呼叫建立期间

完成,且在整个通话期间持续占用。频分多址常与频域双工(Frequency Domain Duplexing, FDD)(见 17.5 节)结合在一起,所以两个有着固定的双工间隔的频带被分配给每个用户,其中一个用于下行链路(基站到移动台)通信,另一个用于上行链路(移动台到基站)通信。

纯频分多址的概念非常简单,并且在实现上具有一些优势:

- 发射机和接收机几乎不需要数字信号处理。然而,这一点实际上不再重要,因为数字信号处理的造价在持续性地降低。
- (时间)同步简单。一旦在呼叫建立时建立起同步,当发送持续进行时,容易借助简单的跟踪算法保持同步。

然而,纯频分多址也有明显的缺陷,特别是用于语音通信时。如下这些问题的出现缘于对频谱有效性的考虑和对多径效应的敏感度。

- 频率同步及稳定比较困难。对于语音通信,每个频率子带是很窄的(典型地为 5 ~ 30 kHz),因而要求本地振荡器非常准确和稳定;载频的抖动将导致邻道干扰。高的频谱效率也要求使用边缘非常陡峭的滤波器来提取期望的信号。准确的振荡器和陡峭的滤波器的成本都很高,因而不符合需求。如果不采用它们,就需要使用保护频带来缓解对滤波器的过高要求,然而这样就降低了系统的频谱效率。
- 对衰落的敏感性。因为每个用户被分配了一个不同于其他用户的频带,并且这些频带比其他多址方式(与时分多址和码分多址相比)的频带都窄,即 5 ~ 30 kHz。对于如此窄的子频带,在实际的所有传播环境中,衰落都是平坦的。这样就有了无须均衡的优点;而缺点就是无法进行频率分集,读者应该记得,频率分集主要由频率间隔超过信道的相关带宽的多个信号分量提供(见第 13 章和第 16 章)。
- 对随机频率调制的敏感性。由于窄的信号带宽使系统对随机频率调制较敏感:随机频率调制造成的误比特率(BER)正比于 $(\nu_{\max} T_s)^2$ (见第 12 章)。因此,也就反比于带宽的平方。幸运的是,适当的信号处理方案不仅能减轻这些效应,甚至还能利用它们获得时间分集。注意,此处的情形对偶于宽带系统的情形,在那里,时间的延迟扩展成为缺点,但是均衡器通过利用频率分集就能将它们转变为优点。
- 互调。基站要发送多个语音信道,且每个信道在整个通话时间内都是持续使用的。典型地,一个基站使用 20 ~ 100 个频道。如果这些信号由同一个功率放大器来放大,就可能生成三阶互调成分,这些成分位于不期望的频率范围,即在发送频带以内。因而,需要为每个语音信道准备一个独立的放大器,或者为复合信号提供高度线性的放大器。每种解决方案都将使基站成本更高。

鉴于这些原因,频分多址大多用在以下场合。

- 模拟通信系统。频分多址是对模拟系统而言唯一行得通的多址方式。
- 频分多址与其他多址方式结合。分配给某项业务(或者某个网络运营商)的频谱被划分

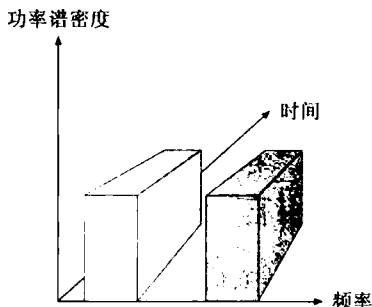


图 17.1 频分多址原理

成较宽的子频带,每个频带服务于一组用户。组内用户的多址借助于另一种多址方式,如时分多址或码分多址。绝大多数现今的无线系统就是这样使用频分多址的(见第 24 章至第 28 章)。

- 高数据速率系统。频分多址的缺点主要与每个用户只需要小的带宽(如 20 kHz)有关。对于无线局域网,情况就不同了。在那种应用场合下,单个用户需要的带宽接近 20 MHz,而且仅有少量频道可供使用。

17.2.2 中继增益

现在来计算频分多址系统的一个基站可以覆盖多少个用户。这一看似简单的问题已成为通信理论的一个单独的分支(常称之为排队论),有几本教材专门讨论它,比如 Gross and Harris [1998]。在这一节将以非常简单的方式描述如何计算为给定数目的用户提供“质量满意”的服务所需的通信信道数。为此目的,假设系统为不需要任何信令信道的纯频分多址系统,还假设系统是纯粹为语音通信设计的。为了后面论述的需要,将提供的业务量定义为呼叫到达率和平均呼叫保持时间(持续时间)的乘积;通常,提供业务量以“爱尔兰(Erlang)”为单位,尽管实际上这是一个无量纲的量(呼叫到达率的单位是 1/s;呼叫保持时间的单位是 s)。

在规划蜂窝网络时,有如下两种极端的情形。

1. 最坏状况的设计。假定所有用户想要同时呼叫。如果网络运营商希望每个小区为 700 个用户提供服务,它就必须提供 700 个话音信道。当然,实际上没必要建设这样的网络,正如设计一个可以同时治疗一个城市里所有居民的医院,都没有必要。
2. 最好状况的设计。一个典型的用户每天使用电话的时间只有 20 分钟;如果每个小区有 700 个潜在用户,则实际用到的呼叫持续时间为 14 000 分钟。对应每个小区有 10 条语音信道的系统,可以提供 $24 * 60 * 10 = 14\ 400$ 分钟的通话时间,因而足以向所需数目的用户提供服务。然而,这一计算假定了所有用户的呼叫是相继的,即旧的呼叫一完成就有新的用户呼叫呼入,而且呼叫均匀分布在 24 小时期间。

显然,两种极端的情形都不现实。在很大程度上,网络设计的技术实质就是预测用户的呼叫行为,并推导出确保得到可接受的业务等级的实际的基础设施条件(即可使用的语音信道数)。

以下几个因素会影响这一规划过程。

1. 呼叫数和持续时间因时而异。因此,可以定义忙时(常常在 10 点钟和 16 点钟左右),即有最多的呼叫发生的一个小时。忙时期间的业务量决定了所需的网络容量。
2. 用户的空间分布是时变的。商业区(城市中心)常常在白天有许多呼叫,市郊和娱乐区则在晚间有更多的业务。
3. 通话习惯在许多年间也在发生着改变。20 世纪 80 年代后期,来自蜂窝电话的呼叫常常限定在几分钟以内,现在以小时计长的呼叫已经非常普遍。
4. 用户习惯的变化也与新业务(如数据链接)的提供及新的价格结构(如晚间时段的免费呼叫)相关联。通话时间打包销售的策略使用户购买后必须按月用完,从而也就导致了比以往采用按分钟计费的价格策略时更长的通话时间。

基于用户习惯的统计经验,现在可以设计出以一定概率允许每小区内给定数量的用户进

行呼叫的系统。即使出现了统计上的意外情况,即更多的用户想要同时呼叫,其中一些呼叫将被阻塞掉。注意,承载业务量是提供业务量与 $1 - \text{Pr}_{\text{block}}$ 相乘的结果。

为计算简化系统的阻塞概率,进行如下假定:(i) 呼叫发起时刻统计独立;(ii) 呼叫持续时间是一个服从指数分布的随机变量;(iii) 如果一个用户被拒绝了,他的下一次呼叫尝试与前一次呼叫尝试统计独立(即当成一个新用户)^①。这样的系统称为 Erlang B 系统;其呼叫阻塞概率可以表示为

$$\text{Pr}_{\text{block}} = \frac{T_{\text{tr}}^{N_c} / N_c!}{\sum_{k=0}^{N_c} T_{\text{tr}}^k / k!}$$

(17.1)

其中, N_c 是(每小区)的语音信道数, T_{tr} 是提供的平均业务量。图 17.2 以曲线图表明了这一关系。可以看出,当 N_c 较小时,需要的信道数和提供业务量的比值非常高,尤其是需要低的阻塞概率时。例如,对于期望的阻塞概率为 1% 的情况,当 $N_c = 2$ 时,可能的提供业务量与可用信道数的比值小于 0.1。如果 N_c 很大,则比值仅略小于 1,且几乎独立于期望的阻塞概率。再次假定期望的阻塞概率为 1%,当 $N_c = 50$ 时,可能的提供业务量与可用信道数的比值约为 0.9。

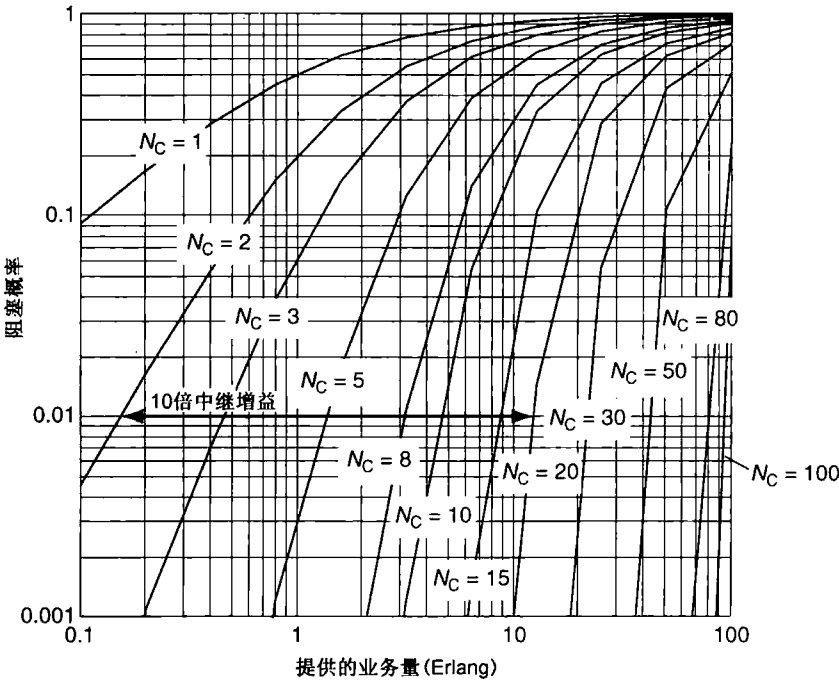


图 17.2 Erlang B 系统的阻塞概率。 N_c 是可用的语音信道数。例如,当阻塞概率为 0.01 时,信道数从 $N_c = 2$ 变为 $N_c = 20$,允许的业务量将从 0.15 Erlang 增长约为 12 Erlang;由于可获得的语音信道数仅增长为原来的 10 倍,故中继增益将增大为原来的 8 倍

另一个模型称为 Erlang C,该模型假定任何不能马上分配信道的用户将进入一个等待循环,只要一有信道就分配给他。用户被延迟的概率为

① 这显然与实际不符。一般情况下,一个被阻塞的用户立即就会重新尝试呼叫。在短时间间隔内被多次阻塞后,她(或他)常常放弃呼叫,然后在相当长的等待时间之后才会发起下一次呼叫。

$$\text{Pr}_{\text{wait}} = \frac{T_{\text{tr}}^{N_C}}{T_{\text{tr}}^{N_C} + N_C! \left(1 - \frac{T_{\text{tr}}}{N_C}\right) \sum_{k=0}^{N_C-1} T_{\text{tr}}^k / k!} \tag{17.2}$$

且平均等待时间为

$$t_{\text{wait}} = \text{Pr}_{\text{wait}} \frac{T_{\text{call}}}{N_C - T_{\text{tr}}} \tag{17.3}$$

此处, T_{call} 是平均呼叫持续时间。

例 17.1 Erlang C 系统, 用户在 50% 的时间处于活跃状态, 平均呼叫持续时间为 5 分钟。希望所有呼叫中不超过 5% 的呼叫会进入等待循环。用户数 n_{user} 分别为 1、8、30 时所需的信道数是多少? 每种情况下的平均等待时间是多少?

解:

因为 $T_{\text{tr}} = 0.5 \cdot n_{\text{user}}$, 它是平均提供的业务量, 我们需要找到满足下式的 N_C :

$$0.05 \geq \frac{(0.5 \cdot n_{\text{user}})^{N_C}}{(0.5 \cdot n_{\text{user}})^{N_C} + N_C! \left(1 - \frac{0.5 \cdot n_{\text{user}}}{N_C}\right) \sum_{k=0}^{N_C-1} \frac{(0.5 \cdot n_{\text{user}})^k}{k!}} \tag{17.4}$$

该式需要用数值方法来求解, 其结果在表 17.1 中给出。

由于 $T_{\text{call}} = 5$ 分钟, 平均等待时间为

$$t_{\text{wait}} = \text{Pr}_{\text{wait}} \frac{5}{N_C - 0.5 \cdot n_{\text{user}}} \tag{17.5}$$

满足以上不等式的所需信道数和相应求得的平均等待时间为

表 17.1 所考虑的 Erlang C 系统的参数

n_{user}	1	8	30
N_D	3	9	23
Pr_{wait}	0.0152	0.0238	0.0380
t_{wait}	0.0304	0.0238	0.0238

一个呼叫可被暂时搁置而不是被阻塞的概率是服务质量的重要组成部分。回忆一下, 服务质量等于 100% 减去被阻塞呼叫的百分比, 再减去丢失呼叫 (即中途掉话的呼叫) 的百分比的十倍 (见 1.3.7 节)。在频分多址系统中, 只要每个用户停留在他 (或她) 的基站覆盖范围以内, 大的系统负荷将导致只有被阻塞的呼叫, 而没有丢失的呼叫 (注意, 本节忽略因过高的误比特率所导致的呼叫阻塞和掉话)。然而, 当正处于通话状态的用户试图进入另一个其基站 (的信道) 已被完全占用的小区时, 掉话是可能发生的。如同 17.6 节将会提到的, 全载系统也会增加对相邻小区的干扰, 从而使得其中的链路对信号强度的波动更敏感, 以至于掉话数目可能会因为不足的信噪比和信干比而有所增加。

总之, 我们发现, 可以提供给定服务质量的用户数随可用语音信道数增加而增加的速度, 比线性增长更快。实际增长量与线性增长量之间的差值称为中继增益。因此, 从纯技术的角度看, 更希望有一个大的语音信道池来服务于所有用户。这种情况其实是可能实现的, 比如整个蜂窝系统只有一个运营商, 他拥有分配给蜂窝业务的全部频谱。之所以未能采用这一方法的原因不在于技术而在于价格和垄断^①。

① 回想美国在蜂窝电话发展初期采用的方法: 将全部可用频谱正好分配给两家服务提供商, 以便在避免垄断和技术的有利原则之间达成妥协。

例 17.2 Erlang B 系统, 可用信道数为 30 个。要求阻塞概率低于 2%。分别计算一个运营商或三个运营商时可以提供的业务量是多少?

解:

1. 将所需的阻塞概率 $P_{\text{block}} = 0.02$ 和信道数 $N_c = 30$ 代入式 (17.1), 得到

$$0.02 = \frac{T_{\text{tr}}^{30}/30!}{\sum_{k=0}^{30} T_{\text{tr}}^k/k!} \tag{17.6}$$

求解该式得到 T_{tr} :

$$T_{\text{tr}} = 21.9 \tag{17.7}$$

2. 三个运营商共享 30 个语音信道时, 每个得到 $N_c = 10$ 语音信道, 类似地可以得到每个运营商的平均业务量 T_{tr} :

$$T_{\text{tr}} = 5.1 \tag{17.8}$$

因此, 总的平均业务量可以将三个运营商的平均业务量加起来得到:

$$T_{\text{tr,tot}} = 3 \cdot 5.1 = 15.3 \tag{17.9}$$

17.3 时分多址

对于时分多址, 不同用户未必用不同频率发送, 更确切地是在不同时间发送(见图 17.3)。时间单位被分割为固定间隔的 N 个时隙, 每个用户被分配了其中一个时隙。在所分配的时隙期间, 用户可以以高的数据速率发送(当可以占用整个系统带宽时); 随后, 在后来的 $N-1$ 个时隙, 该用户不再发送, 而由其他用户占用这些时隙发送。然后, 这样的处理将周期性地重复。初看起来, 这种多址方式与频分多址具有同样的性能: 用户仅在 $1/N$ 的可用时间期间发送, 但却占用了 N 倍的带宽。然而, 有如下一些重要的实际差别。

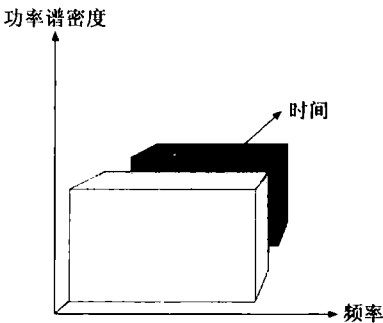


图 17.3 时分多址原理

- 用户占用更宽的带宽。这将允许在系统所分配的带宽内利用频率分集; 而且, 对随机频率调制的敏感度也降低了。相应地, 在多数应用环境下, 需要均衡器来对抗符号间干扰, 这将增加数字信号处理的投入。
- 需要留出保护时间间隔。一个发射机要在很短的时间里将输出功率从 0 提升到“满功率”(典型地, 在 100 mW 到 100 W 之间)。此外, 必须留出足够的保护时间来补偿信号在移动台与基站之间传输的时间。因为可能出现以下情况: 某移动台远离基站, 而使用下一个时隙的另一个移动台离基站却非常近, 以至于后者的传输时间可以忽略。这两个用户的信号在基站处不能重叠, 所以第二个移动台在第一个移动台的信号向基站传播期间就不应该发送^①。然而, 需要指出的是, 并不需要频率保护段, 因为每个用户都完全填满了所分配的频带。
- 当发送不连续时, 每个时隙可能都需要新的同步和信道估计。时隙间隔的优化是具有

^① 通过“时间提前”处理, 需要的保护时间可以缩减(如第 24 章所述)。

挑战性的任务。如果时隙间隔过短,将会有很大百分比的时间被用于同步和信道估计(GSM 中,17%的时隙时间用于此目的);如果时隙间隔过长,传输延迟也将过长(尤其对语音通信用户会带来烦扰),而且在一个时隙期间信道也会有所变化。这样,均衡器必须在一个时隙发送期间能够跟踪信道变化,从而增加了实现的难度(但这是必须的,如在目前已不再使用的暂定标准(IS)136 蜂窝标准中)。如果指派给一个用户的两个时隙之间的时间大于相干时间,信道在这两个时隙之间已经发生了变化,就需要进行新的信道估计。

- 对于干扰受限系统,时分多址有一个重要的好处:在不发送期间,移动台可以“监听”其他时隙的发送^①。这一点在移动台准备从一个基站切换到另一个基站时会非常有用,此时移动台必须找出可能提供更好服务质量并且有可用通信信道的相邻基站。

时分多址被世界范围的蜂窝标准 GSM(见第 24 章)和无绳标准 DECT(数字增强型无绳通信,参见 www.wiley.com/go/molisch)所采用,并且以一种修改后的形式为第四代蜂窝标准 3GPP-LTE(见第 27 章)和 WiMAX(见第 28 章)所采用。时分多址通常与频分多址结合起来,例如在全球移动通信系统中(见图 17.4)。相比之下,纯频分多址主要用于模拟蜂窝和无绳系统。

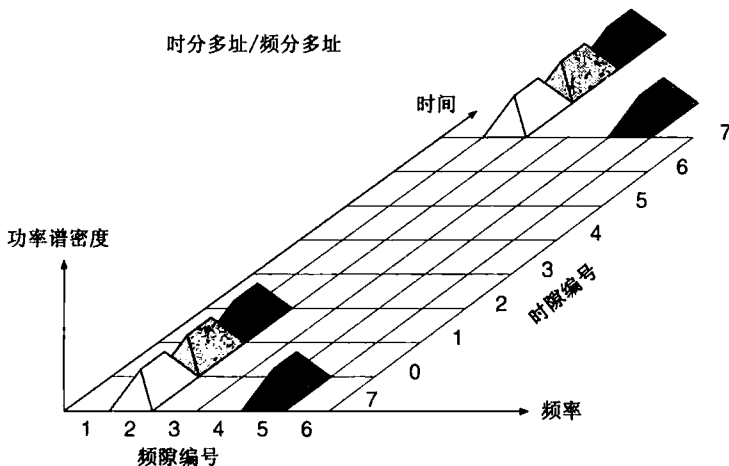


图 17.4 全球移动通信系统如何将时分多址和频分多址结合起来

17.4 分组无线电

分组无线电接入方案将数据分解为分组(packet),且每个分组独立地通过媒介传输。换句话说,每个分组就像一个新的用户,要去争夺它“自认”的资源。当每个用户的数据业务是突发的(如网页浏览、文件下载及类似的数据应用),这种方案让传输媒介得到更有效的利用。

与时分多址和频分多址相比,分组无线电表现出两个主要的不同点。

1. 如上所述,每个分组必须争夺自认的资源。最常见的资源分配方法有 ALOHA 系统、载波侦听多址和分组预约(查询)。这些方法将在 17.4.1 节至 17.4.4 节中描述,其中考虑的情况为:多个用户试图向一个基站发送分组。

^① 在混合的时分多址/频分多址系统中,接收机也可以“监听”其他频率上的发送。

2. 每个分组可以采用不同的路径(即通过不同的中继站)被传送到接收端。这一点在蜂窝系统中不太重要,因为在系统中只可能与最近的基站相连接^①;但却在无线自组织网络和传感器网络中非常重要,因为其中每个无线设备同时又可作为发自另一个无线设备的信息的中继站。合适的路由因此就成为传感器网络的非常重要的方面。这一内容将在第 22 章给以简述。

17.4.1 ALOHA

第一个分组无线系统是夏威夷大学的 ALOHA 系统,该系统用于将夏威夷群岛上不同地点的计算机终端连接到火奴鲁鲁(Honolulu)的中央计算机。下面考虑多个移动台试图向一个指定基站发送分组的情况,所论及的原理也适用于无线自组织网络和传感器网络。

在 ALOHA 系统中,每个用户在数据源形成分组以后就将其发送到基站。一个发射机根本不考虑其他用户是否已经在发送。于是,当几个用户想要同时发送信息时就会出现状况。若两个发射机同时发送分组,则至少有其中之一将遭受到大量干扰而不可用,因此必须重传分组。这样的碰撞降低了系统的有效数据率。所以,在业务负荷较大时,ALOHA 系统将变得效率低下,并且碰撞概率也会因此变得很大。

如果分组发送的起始时刻由发射机完全随机地选择,则这种系统称为纯 ALOHA 或者非时隙 ALOHA 系统。下面来确定非时隙 ALOHA 系统的可能吞吐量。为此先确定可能的碰撞时间,即该时间里可能与其他用户分组发生碰撞。假定所有分组有相同的长度 T_p 。图 17.5 所示为来自发射机 TX2 的分组 A 可能与自己发送之前(或之后)由发射机 TX1 发送的分组发生碰撞的情形。在此假定哪怕是一个短的碰撞也会导致强干扰,以至于分组必须重新发送。为了完全地避免碰撞, TX1 发送的分组必须在分组 A 发送之前的至少 T_p 秒开始发送,或者必须在分组 A 发送完之后开始发送(即必须在分组 A 开始发送后 T_p 秒才开始发送)。因此,可能碰撞的总的时间为 $2T_p$ ^②。

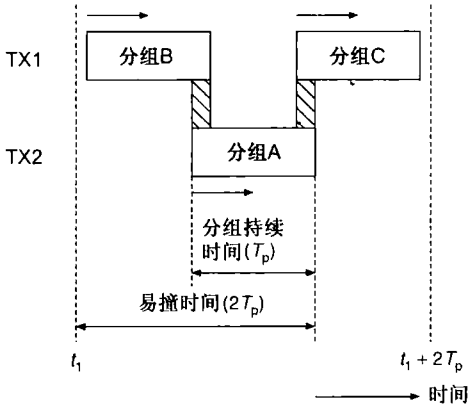


图 17.5 非时隙 ALOHA 系统的碰撞敏感(或碰撞易发生)时间。
引自 Rappaport [1996] © IEEE

为便于数学推导,现在假定所有的发送时刻完全随机,且不同发射机之间彼此独立。所有发射机的平均发送率记为每秒 λ_p 个分组,供给速率 $R = \lambda_p T_p$ 是归一化的信道使用率,数值必须介于 0 到 1 之间。在这些假定下,在持续时间 t 里将有 n 个分组被发送的概率由泊松分布给定 [Papoulis 1991]:

$$\Pr(n, t) = \frac{(\lambda_p t)^n \exp(-\lambda_p t)}{n!}$$

(17.10)

因而, t 时间里,没有产生分组的概率为

① 分组由一个基站向另一基站(或固定线路接收端)的传送,既可以是借助逻辑连接的基于分组的方式,也可以是基于电路的。
② 下面还假定 T_p 包含保护周期,用来体现小区中分组的不同实际传输时延。

$$\Pr(0, t) = \exp(-\lambda_p t) \quad (17.11)$$

有效的吞吐量是信道以有意义的方式使用(即分组被提供后又成功发送)的时间占总时间的百分比。如果在某任意分组的“碰撞敏感”时间内没有竞争性的分组产生,就将实现一次成功的发送。如上所见,可能的碰撞敏感时间是分组长度的两倍,所以无碰撞概率为 $\exp(-2\lambda_p T_p)$ 。因此,有效信道吞吐量就是供给速率乘以分组发送成功概率,即

$$\lambda_p T_p \exp(-2\lambda_p T_p) \quad (17.12)$$

容易获知最大有效吞吐量为 $1/(2e)$, 其中 e 为欧拉常数。

在时隙 ALOHA 系统中, 基站规定了一定的时隙结构, 每个发射机有一个同步时钟, 以确保分组的起始发送时刻与时隙的开始时刻一致。因此, 部分碰撞不再可能发生: 或者两个分组完全碰撞, 或者根本不会碰撞。显而易见, 这种系统的碰撞敏感时间为 T_p , 所以有效吞吐量为

$$\lambda_p T_p \exp(-\lambda_p T_p) \quad (17.13)$$

相应地, 最大可完成吞吐量为 $1/e$, 即二倍于非时隙 ALOHA 系统。

17.4.2 载波侦听多址

基本原理

一个发射机可以判断(侦听)信道当前是否被另一个用户(载波)所占用。这方面的信息可以用来提高分组交换系统的效率: 如果某用户正在发送, 那么其他用户的发送将不被允许。这种方式称为载波侦听多址(CSMA)。它比 ALOHA 更高效, 因为它不会打搅已处于发送状态的其他用户。

载波侦听多址系统最重要的参数是检测延迟和传播延迟。检测延迟是发射端检测信道目前是否被占用所需的时间的相对度量。该参数基本上取决于系统硬件, 但也与期望的虚警概率及信噪比有关。传播延迟是一个数据分组从移动台传送到基站所需的时长。如果在时刻 t_1 , 发射机 TX1 检测出信道空闲, 并开始发送一个分组。在时刻 t_2 另一个发射机 TX2 检测信道。如果 $t_2 - t_1$ 小于数据分组 A 从 TX1 传送到 TX2 所需的时长, 则 TX2 检测到的信道状况是空闲的, 随即它会发送数据分组 B。在这种情况下, 就会发生碰撞。以上描述说明检测延迟和传播延迟应该比分组持续时间小得多。

载波侦听多址的实现

有多种不同的方法来实现载波侦听多址。最流行的方法如下[Rappaport 1996]。

- 非持续载波侦听多址。发送端检测信道。若信道正忙, 则发射机等待一个随机时间以后再次准备发送。
- p -持续载波侦听多址。此方法用于时隙信道。当发送端检测到信道可用后, 它将以概率 p 在随后的时隙里发送分组, 否则将在一个时隙后进行发送。
- 1 -持续载波侦听多址。发送端持续地检测信道, 直到认为信道空闲为止; 然后会立即发送分组。显然, 这是 p 持续载波侦听多址发送在 $p=1$ 时的特例。
- 带碰撞检测的载波侦听多址(CSMA/CD)。此方法中, 节点侦测是否有两个发射机同时开始发送。若存在这一情况, 将立即停止发送。在无线分组通信时这种方法并不适用。
- 数据侦听多址(DSMA)。此方法中, 下行链路包含一条控制信道, 此信道以周期性间隔发送“忙/闲”信号来指示信道状态。如果用户发现信道空闲, 则可以立即发送数据分

组。要注意的是,对于端到端的网络,控制信道的实现比具有中央节点(基站)的场合困难得多。

17.4.3 分组预约多址

在分组预约多址(PRMA)中,每个移动台可以为传送数据分组发出请求。由一种控制机制(可以是集中式或分布式的)来回应请求,以告知移动台何时被允许发送分组。这样就消除了数据分组碰撞的危险,但携带发送请求的信号可能发生碰撞。而且,此系统牺牲了一些发送容量用于预约请求的传送。为保证适当的效率,预约请求必须比实际的数据分组短得多。此方法又称为分离信道预约多址(SRMA)。

此方法的一种变形是用户可以持续使用传输媒体,直到将一个数据块发完为止。其他的一些方法则区分优先级(随时间改变),以避免单个用户长时间独占传输媒体的情况出现。

另一个重要的方法是轮询。此方法中,基站逐个查问(轮询)各移动台是否想要发送数据分组。最短的轮询周期出现在没有移动台要发送信息的时候,这是最低效的情况,既因为轮询牺牲了容量,又没有有效数据被发送。

17.4.4 各方式之间的比较

图 17.6 显示了不同分组交换方式的效率。横坐标为信道的利用率,纵坐标是平均的分组延迟。可以看出 ALOHA 方式仅适用于对信道使用的有效程度要求不高 的情形:时隙 ALOHA 系统的最大可获得容量只有 36%。载波侦听多址和轮询可得到更好的结果。

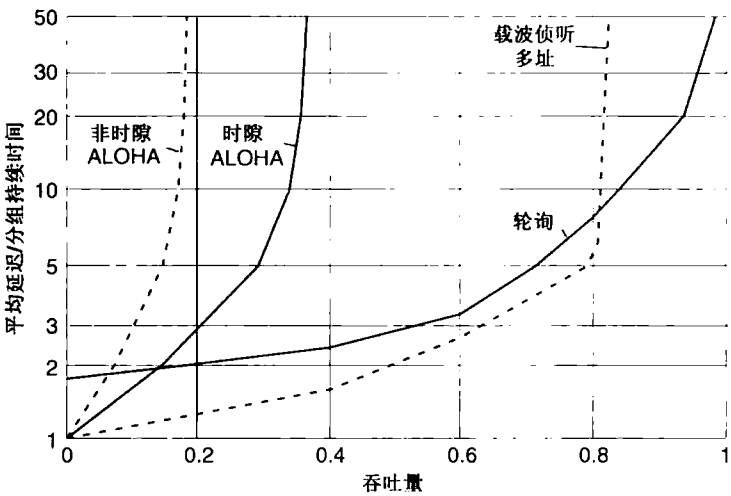


图 17.6 不同分组多址方式的信道使用率和平均分组延迟。引自 Oehrvik[1994]© Ericsson AB

将分组无线电与时分多址和频分多址进行比较时,会发现后两种方案对于传送话音非常有用,因为话音一般要求低的延迟。编码后的语音数据应该在话音被说出之后的 100 ms 以内到达接收端。采用频分多址和适当时隙间隔的时分多址很容易做到这一点。此外,由于信道(频率或时隙)是专门保留的,每个发射机在线路上不发生明显的阻塞或延迟的情况下就能可靠地将数据发送到接收端。另一方面,时分多址和频分多址浪费了资源,尤其是在传送数据时。即使单个用户没有任何数据要发送,也总是要为这一用户保留一个信道(时隙)。

17.4.5 分组无线电的路由问题

现在描述分组无线系统中对数据分组的路由(见图 17.7),此原理对有线和无线系统均适用。一个分组交换系统在发送端和接收端之间建立逻辑连接,但与电路交换系统相比,这里没必要保有固定的物理连接。唯一重要的就是用某种方式把分组由发送端传送到接收端,实际选择的路由(物理连接)则是可以随时间而变的。为达到这一目的,消息被分解为小块(分组),每个分组可以通过不同的路由到达接收端,这取决于当前可获得的传送路径。因而,分组也就可以通过不同的网络节点选择路由,每个节点决定了分组怎样传递。如果目前没有传输路径,则分组将被缓存,直至出现可用路径为止。缓存会导致相当大的传输延迟,并会使数据到达顺序不同于其发送顺序。鉴于此,通过许多节点传递(数据)的分组无线电不易被用于语音传送。然而,基于网际协议的语音(VoIP)电话的出现表明基于分组的语音电话是可以实现的。

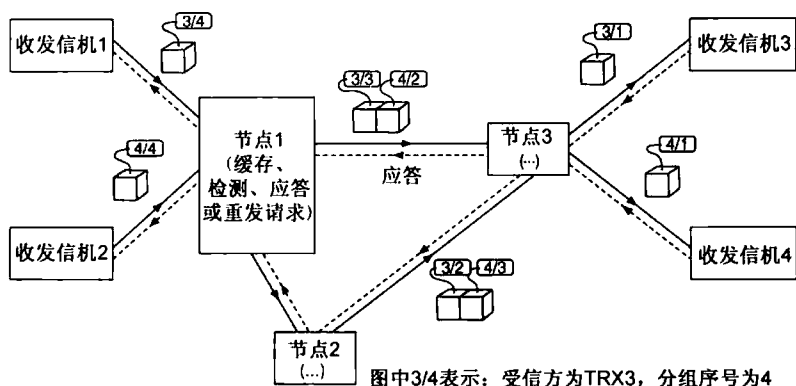


图 17.7 分组无线网络原理。引自 Oehrvik[1994]© Ericsson AB

数据分组包含承载数据和路由信息(见图 17.8)。路由信息清楚地表明了消息的发出方和目的方,并且关于缓存和分组采用的路径的额外信息也会包括在内。

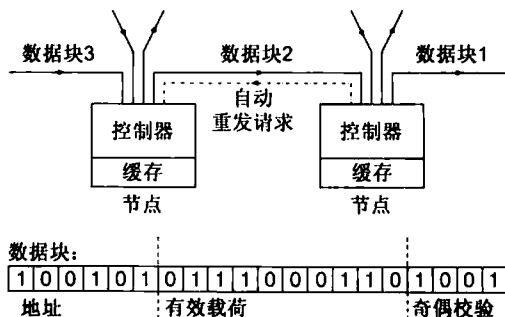


图 17.8 数据块结构及缓冲机制原理。引自 Oehrvik[1994]© Ericsson AB

一般来说,路由方式被分成以下两类。

- 源驱动路由。在这种情况下,分组的报头中包括完整的路由信息,相关的节点只需按照该信息的指示向前传递分组。缺点在于,报头可能会很长,尤其对于有效载荷很少的分组,因此就导致了谱效率的明显下降。

- 表驱动路由。按照这种方法，每个节点保存了一张向前传送分组时的可用节点列表(由分组的来源节点和目的地址决定)。这一方法有较高的谱效率，缺点在于列表可能很庞大，尤其是在位于网络中部的那些节点处。

相关的主题是“路由发现”，即确定分组应采用自发送端到接收端的哪一条路由。路由发现的典型实现方式是通过在整个网络广播特殊的分组，然后记录不同节点之间的链路质量。为保证获得最优的性能，当节点之间的信道发生明显变化时，路由也必须进行相应的变化。

如果想要得到非常低的分组错误率，那么每个节点都将作为一个中继站点，既可以在缓存里保存分组，又可以在收到分组传送的前方节点回应的成功传送信息之后即删除分组(见图 17.8)。

22.4 节将更详细地讨论无线自组织网络中的路由。

17.5 双工

在蜂窝系统中，双工用于分离上行链路(移动台到基站)和下行链路(基站到移动台)。我们将其区分为时域双工(Time Domain Duplexing, TDD)和频域双工(Frequency Domain Duplexing, FDD)。在时域双工中，上行链路的数据发送时段不同于下行链路的发送时段，如图 17.9(a)所示。在频域双工中，上行链路和下行链路在不同的频带发送，如图 17.9(b)所示。

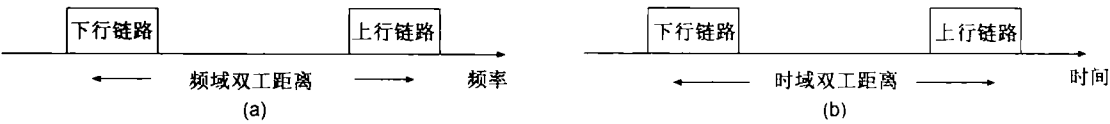


图 17.9 频域双工和时域双工

时域双工常与时分多址一同使用。在这种情形下，可用时间被分为 $2N$ 个时隙，而不是 N 个时隙，这样 N 个用户中的每一个都将被分配一个用于上行链路的时隙及另一个用于下行链路的时隙。时域双工和分组无线电方案相结合也可以很好地发挥作用。另一方面，时域双工用于频分多址系统会抵消频分多址固有的很多优点(如连续的发和收，使同步得以简化)，却保留了其缺点。

频域双工可以与任何多址方式结合使用。大多数情况下，频域双工有固定的双工距离(即，上行链路频带与下行链路频带之间的频率差是固定的)。

对于时域双工和频域双工，关于上行链路和下行链路是否一致的问题是人们所感兴趣的。对于时域双工系统，就是要求双工的时间间隔比相干时间小得多；而对于频域双工系统，就是要求频域双工距离比相关带宽小得多(更多细节在 20.1.6 节给出)。考虑实际系统参数，前一个条件可以很好地得到满足，尤其是在准静态(基站和移动台都不移动)的环境下运作的无线局域网。后一个针对频域双工的条件实际上根本无法满足。然而，当上、下行链路的信道相一致时将带来相当大的好处：在第 19 章将会看到保证这一点可以简化自适应调制的实现，在第 20 章会看到它会有利于不同的智能天线系统的实现。

另一方面，时域双工系统在发送和接收之间将经历一个无效时间(dead time)。假定用户在离基站距离为 d 的地方(因而传输时延为 d/c_0)发送，在时刻 T 停止发送，则数据将在时刻 $T + d/c_0$ 完整到达基站。仅当此时基站才能切换到发送模式来发送它自己的数据块。因为数据又将传输回移动台，它们将在 $T + 2d/c_0$ 时刻开始到达移动台处。因此，移动台已经历了一个长

为 $2d/c_0$ 的无效时间。当小区尺寸较大时,这将对系统造成相当可观的效率损失。

频域双工和时域双工的结合发生在半双工系统情形中。在半双工系统中,单个移动台处的发送和接收在时间和频率上都将加以区分。虽然在采用频域双工时同时发送是能够做到的,但需要高质量的双工滤波器。如果还能确保发送在不同的时间进行,也就不必使用那么好的双工滤波器了。半双工的一个例子就是全球移动通信系统,在该系统中发射机和接收机工作所采用的频带是不同的,而且收发彼此之间存在 2 个时隙的偏移(见第 24 章)。

17.6 蜂窝网络原理

17.6.1 重用距离

现在讨论一个无线系统如何覆盖一个大的区域,并能向该区域内尽可能多的用户提供服务。第一个移动无线系统实际上是服务于少量用户的噪声受限系统。因此,将每个基站设置于山巅或高塔上是有好处的,这样就能为大区域提供无线覆盖。下一个基站离得如此之远,以至于干扰弱到不构成影响。然而,这种办法严重限制了可同时通信的用户数。本节将要描述的蜂窝原理为上述问题的解决提供了答案。本节将使用在每个小区中用频分多址作为多址方案的例子,同样的论述也适用于时分多址。码分多址系统将在第 18 章进行讨论。

在蜂窝系统中,覆盖区域被分成许多小的区域,称为小区(cell)。在每个小区中,有一个基站为该小区的区域提供覆盖,且仅覆盖该小区区域。这样,每个频道就可用于多个小区。自然会提出的问题是:是否可将每个频道都用到每个小区中?一般地,答案是“不可以”。设想用户 A 处在当前向其提供服务的小区边界处的情形,这样从“有用”基站及其一个相邻基站到用户的距离是相等的。如果相邻基站采用相同频道(目的是与自己小区中的用户 B 通信),那么用户 A 处的信干比(SIR)为 $C/I=0$ dB。这肯定不足以支持可靠通信,尤其因为 0 dB 还是中值信干比,衰落会使得 50% 的时间里情况更差。

这一问题的解决方案是不在每个小区中重用每个频道,而仅在相互之间有一定的最小距离的小区中这样做。两个可使用相同频道的小区之间的归一化距离称为重用距离, D/R 。重用距离可以从链路预算(见 3.2 节)计算得到。还可以定义由完全使用不同频率的若干个相互邻接的小区形成的区群(cluster);因此,在一个区群里就可以不存在同频干扰。一个区群中包含的小区数称为区群大小。整个覆盖区域被划分为许多这样的区群。

区群大小也决定了蜂窝系统的容量^①。一个获得 35 个频道牌照的运营商,采用区群大小 7 时,可在每个小区中同时支持 5 个用户。因而,容量的最大化就要求区群大小的最小化。区群大小为 1(即,在每个小区中使用每个频率)是终极目标;然而,如上所述,在频分多址系统中,由于需要考虑链路容限,这是不可能做到的。模拟频分多址系统的信干比典型值为 18 dB,这就导致了其区群大小为 21 的结果(下文将详述)。数字系统(如全球移动通信系统)需要 10 dB 以下的值,从而将区群大小减小为 7 或更小的数。这就使容量能有一个惊人的提升,从而成为 20 世纪 90 年代初蜂窝系统从模拟向数字迁移的一个最重要的原因。

17.6.2 小区形状

小区通常会呈现什么形状?先考虑路径损耗与方向无关而仅取决于与基站之间的距离的

^① 本章提到的容量指的是可同时支持的用户或通信设备的数目,而不是信息论意义上的容量。

理想情况。最自然的选择就是盘状(圆形),因为它提供了小区边界上处处相同的接收功率。然而,盘状小区不能既无缝隙又无重叠地填满一个面。另一方面,正六边形形状近似于圆形,并且可以像在蜂房图案中一样填满一个面。因而,正六边形常常作为“基本”的小区形状,尤其是在理论研究中。

然而,必须强调,正六边形结构仅在以下情形下才成为可能。

- 需要的业务密度独立于位置。当人口密度变化时,此条件显然会被违背。
- 地形是完全平的,并且没有高大的建筑物,这样路径损耗才会仅受到与基站之间距离的影响。这一条件在欧洲是无法满足的,而且在世界其他地方也仅有少数地方可以满足,如美国中东部、西伯利亚冰原、沙漠。图 17.10 显示了地形和从位于典型的丘陵地形的基站获得的功率情况。按上述简单化的考量,等功率线应该是围绕基站的同心圆。实际的结果决非如此,功率在一些方向上甚至并不必然是距离的单调函数。因此,实际的小区规划需要计算机仿真或者对接收功率的测量。第 7 章描述的建模技术因而成为现实小区规划的至关重要的基础。

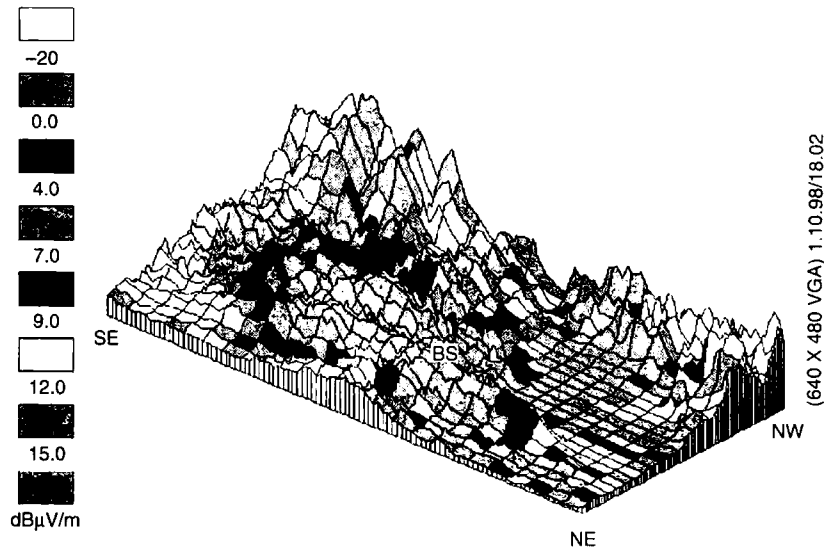


图 17.10 典型的丘陵地形接收功率状况示例(奥地利维也纳西部)(引自 Buehler[1994]© H. Buehler)

17.6.3 六边形小区的小区规划

对于六边形情形,关于链路容限和重用距离之间的关系可以得出一些有趣的结论。考虑中心位于坐标系原点的六边形小区,沿 y 轴方向前进 i 个小区,然后逆时针旋转 60° 在新的方向上再前进 k 个六边形(见图 17.11)。这样,所到达的小区中心与原点之间的距离如下:

$$D = \sqrt{3} \sqrt{(iR + \cos(60^\circ)kR)^2 + (\sin(60^\circ)kR)^2} \tag{17.14}$$

注意,两个相邻小区的中心的距离为 $\sqrt{3}R$, R 是六边形中心到其最远的顶点的距离。还应注意, i 和 k 只能取整数值。

频率规划的任务就是寻找那些使得由式(17.14)确定的距离大于所需重用距离的 i 和 k 值。当然,有无限多种这样的数值对(大的 i 和 k 值肯定是满足条件的)。我们想要寻找的是可使区群大小最小化的数值对,因为这样可以在满足最小重用距离时,使频谱效率最大化。

由于六边形的布局, 区群大小 N 和参数 i 和 k 之间的关系为

$$N = i^2 + ik + k^2, \quad i, k = 0, 1, 2, \dots \quad (17.15)$$

该公式也表明了并非所有整数都能成为可能的区群大小值。区群大小的可能取值有 $N = 1, 3, 4, 7, 9, 12, 13, 16, 19, 21, \dots$ 。由式(17.14)和式(17.15), 重用距离 D/R 和区群大小之间的关系为

$$D/R = \sqrt{3N} \quad (17.16)$$

图 17.12 也表明了这种关系。表 17.2 给出了一些典型的区群大小和重用距离。

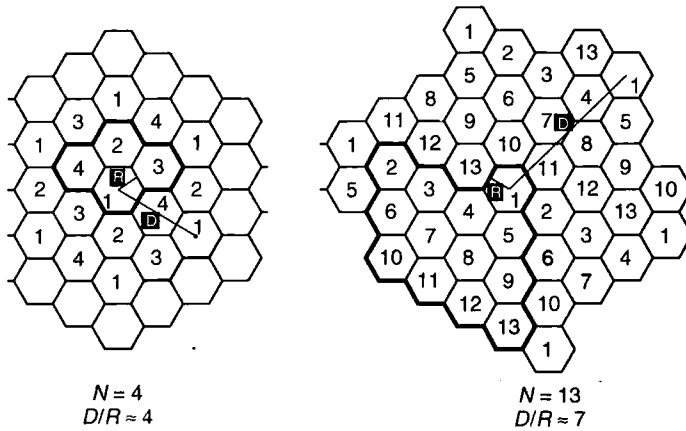


图 17.11 重用距离 D/R 和区群大小 N 的关系。引自 Oehrvik[1994]© Ericsson AB

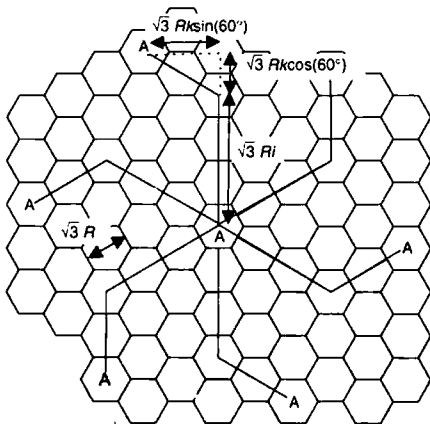


图 17.12 最小重用距离

表 17.2 全向天线下的典型区群大小和重用距离

N	D/R	
1		码分多址系统
3	3	
4	3.46	
7	4.58	时分多址系统
9	5.2	
12		
13	6.24	
16	6.93	
19	7.55	
21	7.94	模拟系统(NMT, AMPS)

NMT 代表北欧移动电话; AMPS 代表先进移动电话系统

因此, 小区规划的步骤如下: 先由最低传输质量的指标, 用链路预算(见第 3 章)确定选定基站和干扰基站之间的最小距离; 然后, 由式(17.16)可以得到区群大小, 注意, 该区群大小应该是由式(17.15)定义的一组值中最小的整数值; 再采用前述方法(即在一个方向上行进 i 个小区, 转 60° , 再行进 k 个小区)确定最近的同频小区, 每个小区使用的频率随之得到确定。

例 17.3 干扰受限系统的小区规划。

解:

小区规划的原理和典型数值可以通过一个具体的例子得到最好的理解。为简化分析和探

讨论频分多址的实际系统,考虑模拟先进移动电话系统的有关系统数据。每个频道为 30 kHz 宽,信干比为 18 dB 就可以达到满意的语音质量。衰落容限(阴影加瑞利衰落)为 15 dB。这意味着在小区边界处,信号功率的中值应该比干扰功率强 33 dB(或 2×10^3 倍)。选定基站与六边形最远顶点之间的距离为 R ,干扰基站与这一顶点之间的距离近似为 $D - R$ 。进而假定功率按照 d^{-4} 衰减,可得

$$\frac{D - R}{R} = (2 \cdot 10^3)^{1/4} = 6.7 \quad (17.17)$$

因此,重用距离应该为 $D/R = 7.7$ 。按照式(17.16),重用距离等于 $(3N)^{0.5}$,所以区群大小为 19.8。满足式(17.15)且 $N \geq 19.8$ 的最小整数值 $N = 21$ 。现在假定某运营商获得了 5 MHz 频谱执照,这样共有 $5000/30 = 167$ 个可用频道,每个小区仅可使用 $167/21 \approx 8$ 个频道。

17.6.4 增加容量的方法

系统容量是蜂窝网络最重要的指标。于是,增加容量的方法就成为了基本的研究领域。下面将给出一个简明的概括,当然这一概括常常会涉及到本书的其他章节。

1. 增加可用频谱数量。这是一种“蛮力”方法。此方法代价高昂,因为频谱是稀缺资源,并且其使用权常常由政府以极高的价格拍卖掉。而且,分配给无线系统频谱总量的变化却非常缓慢,所分配频谱的改变必须获得国际管制会议的许可,而这种会议常常十年或更长时间才召开一次。
2. 更高效的调制制式和编码。采用需要更少的传输带宽(高阶调制)或兼具更强的抗干扰能力的调制制式。前者将允许增加每个用户的数据速率(或者当每个用户数据率为常数时,可以增加小区内的用户数)。然而,高阶调制的可能优势将由于以下原因而受到限制:它们对噪声和干扰更敏感(见第 12 章),因此可能不得不增加重用距离。抗干扰调制方式的采用将允许重用距离的减小。另一种获得更好的抗干扰性能的方法是,引入像 Turbo 码和低密度奇偶校验码(LDPC)这样的接近于香农容量限的码(见第 14 章),因而也可以增加系统容量。
3. 更好的信源编码。取决于需要的语音质量,目前的语音编码速率在 32 kbps 和 4 kbps 之间。更好的语音属性模型将允许数据速率降低而同时不带来语音质量的下降(见第 15 章)。数据文件和音乐/视频的压缩也将允许为更多用户提供服务。
4. 非连续传输(DTX)。非连续传输的采用,基于这一事实:电话交谈期间每个通话者的实际讲话时间仅为其全部通信持续时间的 50%。藉此,一个时分多址系统可以建立多于可用时隙数的通话连接。在通话期间,某一时段正在讲话的用户可以与相同时段不讲话的用户(系统将在此时不再分配无线资源给他)复用可用的时隙。
5. 多用户检测。多用户检测将大大降低干扰的影响。因而,在码分多址系统中,也就允许每小区支持更多的用户;或者,在频分多址系统中,可以采用更小的重用距离(详见第 18 章)。
6. 自适应调制和编码。发送端利用传输信道的信息,选择“刚好”适用于当前链路状况的调制制式和编码速率。这一方法更好地使用了可用功率,并且作为多种效果之一,降低了干扰(详见第 19 章)。
7. 减小小区半径。这是一种效果明显却又非常昂贵的增大容量的方法,因为额外的每个

小区都需要建立新的基站。对于频分多址系统,也将意味着重新制定一个很大区域的频率规划^①。而且,对于移动的用户,更小的小区也需要更多的切换,这样做不仅复杂,而且会降低频谱效率(由于在切换期间有大量信令信息必须被传送)。

8. 使用扇区。六边形(或相似形状)的小区可划分为多个(典型地为 3 个)扇区。每个扇区有一个扇区天线来实现覆盖。因此,小区(指扇区)数目成为以前的 3 倍,与基站天线数目相同。然而,基站站址数不变,因为 3 个天线位于同一位置(见图 17.13)。

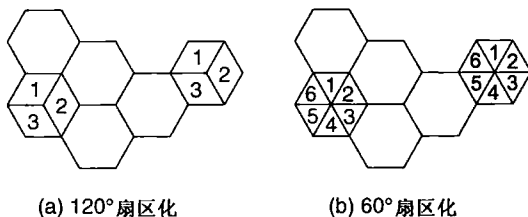


图 17.13 扇区原理

9. 采用重叠结构。重叠结构组合了不同尺寸、不同业务密度的小区,所以有些位置会有几个小区同时向其提供服务。伞状小区对大的区域提供基本覆盖,在其覆盖区域内多个微小区被布置在高业务密度的区域。在微小区的覆盖区域中,大多数用户由微小区的基站向其提供服务,而快速移动的用户将由伞状小区提供服务,以便减少小区之间的切换次数。2010 年前后,引入了意在安装于公寓和办公室的所谓飞蜂窝(femtocell)。这些飞蜂窝的基站利用飞蜂窝所在的用户房屋中的互联网连接(电缆、数字用户线)来接入蜂窝网络。这种飞蜂窝的主要目的是给安装它的公寓/办公室提供好的覆盖/高的数据速率。主要的难点就在于如何将飞蜂窝集成到整个网络中;由于可以预期在不远的将来会出现大量这类蜂窝小区,看上去自组织网络(Self Organizing Network, SON)是进行这种集成的最佳方式。
10. 多天线。可藉由以下不同的情形来增进容量。
 - (a) 分集(见第 13 章)提高了接收信号的质量,可用来提高容量,如采用高阶调制制式或减小重用距离。
 - (b) 多输入多输出系统(见 20.2 节)增加了每个链路的容量。
 - (c) 空分多址(见 20.1 节)允许同一小区里获得服务的几个用户采用相同的频率。
11. 部分承载。这种系统采用小的重用距离,但每个小区仅使用一小部分的可用时隙。这将导致它能够获得与“传统的”方案(大的重用距离和每个小区的“完全承载”)近似相等的平均容量。然而,部分承载具有较高的灵活性,因为当另一些小区吞吐量较低时,某些小区可以有更高的吞吐量。
12. 部分频率重用。在这一方案中,可获得的频谱被分成 $N + 1$ 个子频带。一个子频带被用于所有的小区中心(cell center),而采用传统的频率重用方式(区群大小为 N),将其他子频带都用在小区边沿(cell edge)。“小区边沿”必须足够地大,以保证从一个小区中心到另一个小区中心的干扰足够微弱。还需要指出,所有的子频带不必具有相

^① 对于六边形小区的情形,可以得到进行准确的小区分裂的同时又不必对频率配置进行大规模的重新规划的算法 [Lee 1986]。

同的带宽。根据小区中心的不同尺寸,小区中心采用的子频带可能宽于小区边沿处采用的子频带。

17.7 附录

请访问 www.wiley.com/go/molisch。

深入阅读

频分多址和时分多址系统是“经典”的多址方案,如今它们大多在特定系统(特别是 GSM,见第 24 章)的语境中被分析;Falconer et al. [1995] 中同样有着非常好的总结。Xu et al. [2000] 中给出了干扰下的有趣的性能分析;Rubin [1979] 中分析了时分多址和频分多址系统中的消息延迟。Sari et al. [2000] 讨论了由时分多址/频分多址向码分多址迁移的问题。排队问题在 Gross and Harris [1998] 这本优秀教材中讨论。关于业务分布的讨论可以在诸如 Ashtiani et al. [2003] 的文献中找到。

ALOHA 系统在 Abramson [1970] 中首先被提出,其无线应用的分析由 Namislo [1984] 给出。载波侦听多址的经典描述由 Kleinrock and Tobagi [1975] 及 Tobagi [1980] 给出。分组预约多址在 Goodman et al. [1989] 中进行了描述。

蜂窝原理和蜂窝无线电的基本概念在 Lee [1995] 中有详尽的描述。这一描述试图涵盖模拟和数字系统,尽管现在多少有些过时了,但所阐述的原理仍然有效;Alouini and Goldsmith [1999] 论及更高级的方面,包括衰落效应。Chan [1992] 讨论了扇区化的影响。Frullone et al. [1996] 给出了高级小区规划技术的全面评述;Woerner et al. [1994] 描述了关于蜂窝系统和网络规划仿真的问题。Katzela and Naghshineh [2000] 中(以 2000 年的研究状态)描述了意在减轻干扰的频率规划与调度。尤其是与第四代系统有关的更新一代的技术,在 Boudreau et al. [2009] 和 Necker [2008] 中进行了综述。在诸如 Chandrasekhar et al. [2008] 和 2009 年 9 月号的 *IEEE Communications Magazine* 的几篇文章中,对飞蜂窝进行了描述。

第18章 扩展频谱系统

扩展频谱技术将信息扩展到非常宽的带宽上,确切地说,是比数据速率大得多的带宽。本章将讨论通过扩展频谱来提供多址的不同方式。首先从概念上最简单的方式,即跳频开始,然后对应用最广泛的扩展频谱方式,即直接序列码分多址(DS-CDMA)展开讨论。最后,将详细描述跳时脉冲无线电,这是一种相对较新的方案,因其在超宽带系统中的应用而成为近年来人们研究兴趣的焦点。

前面的章节已强调了频谱效率的重要性,即希望在单位可用带宽上传输尽可能多的信息。所以,在商业无线系统中,将信息扩展至较宽的带宽,看起来是个有些奇怪的想法。毕竟,术语“扩展频谱”来源于军事领域,在这一领域里人们的兴趣所在集中于通信的保密性、抗截获性及抗敌方发射机干扰作用的能力,而这些问题在商业蜂窝系统运营商所关注的问题中并不是最重要的^①。因此,扩谱方式在无线通信中已取得如此重要的位置好像有些不可思议。

当我们意识到不同的用户可用不同的方式被扩展到同一频谱时,这个看起来似非而是的困惑迎刃而解。扩频方式下允许多个用户采用同一频带同时发送,接收机可以通过仅对具有特定扩展规律的信号的查找来确定在总的接收成分中哪一部分来自某一特定用户。因此,每单位带宽的容量在采用扩频技术的情况下未必会有所下降,相反,通过充分发挥扩频技术的特定优势,用户容量甚至会有所增加。

18.1 跳频多址

18.1.1 跳频原理

跳频(FH)的基本思想是改变一个窄带传输系统的载波频率,使得在某一频带上的传输仅持续较短的时间。载波频率跳变的总的带宽范围和窄带传输带宽的比值是其扩频因子。

跳频源于军事通信,由女演员海蒂·拉玛(Hadi Lamar)在二战期间发明。其灵感来源于以下问题:无线发射机的发射可能被敌方用来对发射机进行三角定位,同时窄带发送也可能被敌方的大功率(窄带)发射机干扰。通过频繁地改变发射载频,信号在每个“脆弱的”(即容易被发现和干扰的)频带上仅持续较短的时间。跳频图案(频率跳变的规律)必须为本方接收机所知,而又是敌方所无法预测的,使得敌方无法“跟踪”跳变。除了可以抑制窄带干扰,跳频也有助于减轻深度衰落沉陷的影响。这样一来,有时系统在一个“好”的频率上发送,即发射机和接收机之间的衰减较小且干扰较弱;有时又会工作在一个“坏”频率上,即经历衰落引起的陷落甚或还有较强的干扰。

跳频有两种基本类型:慢跳频和快跳频。快跳频在一个符号传送期间多次改变载频;换句话说,单个符号的发送被扩展到较大的带宽上,从而使衰落或干扰的影响能在每个符号上分别得到缓解。根据基本傅里叶变换关系,传输一个符号的每一部分将需要比窄带系统更宽的带

① 虽然安全性(抗截获性)的确重要,但也可通过加密手段来保证这一点,而未必要采用扩频技术。

宽。此外,要整合属于同一符号的不同成分就必须采用比符号速率更快的处理手段。快跳频在民用无线系统中未能得到广泛应用,主要已被码分多址所替代。

慢跳频在一个频率上传送一个或多个符号。此方式常与时分多址结合使用,即每个时隙在一个给定载频上发送,下一个时隙则改变至一个不同频率。在这一情况下,同步所需的额外开销较小,因为接收机在下一时隙无论如何必须同步。为了发挥出跳频的有效性,交织和编码用于将来自同一源比特的信息分布到几个时隙上。假定采用简单的重复编码,则每比特在不同时隙上(也就是在不同载频上)被重复发送两次。即使第一个时隙的发送经历了深度衰落,第二个时隙发送使用的频率极有可能经历较小的信道衰减;这样一来,信息就可能得到有效恢复。例如,慢跳频在全球移动通信系统(GSM)有所应用(见第24章)。

一般来说,跳频导致接收信号特征的“白化”。这种作用具有隐含的信道衰减平均化效果。而且,在每一个载频上,出现一个不同的干扰,以至于跳频也带来了对所有干扰的平均化。对于许多类型的接收机来说,这是有益的,因为这样降低了出现“灾难性”状况的概率,同时降低了所需的链路容限。然而,在18.4节和20.1节将会看到,有些接收机构造实际上可以利用干扰本身的特定(已知)结构来消除它。例如,智能天线就可以利用干扰信号的空间结构来形成天线的方向图,以使干扰来源方向上出现零陷。跳频会使这些技术很难应用。

18.1.2 跳频多址(FHMA)

前一节从抑制干扰和引入频率分集的角度讨论了跳频。得到以上好处所付出的代价是传输时不得不采用更宽的带宽,这看上去有些浪费。下面将会说明跳频可以作为一种多址方式来应用,并能获得类似时分多址和频分多址的频谱效率。为此,将跳频系统区分为同步的和异步的。

首先考虑同步的情况,例如,在蜂窝系统的下行链路,基站总是可以保证同时向所有移动台发射。图18.1显示了使用3个可用频带(载频)的例子。显然,一个时间间隔期间,基站可以同时向3个用户发送,但为了便于讲解,假定只有两个用户(用户A和用户B)正在使用系统。在第一个时间间隔(跳频周期)里,基站用频带2向移动台A发送。同时,频带3空闲,因此基站向移动台B发送时可以使用此频带。在下一个时间间隔里,基站则采用频带1向移动台A发送,而用频带2向移动台B发送。在第三个时段里,用频带3为移动台A服务,对移动台B则使用频带1。在随后的时间间隔里重复整个跳频序列。所有移动台的信号都采用同一跳频序列^①,这样可以确保跳变时相互之间不会发生任何碰撞。因此,同步的跳频多址显然具有与频分多址相同的容量,同时还增加了频率分集的好处。为了将同样的概念用于上行链路,所有的移动台必须采用某种方式来发送各自的信号,这种方式可以保证信号同步到达基站,因而就能在上行链路复现图18.1所示的情况。要达到这一效果,需要每部移动台到基站的空中传输时间信息,此信息使每个移动台确切地知道自己的发送该始于何时(定时提前,见第24章)。

当用户之间不同步时,情况将有所不同,这种情况发生在定时提前量无法预测(当虑及小区之间的干扰时)^②的简单网络中,或者发生在无线自组织网络中。对于异步的情况,所有用

① 每个移动台使用的是同一序列的一个不同的时移版本。——译者注

② 对于某一基站而言,通过适当的设计可以保证各个用户的上行链路信号同步到达,但对于另一基站而言,这些信号却未必能同步到达。

户采用相同的跳频序列是不明智的。由于同步的缺失,不同用户信号之间的延迟可能为任意值,包括零延迟的情况。如果所有用户使用同一跳频序列,则零延迟将导致“灾难性的碰撞”,即不同用户之间时时刻刻都相互干扰。为了避免出现这类问题,每个用户将采用不同的跳频序列(见图 18.2)。这些跳频序列按照每个跳变循环(即跳频序列重复出现的时间间隔,本例中的跳变循环是跳频周期的 3 倍)里恰好有相当于一个时段(跳频周期)的时间按被干扰的要求来进行设计,而其余的时间则不会发生碰撞^①。显然,这种系统的性能要比同步系统(或频分多址系统)差。提供低碰撞概率保证的跳频序列设计是十分富有技巧性的工作,仍然是目前的研究热点。

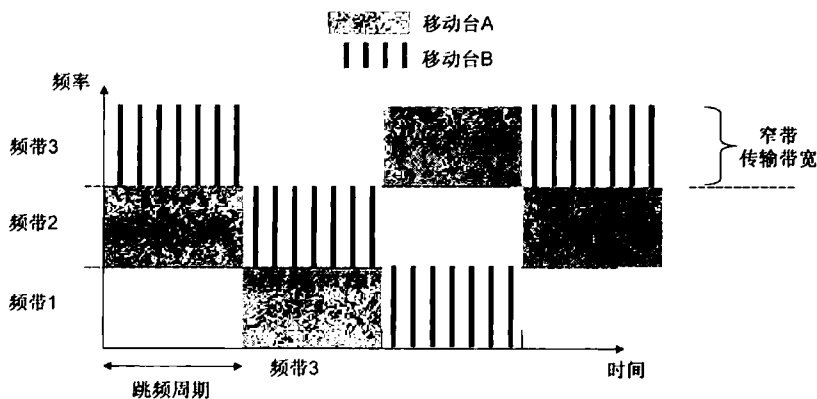


图 18.1 同步用户跳频多址原理

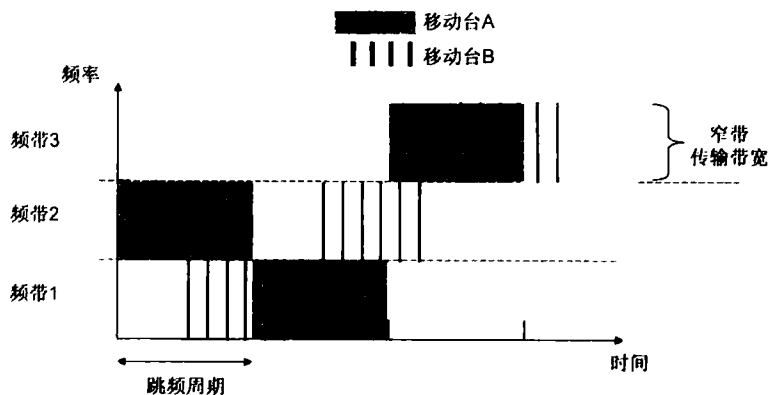


图 18.2 异步用户跳频多址原理

18.2 码分多址

码分多址(CDMA)的起源,特别是直接序列扩频(DS-SS)的发展也可以追溯到军事通信的研究。在 18.2.1 节将讨论基本的扩频操作。然后,18.2.2 节和 18.2.3 节描述码分多址系统如何采用直接序列扩频实现多址的原理和数学表示。18.2.4 节分析码分多址在频率选择性信道中的表现。18.2.5 节和 18.2.6 节讨论同步和码分多址扩频码的选择。在蜂窝网络中应用的更多细节在 18.3 节叙述。

^① 此处,不发生碰撞的意思是“用户 B 不会引发碰撞”,其他用户仍然可能造成干扰。

18.2.1 直接序列扩频的基本原理

直接序列扩频通过将发送信号与具有很宽带宽的第二个信号相乘来实现对信号的扩展。这个合成信号的带宽与宽带扩频信号(即上述第二个信号)的带宽近似相等。新信号的带宽与原始信号带宽的比值也称为扩频因子。由于扩频信号的带宽很宽,而发送功率却保持不变,因而发送信号的功率谱密度非常低,取决于扩频因子和基站与移动台之间距离的具体取值,甚至有可能低于噪声功率谱密度。这一点对于军事应用很重要,因为非法接收机因此就无法确定是否有信号正在被传送。另一方面,合法接收机则可以实现扩频操作的逆操作,也就可以恢复出窄带信号(其功率谱密度远大于噪声功率)。

图 18.3 显示了直接序列扩频发射机的框图。信息序列(可能已进行过编码)与一个用扩频序列调制正弦载波信号后生成的宽带信号相乘。也可以等效地将这一过程理解为:在调制之前先把每个持续时间为 T_s 的信息符号与一个扩频序列 $p(t)$ 相乘^①。此后都假定扩频序列长为 M_c 个码片,这里每个码片的持续时间为 $T_c = T_s/M_c$ 。由于带宽是码片持续时间的倒数,于是合成信号的带宽就也是 $W = 1/T_c = M_c/T_s$,即带宽宽至窄带调制信号带宽的 M_c 倍。由于假定扩频操作不改变总的发射功率,也就意味着功率谱密度将减小为原来的 $1/M_c$ 。

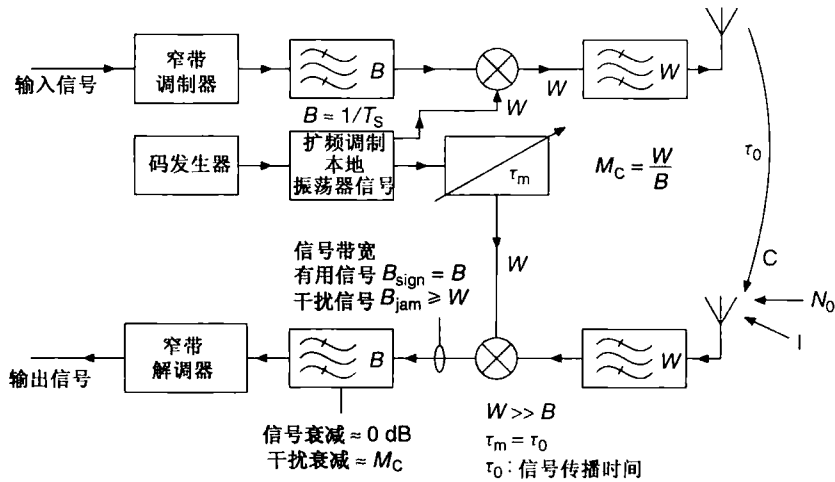


图 18.3 直接序列扩频发射机和接收机框图

在接收机处就应该进行扩频操作的逆操作。这可以通过将接收信号与扩频序列进行相关运算来完成。这一过程可以逆转发送端对带宽的扩展,因此在相关运算之后,所需的信号又恢复了 $1/T_s$ 的带宽。除了所需的信号以外,接收信号里还含有噪声、其他的宽带干扰及可能存在的窄带干扰。应该注意到解扩(despreading)操作不会对噪声和宽带干扰的有效带宽带来明显的影响,而同时窄带干扰实际上却被扩展到带宽 W 上了。作为解扩操作的一部分,相关后的信号将通过一个带宽为 $B = 1/T_s$ 的低通滤波器。这将使所需的信号在通过滤波器之后基本保持不变,而噪声、宽带干扰和窄带干扰的功率降低为滤波前的 $1/M_c$ 。在符号解调器处,直接序列扩频也就具有了和窄带系统一样的信噪比。对于窄带系统,解调器处的噪声功率为 N_0/T_s 。对于直接序列扩频系统,接收机输入端的噪声功率为 $N_0/T_c = N_0 M_c/T_s$,会被窄带滤波减

① 此理解基于窄带信号和宽带信号采用相同的调制方式。

弱(减小为 $1/M_c$)，因此检测器的输入端的噪声功率也为 N_0/T_s 。对宽带干扰而言，也会产生相似的效果。

下面来讨论直接序列扩频系统的扩频序列。为了通过相关运算完全地逆转扩频操作，我们希望扩频序列的自相关函数呈现出狄拉克 δ 函数的特性。在这种情况下，原始信息序列与扩频器和解扩器的级联系统相卷积后的输出就是原始序列本身。因此，我们希望 $p(t)$ 在时刻 iT_c 的自相关函数值 $ACF(i)$ 为

$$ACF(i) = \begin{cases} M_c, & i = 0 \\ 0, & \text{其他} \end{cases} \quad (18.1)$$

实际上，这些理想化特性只能近似达到。一类具有这种近似特性的码序列是称为最大长度序列(m 序列)的一种伪噪声序列。该伪噪声序列具有如下的自相关函数：

$$ACF(i) = \begin{cases} M_c, & i = 0 \\ -1, & \text{其他} \end{cases} \quad (18.2)$$

如图 18.4 所示。

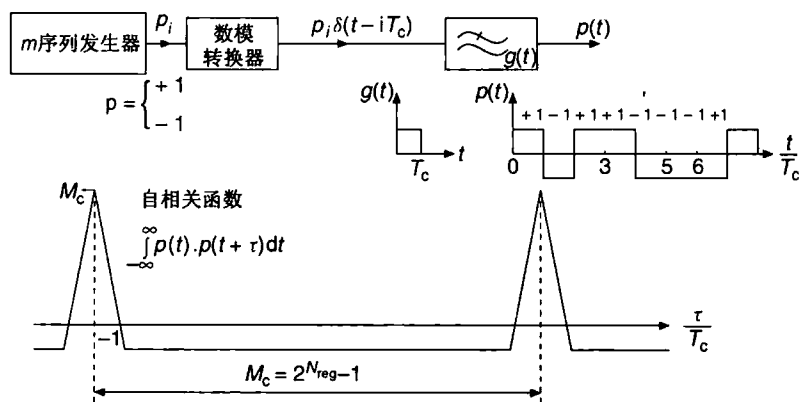


图 18.4 m 序列的自相关函数。引自 Oehrvik[1994]© Ericsson AB

18.2.2 多址

直接序列扩频操作本身，即与宽带信号相乘，可视为隐蔽通信的一种调制方式，并因此具有重要的军事应用价值。人们采用基于直接序列扩频的码分多址，利用扩频来获得多址能力。每个用户被分配以不同的扩频码，该扩频码决定了与信息符号相乘的宽带信号。因而，许多用户可以同时在一个宽频带上发送(见图 18.5)。

在接收机处，可以通过将收到的信号与拟接收用户的扩频信号进行相关来获得所需的信号。因此，其他用户信号就成为了宽带干扰；通过解扩器之后，到达检测器输入端的干扰功率数值等于干扰用户的扩频序列和拟接收用户扩频序列的互相关函数(CCF)值。因此，最理想的情况下，我们希望不同用户扩频序列的互相关满足下式：

$$CCF_{j,k}(t) = 0, \quad j \neq k \quad (18.3)$$

上式应该对所有的用户 j 和 k 都满足。换句话说，需要码序列相互正交。就至多 M_c 个扩频序

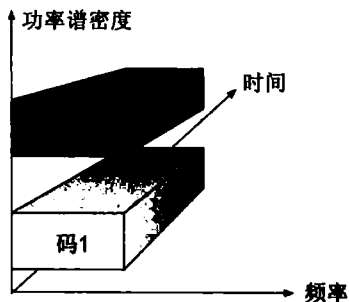


图 18.5 码分多址原理

列而言,是可以做到相互之间完全正交的;这是因为如下事实: M_c 个正交序列可以构建一个 M_c 维空间,任何持续期为 M_c 个码片的其他序列都可以表示为它们的线性组合。

如果扩频序列相互不正交,接收机就只能做到有限的干扰抑制,即干扰抑制因子为自相关函数与互相关函数之比。如果不同的扩频序列是移位 m 序列,则抑制因子为 M_c 。

基于以上描述,可以得出如下一些重要结论。

- 扩频序列的挑选是保证码分多址系统质量的根本因素。序列必须具有良好的自相关函数(类似于狄拉克 δ 函数)和小的互相关值。 m 序列是一个可能的选择;采用长度为 N_{seq} 的移位寄存器就可能产生出 $2^{N_{\text{seq}}} - 1$ 个^①序列,这些序列具有式(18.2)给出的自相关函数,且自相关函数与互相关函数之比等于 M_c 。其他序列还包括 Gold 和 Kasami 序列,将在 18.2.6 节讨论。
- 码分多址需要精确的功率控制。如果自相关函数与互相关函数之比为有限数值,接收机将不能很好地抑制干扰。当接收到的干扰功率比来自于所期望的发射机的功率大得多时,就超出了解扩接收机的干扰抑制能力。因此,每个移动台都必须按照如下原则来调整自身的发射功率:使到达基站的所有信号的功率大致相等。经验表明,要完全发挥出码分多址系统的理论容量,功率控制必须精确到约 ± 1 dB 以内。

18.2.3 数学表示

下面把以上的定性描述转变为数学体系[Viterbi 1995]。为此,假定采用二进制相移键控调制方式,并假定发射机和接收机之间可以达成完全同步(更多细节下面会讨论到),然后就可以在接收机处得到如下 4 个信号分量。

- 所需的(用户)信号。以 $c_{i,k}$ 表示第 k 个用户的第 i 个信息符号, $r_{i,k}$ 表示相应的接收信号。假定符号间干扰、码片间干扰(随后定义)及噪声都是零均值过程,则接收信号的期望值将正比于发送符号:

$$E\{r_{i,k}|c_{i,k}\} = \sqrt{(E_c)_k} c_{i,k} \int_{-\infty}^{\infty} |H_R(f)|^2 df \quad (18.4)$$

其中, $(E_c)_k$ 是第 k 个用户的码片能量,而 $H_R(f)$ 是归一化为 $\int_{-\infty}^{\infty} |H_R(f)|^2 df = 1$ 的接收滤波器的传递函数。

- 码片间干扰。接收滤波器具有有限持续时间的冲激响应,所以一个码片与这一冲激响应相卷积后,卷积结果的持续时间将比一个码片长。因此,接收滤波器以后的信号就表现出码片间干扰(稍后将会看到信道的延迟扩展也会导致码片间干扰)。如果扩频序列都是零均值的,则码片间干扰会使接收信号的方差增大,所增加的数值为

$$(E_c)_k \sum_{i \neq 0} \left[\int_{-\infty}^{\infty} \cos(2\pi i f T_C) |H_R(f)|^2 df \right]^2 \quad (18.5)$$

噪声使接收信号方差增加了 $N_0/2$ 。

① 依照线性反馈移位寄存器序列理论,此处的序列数应该有误,因为移位寄存器阶数一定时的最大长度序列个数应该没有那么多!但原作者此处的意思为,某一 m 序列的任何一个移位序列都能与原始序列保持好的互相关特性,故如果将它们都用来作为扩频序列,也是可行的。这样可用序列个数就是 $2^{N_{\text{seq}}} - 1$ 。——译者注

- 同道干扰 (CCI)。当施加干扰的用户们的发送码片独立于所需用户的数据符号和码片时, 假定同道干扰不改变接收信号的均值。同道干扰使得方差增加了:

$$\sum_{j \neq k} \frac{(E_C)_j}{2T_C} \int_{-\infty}^{\infty} |H_R(f)|^4 df \quad (18.6)$$

现在, 如果假定码片间干扰和同道干扰都是近似高斯的, 则计算错误概率的问题就简化为在高斯噪声中检测信号的标准问题:

$$\text{BER} = Q \left(\sqrt{\frac{(E_C)_k M_C}{\text{总的方差}}} \right) \quad (18.7)$$

在衰落信道中, 所需信号的能量是不断变化的。对于上行链路, 功率控制确保这些变化得到相应的补偿。

18.2.4 码分多址情况下的多径传播效应

以上关于码分多址系统的相当简化的描述都假定信道是平坦衰落信道。在所有实际环境下, 这一假定都是不成立的。码分多址系统的基本属性就是把信号扩展到很宽的带宽上, 因此可以预期的是, 信道的传递函数在这个带宽上将呈现出某些变化。

频率选择性(延迟扩展)对于一个码分多址系统所产生的效应可以通过探究扩频器、信道、解扩器的级联系统的冲激响应来理解。如果信道是随时间缓慢变化的, 那么这个等效的冲激响应可以写成如下形式^①:

$$h_{\text{eff}}(t_i, \tau) = \tilde{p}(\tau) * h(t_i, \tau) \quad (18.8)$$

其中, 等效系统冲激响应是发送和接收扩频序列的卷积:

$$\tilde{p}(\tau) = p_{\text{TX}}(\tau) * p_{\text{RX}}(\tau) = \text{ACF}(\tau) \quad (18.9)$$

以下假定扩频序列是满足式(18.1)的理想序列, 解扩器输出就会呈现出多个峰值, 更确切地说, 接收机可分辨的每个多径分量(间隔至少为 T_C)对应于一个峰值。每个峰值都含有与发送信号有关的信息。因此, 所有峰值在检测过程中都应该被用到, 因为仅使用最大相关峰值就意味着丢弃了到达信号中的许多有用信息。一种可以利用多个相关峰值的接收机称为 Rake 接收机^②, 它搜集[把集(Rake up)]来自不同多径分量的能量。如图 18.6 所示, 一个 Rake 接收机包含一组相关器。每个相关器获取不同时间段上的接收数据(不同时间段之间具有时延 τ), 也就是搜集彼此时延为 τ 的不同多径分量的能量。之后, 对各相关器所获取到的数据进行加权与合并。

此外, 还可以将 Rake 接收机解释成抽头延迟线, 各级延迟器的输出经加权后再合并起来。各级延迟器的延迟时间和抽头处的权值都是可调的, 并与信道相匹配。注意, 抽头之间通常至少间隔一个码片的持续时间, 但抽头之间不必是均匀间隔的。接收滤波器和 Rake 接收机合起来构成了一个与接收信号匹配的滤波器。接收滤波器与发送信号匹配, 而 Rake 接收机与信道匹配。

不管怎样解释, 接收机都会将来自不同 Rake 分支(Rake fingers)的已加权信号以相干的方式叠加起来。因为这些信号对应于不同的多径分量, 所以它们的衰落是(近似)统计独立的。

① 注意, 这一表达式与相关信道探测器(见第 8 章)的表达式完全相同。相关信道探测器和码分多址系统具有相同的结构, 只不过目标不同: 信道探测器试图由接收信号和已知的发送数据来确定信道的冲激响应, 而码分多址接收机则试图由接收信号和估计出的信道冲激响应来确定发送数据。

② Rake 的英文含义为耙子。——译者注

换句话说,它们提供了延迟分集(频率分集)。因此,Rake 接收机就是一种分集接收机,而且所有用于分集性能分析的数学方法同样适用于 Rake 接收机。至于 Rake 接收系统的性能,可以直接参照 13.4 节和 13.5 节的有关公式。

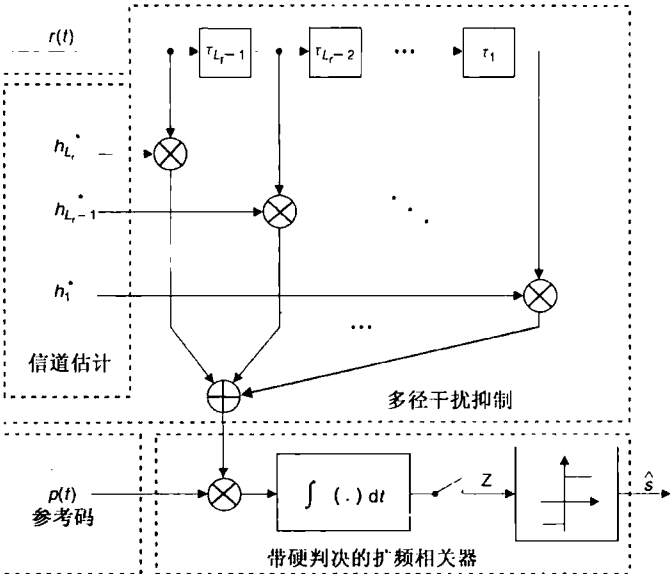


图 18.6 Rake 接收机。引自 Molisch[2000]© Prentice Hall

例 18.1 Rake 接收机的性能: 计算以下两种情况的二进制相移键控的误比特率: (i) 窄带系统; (ii) 码分多址系统: 使用 6 分支的 Rake 接收机, 可以分辨所有的多径, 接收机前端信噪比为 15 dB, 信道模型为 ITU Pedestrian-A^①。

解:

ITU Pedestrian-A 信道的抽头延迟线模型为

$$|h(n)|_{\text{dB}} = [0 \quad -9.7 \quad -19.2 \quad -22.8] \tag{18.10}$$

$$|h(n)| = [1 \quad 0.3273 \quad 0.1096 \quad 0.0724] \tag{18.11}$$

平坦衰落信道的平均信道增益为

$$\Sigma |h(n)|^2 = 1 + 0.33^2 + 0.11^2 + 0.07^2 = 1.1 \tag{18.12}$$

从而发送信噪比应该为

$$\bar{\gamma}_{\text{TX}} = \frac{10^{1.5}}{1.1} = 28.75 \tag{18.13}$$

以保证 15 dB 的接收机前端信噪比。

由第 12 章可知, 平坦衰落信道下的误比特率为

$$\overline{\text{BER}} = E[P_{\text{BER}}(\gamma_{\text{Flat}})] = \int_0^{\pi/2} \frac{1}{\pi} M_{\gamma_{\text{Flat}}} \left(-\frac{1}{\sin^2 \theta} \right) d\theta \tag{18.14}$$

进而

$$\overline{\text{BER}} = \int_0^{\pi/2} \frac{1}{\pi} \frac{\sin^2 \theta}{\sin^2 \theta + \bar{\gamma}_{\text{TX}} \Sigma |h(n)|^2} d\theta = 7.724 \times 10^{-3} \tag{18.15}$$

^① pedestrian 意为步行者。——译者注

当将信号合并起来(正如 Rake 接收机中所做的)时有

$$\gamma_{\text{Rake}} = \gamma_1 + \cdots + \gamma_6 \quad (18.16)$$

由于只有 4 个多径分量携带有能量, 所以实际上只使用了 4 个 Rake 分支。如果 $\gamma_1, \cdots, \gamma_4$ 是相互独立的, 则联合概率密度函数 $f_{\gamma_1, \cdots, \gamma_4}(\gamma_1, \cdots, \gamma_4) = f_{\gamma_1}(\gamma_1), \cdots, f_{\gamma_4}(\gamma_4)$ [也可以参照式(13.39)]:

$$\begin{aligned} \overline{\text{BER}} &= \int d\gamma_1 \text{pdf}_{\gamma_1}(\gamma_1) \int d\gamma_2 \text{pdf}_{\gamma_2}(\gamma_2) \cdots \int d\gamma_4 \text{pdf}_{\gamma_4}(\gamma_4) \int_0^{\pi/2} d\theta f_1(\theta) \prod_{k=1}^{N_r} \exp(-\gamma_k f_2(\theta)) \\ &= \int_0^{\pi/2} \frac{1}{\pi} \prod_{k=1}^4 \int_{\gamma_k} f_{\gamma_k}(\gamma_k) e^{\left(-\frac{\gamma_k}{\sin^2 \theta}\right)} d\gamma_k d\theta \\ &= \int_0^{\pi/2} \frac{1}{\pi} \prod_{k=1}^4 M_{\gamma_k} \left(-\frac{1}{\sin^2 \theta} \right) d\theta \end{aligned} \quad (18.17)$$

因此,

$$\overline{\text{BER}} = \int_0^{\pi/2} \frac{1}{\pi} \prod_{k=1}^4 \left[\frac{\sin^2(\theta)}{\sin^2(\theta) + \bar{\gamma}_k} \right] d\theta \quad (18.18)$$

对于同样大小的发送信噪比 $\bar{\gamma}_{\text{TX}}$, 可得

$$\begin{aligned} \overline{\text{BER}} &= \int_0^{\pi/2} \frac{1}{\pi} \frac{\sin^2 \theta}{\sin^2 \theta + \bar{\gamma}_{\text{TX}}} \frac{\sin^2 \theta}{\sin^2 \theta + 0.33^2 \bar{\gamma}_{\text{TX}}} \frac{\sin^2 \theta}{\sin^2 \theta + 0.1^2 \bar{\gamma}_{\text{TX}}} \frac{\sin^2 \theta}{\sin^2 \theta + 0.07^2 \bar{\gamma}_{\text{TX}}} d\theta \\ &= 9.9 \times 10^{-4} \end{aligned} \quad (18.19)$$

采用延迟分集解释之后, 要做的另一件事就是为各 Rake 分支输出的合并确定分支权重。最优权值是最大比值合并方式所对应的权值, 即每个 Rake 分支所对应的多径分量幅度的复共轭值。然而, 只有在可以为每个可分解多径分量分配一个 Rake 分支时, 才可能取得最优权重(在文献中已广泛使用术语“全 Rake”(all Rake)来命名这种接收机)。这种情况下需要多达 $L_r = \tau_{\text{max}}/T_c$ 阶, τ_{max} 是信道最大附加时延(见第 6 章)。尤其对于室外环境, 这个阶数很容易超过 20。然而, 实际 Rake 合并器的可实现阶数受到功耗、设计复杂度和信道估计的限制。一个仅处理能够使用的 L_r 个可分辨多径分量的某个子集的 Rake 接收机可以达到降低复杂度的目的, 同时仍然可以提供优于单径接收机的性能。选择式 Rake (SRake) 接收机选择 L_r 个最佳路径 (L_r 个能够使用的可分辨多径分量的一个子集), 然后采用最大比值合并方式来合并被选出的子集中的各个多径分量。这种合并方式是“选择/最大比值混合式合并”(如第 13 章讨论到的)。然而应当注意, 不同分集支路的平均功率是不同的。要注意 SRake 仍需所有多径分量的瞬时取值信息, 以便进行恰当的选择。部分 Rake (PRake) 是另一种可能的形式, 它使用前 L_r 个多径分量。尽管所提供的性能差一些, 但它只需要估计 L_r 个多径分量。

Rake 接收机的另一个具有普遍性的重要问题是径间干扰(interpath interference)。在抽头延迟线中, 相对于经调节所得的某 Rake 分支的延迟值, 还具有额外的时间延迟(设延迟量为 τ_i)的那些多径分量, 在经过该分支的相关处理之后, 它们的影响将受到抑制, 抑制因子为 $\text{ACF}(0)/\text{ACF}(\tau_i)$, 只有在扩频序列具有理想自相关函数特性时, 该抑制因子才为无穷大。因此, 采用非理想扩频序列的 Rake 接收机将会发生径间干扰。

最后要指出, 为使 Rake 接收机达到最优, 还应当不存在符号间干扰(ISI)。也就是说, 尽

管信道最大附加时延可以比 T_c 大,但它必须远小于 T_s 。如果存在符号间干扰,则在接收机中除了 Rake 接收机以外还必须具备均衡器(工作于 Rake 输出之后,即均衡器的信号采样间隔为 T_s)。一种针对 Rake 接收机和符号间隔均衡器的这一合并形式的替代方案就是基于码片的均衡器,这种均衡器直接工作于以码片速率采样的解扩器的输出端。这种方式是最优的,但非常复杂。正如在第 16 章所指出的,随着采样频率和信道时延扩展乘积的增大,均衡器的运算量迅速增加。

18.2.5 同步

同步是码分多址系统的一个最重要的实际问题。从数学角度讲,同步就是从由可能数值组成的无穷大集合(即连续时间)中确定最佳采样时刻的估计问题。将同步问题分成如下两个局部问题会更有利于同步的实现。

- 捕获(Acquisition)。这是第一步,用来确定最佳取样时刻位于哪个(持续时间为 T_c 或 $T_c/2$ 的)时间间隔。这是一个假设-检验(hypothesis-testing)问题:将检验有限数目的假设,其中每个假设都假定了取样时刻位于某特定间隔内的一种情况。对于众多假设的检验可以并行地进行,也可以串行地进行。
- 跟踪(Tracking)。一旦时间间隔被确定了,可以采用一个控制循环来将取样时刻精确调整至其准确数值。

就捕获阶段而言,可以采用一种特殊的同步序列,该序列要比数据传输期间所使用的扩频序列短。这就减少了必须被检验的假设数目,因而也就减少了同步所必须花费的时间。此外,还将同步序列设计成具有非常好的自相关特性的序列。就跟踪部分而言,可以直接利用用于数据传输的普通扩频序列。

对系统设计的许多方面来说,确定来自不同用户的信号是否彼此同步也是很重要的。

- 小区内同步。一个基站发出的多个信号总是同步的,因为基站在发送它们时具有控制权。对于上行链路,多个信号同步抵达基站通常是无法实现的。因为这需要所有移动台安排它们的定时提前量(timing advance),即当它们开始发送一个码序列时,用这种方式来使所有信号同时到达基站。定时提前量应该精确到一个码片持续时间以内。对于大多数应用来说,这显得过于复杂了,尤其是因为移动台的运动会导致所需定时提前量的变化。
- 基站之间的同步。基站之间彼此可以同步。这通常借助全球定位系统(GPS)提供的定时信号来实现,所以每个基站都需要一部 GPS 接收机,还应当存在与几个 GPS 卫星之间的自由视距。虽然前一条件往往并不成为问题,后一条件在微蜂窝(microcell)和皮蜂窝(picocell)中却不易得到满足。IS-95 系统(见第 25 章)采用了这种同步方式。

18.2.6 码族

选择标准

扩频码的选择对于码分多址系统性能有着至关重要的影响。前面作为例子使用的 m 序列至今仍然是众多扩频码中的一员,但的确并非唯一的可能选择。一般地,扩频码的质量由以下性质决定。

- 自相关。理想情况下, $ACF(0)$ 应该等于每符号的码片数目 M_c , 而在其他所有场合下的自相关函数都等于 0。就 m 序列而言, $ACF(0) = M_c$, 而对于 $n \neq 0$, $ACF(n) = -1$ 。好的自相关函数特性对于同步也是有用的。此外, 如上面所讨论到的, 自相关性会影响 Rake 接收机中的码片间干扰: 相关器的输出是各延迟分量的自相关函数之和, 自相关函数中的假相关峰看上去像是额外的多径分量^①。
- 互相关。理想情况下, 所有的用户码应该彼此正交, 这样就可以将来自其他用户的干扰完全抑制掉。下文将讨论一个正交码族 (Walsh-Hadamard 码)。对于非同步系统, 必须不同用户之间存在任意延迟的情况下都能够满足正交性。注意, 带宽扩展和用户区分可以采用不同的码来完成; 这时将用到的码区分为扩频码和扰码。
- 码的数目。一个码分多址系统应该支持尽可能多的用户同时通信。这就意味着要有大量的可用码。正交码的数目受到 M_c 的限制。如果需要更多的码, 就必须接受更糟的互相关特性。在相邻小区所使用的码也应有所不同的现实要求下, 情形会变得更加复杂: 不同的用户终究只能以不同的码来区分。这样一来, 如果小区 A 有 M_c 个用户使用了所有的正交码, 那么相邻小区 B 中使用的码就无法做到与小区 A 中使用的码保持正交关系; 因此, 小区之间的干扰是无法完全抑制掉的。然而, 我们可以按照如下原则来选择在小区 B 中使用的码: 这些码对原始小区 (如小区 A) 中用户所形成的干扰应具有类噪声的特点; 这种选取方式就要求以码规划来替代频率规划。另一个可行方案是产生大量的具有次优互相关函数的码 (下文将会讨论到), 然后将它们分配到任何需要使用的地方。在这种情形下, 不需要进行码规划, 但系统容量会更低。

伪噪声序列

m 序列、Gold 序列和 Kasami 序列是码分多址系统上行链路最常用的扩频序列。 m 序列的自相关性质优异; 一个 m 序列的移位版本仍然是具有近乎理想的互相关性质的有效码字。可通过对几个 m 序列进行适当合并来生成 Gold 序列; Gold 序列的自相关函数可以呈现出 3 种可能数值, 其自相关函数和互相关函数的非峰值函数值都可以被上界所限定。

一个更具普遍性的序列族是 Kasami 序列。下面会区分 S (小) 型、L (大) 型和 VL (非常大) 型 Kasami 序列。这些大写字母说明了码族中码数的多少。S-Kasami 序列具有最佳的互相关函数, 但只存在数量相当的这种序列。VL-Kasami 序列互相关函数最差, 但这种序列的数目几乎是无限的。表 18.1 给出了不同码族的性质。

表 18.1 码分多址所使用码字的性质

序列	码数	最大互相关函数 (dB)	注解
m 序列	$2^{N_{reg}} - 1$		良好自相关函数
Gold	$2^{N_{reg}} + 1$	$\approx -3N_{reg}/2 + 1.5$	
S-Kasami	$2^{N_{reg}/2}$	$\approx -3N_{reg}/2$	在所有 Kasami 序列中具有最佳互相关函数
L-Kasami	$2^{N_{reg}/2} (2^{N_{reg}} + 1)$	$\approx -3N_{reg}/2 + 3$	
VL-Kasami	$2^{N_{reg}/2} (2^{N_{reg}} + 1)^2$	$\approx -3N_{reg}/2 + 6$	近乎无限的数目

N_{reg} 是用于生成序列的移位寄存器的阶数。

① 对接收机而言自相关函数是已知的, 因此能够从剩余信号中将属于第一个检测到的多径分量的假相关峰去除, 从而能够消除它们可能产生的影响 (这实质上就是干扰抵消技术, 18.4 节会讨论到它)。

Walsh-Hadamard 码

在下行链路中,由于通过同一部基站发射机来发射所有的信号,所以能使属于不同用户的信号完全同步。为此,一个所有码字彼此完全正交的码族可以用 Walsh-Hadamard 矩阵给出。由 n 阶 Hadamard 阵可以定义 $n+1$ 阶阵 $\mathbf{H}_{\text{had}}^{(n+1)}$:

$$\mathbf{H}_{\text{had}}^{(n+1)} = \begin{pmatrix} \mathbf{H}_{\text{had}}^{(n)} & \mathbf{H}_{\text{had}}^{(n)} \\ \mathbf{H}_{\text{had}}^{(n)} & \bar{\mathbf{H}}_{\text{had}}^{(n)} \end{pmatrix} \quad (18.20)$$

其中, $\bar{\mathbf{H}}$ 是 $\mathbf{H}$ 的模 2 取补。这一递推公式的初始矩阵是

$$\mathbf{H}_{\text{had}}^{(1)} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (18.21)$$

这个矩阵的各列代表了长度为 2 的所有可能 Walsh-Hadamard 码字。各列相互正交是显而易见的。由递推公式可得

$$\mathbf{H}_{\text{had}}^{(2)} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \quad (18.22)$$

这个矩阵的各列是周期(duration)为 4 的所有可能码字;并且容易看出它们彼此都是正交的。进一步迭代可以得到另外的码字,每次所得序列的长度是前一矩阵所对应序列长度的两倍。

如果信号在加性高斯白噪声信道中传输,则正交码可以保证在接收机处实现完美的多用户抑制。时延色散破坏了码字之间的正交性。于是,接收机或者可以承受延迟扩展带来的附加干扰(用正交性因子来描述干扰程度),或者可以在进行相关运算(用以实现用户分离)之前先让接收信号通过一个码片间隔均衡器,以消除延迟扩展的影响。

例 18.2 频率选择性信道中的正交可变扩频因子(OVSF)码的正交性表现为:码字 $[1 \ 1 \ 1 \ 1]$ 和 $[1 \ -1 \ 1 \ -1]$ 经信道冲激响应为 $(0.8, -0.6)$ 的信道发送。假定发送的是 $[1 \ 1 \ 1 \ 1]$, 计算到达接收机的信号与可能的发送信号之间的相关系数。

解:

假定基站与两个移动台通信,所发送的扩频码是 $t_1(n) = [1 \ 1 \ 1 \ 1]$ 和 $t_2(n) = [1 \ -1 \ 1 \ -1]$ 。 $t_1(n)$ 和 $t_2(n)$ 彼此正交,有

$$t_1(n) \cdot t_2(n)^T = 0 \quad (18.23)$$

然而,时间色散信道(多径信道) $h(n) = [0.8 \ -0.6]$ 将影响两个码字之间的正交性。接收信号等于冲激响应和发送信号的线性卷积(假定信道是线性时不变的),有

$$r(n) = t(n) * h(n) = \sum_{k=-\infty}^{\infty} t(k)h(n-k) \quad (18.24)$$

当发送 $t_1(n)$ 时,接收信号 $r_1(n)$ 计算如下:

$$r_1(0) = 1 \cdot 0.8 = 0.8 \quad (18.25)$$

$$r_1(1) = 1 \cdot (-0.6) + 1 \cdot 0.8 = 0.2 \quad (18.26)$$

$$r_1(2) = 1 \cdot (-0.6) + 1 \cdot 0.8 = 0.2 \quad (18.27)$$

$$r_1(3) = 1 \cdot (-0.6) + 1 \cdot 0.8 = 0.2 \quad (18.28)$$

$$r_1(4) = 1 \cdot (-0.6) = -0.6 \quad (18.29)$$

因此, 将上面得到的接收信号 $r_1(n)$ 与 $t_1(n)$ 和 $t_2(n)$ 做相关, 可以得到相关系数如下:

$$\rho_{r_1 t_1} = \sum_{n=0}^3 r_1(n) t_1(n) \quad (18.30)$$

$$= [0.8 \ 0.2 \ 0.2 \ 0.2] [1 \ 1 \ 1 \ 1]^T = 1.4 \quad (18.31)$$

$$\rho_{r_1 t_2} = \sum_{n=0}^3 r_1(n) t_2(n) \quad (18.32)$$

$$= [0.8 \ 0.2 \ 0.2 \ 0.2] [1 \ -1 \ 1 \ -1]^T = 0.6 \neq 0 \quad (18.33)$$

类似地, 发送 $t_2(n)$ 时, 接收到的信号 $r_2(n) = [0.8 \ -1.4 \ 1.4 \ -1.4 \ 0.6]$ 。相关系数为

$$\rho_{r_2 t_1} = \sum_{n=0}^3 r_2(n) t_1(n) = 0.6 \neq 0 \quad (18.34)$$

$$\rho_{r_2 t_2} = \sum_{n=0}^3 r_2(n) t_2(n) = 5 \quad (18.35)$$

两个接收信号之间的相关结果为

$$\rho_{r_1 r_2} = \sum_{n=0}^3 r_1(n) r_2(n) \neq 0 \quad (18.36)$$

如果不同用户要求不同的数据速率, 将引出另一个问题: 需要采用不同长度的码字来进行扩频。正交可变扩频因子码是满足这些条件的一类码, 由 Walsh-Hadamard 码导出。

先来定义周期不同的码字之间的正交性。码片间隔对于所有码来说是相同的, 这由可用系统带宽确定, 而与要发送的数据速率无关。考虑两个码片长(码周期为 2)的码 $A(1, 1)$ 和四个码片长(码周期为 4)的码 $B(1, -1, -1, 1)$ 之间的正交性。如果码 B 作为相关器输入, 为了满足正交性, 则相关器 A 的输出必须是 0。也就是码 A 和码 B 的第一部分的相关结果必须为 0。这里的确如此: $1 \cdot 1 + 1 \cdot (-1) = 0$ 。同时正交性还要求, 码 A 和码 B 的第二部分的相关结果也必须为 0。这一点同样满足: $1 \cdot (-1) + 1 \cdot 1 = 0$ 。这时称码 A 和码 B 相互正交。

现在将不同 Walsh-Hadamard 矩阵的所有码字写到一个“码树”(见图 18.7)中。树中同一个级别(共同点是码周期相同)上的所有码彼此正交。如果码 A 和码 B 周期不同, 则仅当它们位于树的不同分支时才会彼此正交。当一个码(码 A)是另一个码(码 B)的“母码”时, 也就是码 A 位于从码树“树根”到码 B 的路径上时, 它们将不会彼此正交。这种不正交关系的例子是图 18.7 中的码 $p_{2,2}$

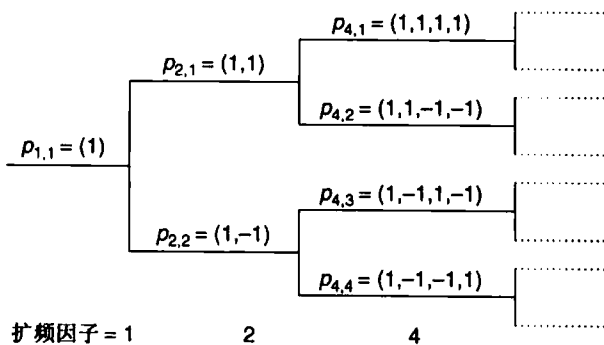


图 18.7 正交可变扩频因子码的码树

和码 $p_{4,4}$, 而与之相反, 码 $p_{2,2}$ 和码 $p_{4,1}$ 则是相互正交的。

18.3 蜂窝码分多址系统

18.3.1 码分多址原理回顾

分析一个系统的多址能力时,实质上是在提出这样一个问题:“是什么原因妨碍了同时向无穷多用户提供服务?”。在时分多址/频分多址系统中,问题的答案就是可用时隙/频率的有限数目。用户可以相互不发生干扰地占用那些时隙。而一旦所有可能的时隙都已经分配给了用户,就不再有空闲的可用资源了,系统也就不可能接受更多的用户。

在码分多址系统中,多址机制存在着微妙的差别。下面先分析上行链路,在上行链路中系统使用不完美但却拥有很大数量的扩频码。不同用户以不同的扩频码来区分;可是,由于用户分离不够完美,小区中的每个用户对其他所有用户都将产生干扰。因此,随着用户数目的增加,每个用户所遭受到的干扰也会增加。结果,传输质量逐渐下降(适度地降低),直到用户发现传输质量已经糟糕到不足以打出电话(或继续通话)为止。因此,码分多址在用户数目上所设定的是一个软的限制,而不是像时分多址那样的硬限制。从而,一个系统的用户数目从根本上取决于接收机所需的信干噪比(SINR)。同时也就意味着接收机处任何的 SINR 提升或接收机所需 SINR 门限值的下降,都能直接转变成更高的容量。

大多数干扰来自目标用户所在的同一小区内部,因此以小区内干扰(intracell interference)来命名。总的小区内干扰是许多独立作用之和,因此其表现近似类同于高斯噪声,小区内干扰所产生的影响类似于热噪声。通常用噪声提升(noise rise)来描述干扰的这一作用,即“有效”噪声功率(噪声和干扰功率之和)相对于单独的噪声有所增加: $(N_0 + I_0)/N_0$ 。图 18.8 给出了噪声提升作为系统负荷的函数的关系曲线,系统负荷定义为在用用户数与最大可能用户数 M_c 的比值。从图中可看出,当系统负荷接近 100% 时噪声提升也变得非常强。因此,通常在噪声提升为 6 dB 时将小区判定为“满员状态”。然而,如同上面所指出的,对在用用户的数目没有硬性限制。

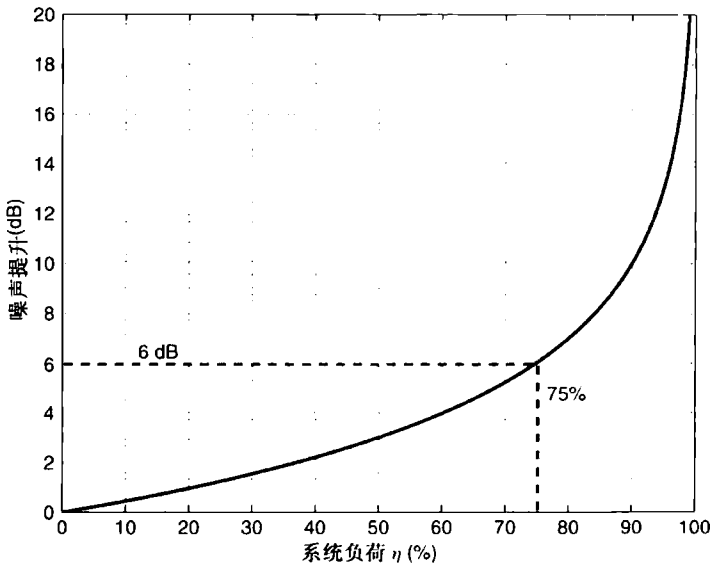


图 18.8 在码分多址系统中,噪声提升作为系统负荷的函数

有些干扰来自于相邻小区,因而称之为小区间干扰(intercell interference)。码分多址系统的一个主要特性就是它采用了完全的频率重用(universal frequency reuse),即重用距离为 1^①。换句话说,所有小区使用相同的频段,不同小区中的用户仅用不同的码来区分。正如 18.2.6 节所讨论的,干扰的程度主要由不同小区所采用的互不相同的码来决定。

码分多址系统的许多好处都与干扰有着近乎噪声的表现这一事实有关,尤其是在上行链路。这种类噪声的表现是由于以下几个原因。

- 每个小区的用户数目(也就是干扰数目)很大。
- 功率控制保证了到达基站的所有信号具有大致相等的强度(见本节后面的内容)。
- 相邻小区所引起的干扰也来自于大量用户。用于相邻小区的扩频码的设计原则是:某小区中的所有信号与所有相邻小区中的每一个信号都具有近似相等的互相关值。注意,这就意味着不能在每个小区中简单地重用同一码集(codeset);否则相邻小区中就会有个别用户造成比别的用户更多的干扰(由于个别用户采用了与目标小区中的目标用户完全一样的码)。

由于上述效应,总干扰功率显得几乎没有任何起伏。同时,功率控制保证了来自目标用户的信号强度总是恒定不变的。因此, SINR 也是恒定不变的,而且在链路预算中不需要考虑衰落容限。然而,要注意的是,就获得最大数据吞吐量而言,使干扰尽可能地表现出高斯型的特点并不总是最佳的策略。多用户检测(见 18.4 节)主动利用了干扰中的结构,并可以在只有少量强干扰源时取得最佳的用户检测效果。

在下行链路,扩频码都是彼此正交的,所以(至少在理论上)不同用户可以被完全分离。然而,对于下行链路的情况,小区中的用户数目会受到 Walsh-Hadamard 码数目的制约。这种情形就变得与时分多址系统有些相似了:如果 M_c 个可用的 Walsh-Hadamard 码用完了,另外的用户就不可能再得到服务。此外,还有相邻小区的 Walsh-Hadamard 码所引起的干扰。通过将 Walsh-Hadamard 码与扰码相乘可以使小区之间的干扰得到改善。与同一扰码相乘的 Walsh-Hadamard 码之间仍然保持了正交关系,与不同扰码相乘的码也不会引发灾难性的干扰。因此,不同小区采用不同的扰码(见第 26 章)。

在下行链路中,小区间干扰并非来自于大量独立的干扰源。所有干扰都来自于所考虑的目标移动台周围的基站,即少量(至多 6 个)基站构成了主要的干扰源。这些基站中的每一个都是为各自的小区服务的,向自己小区中的大量用户发送信号。这并不能改变如下现实:从被干扰移动台的角度看,某干扰信号仍旧出自于某个单独地理位置的干扰源^②,而每个干扰源与被干扰移动台之间具有一条单独的传播信道。因此,在下行链路的链路预算中就应该考虑衰落余量(fading margin)。

18.3.2 功率控制

如之前所指出的,功率控制至关重要,对于确保目标用户具有时不变的信号强度是这样的,对于保证来自其他用户的干扰呈现出类噪声特性也是这样的。为了进一步的研讨,必须区分上行链路和下行链路的功率控制。

① 此处括号中的原文为 also known as reuse distance one,意思是完全频率重用的情况也可以被称为 reuse distance one。在本书中,原作者对于重用距离的定义(见 17.6.3 节)是:重用距离 = D/R ,其中 R 为小区半径, D 为实际发生频率重用的两个最接近小区之间的距离。照此定义,完全频率重用就是 $D = R$ 的情况,所以重用距离为 1。但应该指出的是,其他的资料文献中也常将 D 直接定义为重用距离。——译者注

② 即上述各个基站,它们与被干扰台之间的地理位置关系有所不同。

- 上行链路功率控制。就上行链路而言,功率控制对于码分多址的正常运作是至关重要的。功率控制以闭环控制方式实现:先由移动台以特定功率发送,然后基站会告知移动台其先前的发送功率是偏高还是偏低了,移动台则据此调整其发送功率大小。控制环路的带宽要有所选择,以利于补偿小尺度衰落,即必须接近于多普勒频率。由于信道的时变性和信道估计中引入的噪声,到达基站的功率会存在一个剩余方差,这一方差的典型值约为1.5~2.5 dB,而待补偿的动态范围达到60 dB或更多。与具有理想功率控制的情况相比,这一方差的存在将导致码分多址蜂窝系统容量的下降,容量损失至多可达20%。

注意,在频域双工(FDD)系统中,无法用开环控制(移动台基于自己的信道估计来调整其发射功率)来补偿小尺度衰落:移动台在接收来自基站的信号时所面对的信道不同于它的发送信道(见17.5节)。然而,可以把开环和闭环结合起来使用。开环用于补偿信道中的大尺度变化(路径损耗和阴影),这种变化在上、下行链路的频率上是大致相同的。闭环则用于补偿小尺度变化。

- 下行链路功率控制。就下行链路而言,功率控制并非码分多址运行所必需的,因为来自基站的所有信号都以相同的功率到达同一移动台(信道同等对待所有的信号)。然而,为了保持较低的总发送功率,使用功率控制依然可以带来好处。将一个小区中所有用户的发送功率都降低相同的数量值,可以使有用信号功率和小区内干扰(即针对小区中其他用户所发送的那些信号对目标用户形成的干扰)的比值保持不变。然而,这的确降低了对其他小区所形成的总干扰功率。另一方面,由于信噪比绝对不能低于某一门限值,所以不能任意地降低信号功率。因此,下行链路功率控制的目标就是在保持误比特率或信干噪比值高于给定门限值的同时使总发送功率最小化。下行链路的功率控制的准确性不必像上行链路那么高,许多情况下,开环控制就足够了。

值得注意的是,相邻小区内部的用户功率控制并不能使小区间干扰的干扰功率保持恒定。某个相邻小区内的用户由该小区自己的基站进行功率控制,换句话说,用户依照到达目标基站的信号功率保持恒定的原则进行功率的调整。然而,用户将“面对”到达非目标基站的完全不同的信道,目标基站既不了解也不关心该信道随时间波动的具体情形。因此,小区间干扰是时变的。

只有当所有用户采用同样的数据速率时,来自他们的干扰功率才是相等的。高数据速率用户将“贡献”出更多的干扰功率,因此成为主要的干扰源。当通过向同一用户分配多个扩频码的方式来提高用户速率时,可以使上述结论得到最直观的理解。很显然,在这种方式下,被分配了多个扩频码的用户所“贡献”出的干扰功率将随着数据速率的升高而线性增加。虽然这种情况在只有语音用户的第二代蜂窝系统中不会发生,但却与宣称可提供高数据速率业务的第三代蜂窝系统必定有关。

还应指出的是,功率控制并非码分多址系统的专用属性;频分多址或时分多址系统也可以通过功率控制来降低小区间干扰,从而提高系统容量。主要差别在于功率控制对于码分多址是必不可少的,而对于时分多址或频分多址则是可选的。

软切换

当所有小区采用相同的频率时,一个移动台可以同时与两个基站保持联络。如果一个移动台靠近某个小区边界了,它会从两个或者更多的基站接收信号(见图18.9),同时也向所有

这些基站发送信号。来自不同基站的信号具有不同的时延,但可以用 Rake 接收机来补偿其影响,从而来自不同小区的信号可以彼此相干地叠加起来^①。这与基于频分多址的系统所采用的硬切换形成鲜明的对比,在那里一个移动台在某一时刻只能与一个基站保持联络,因为这一时刻它只能在一个频率上通信。

现在考虑如下情况:一个移动台起初在小区 A 中,并已与基站 A 进入通信状态,但也已与基站 B 建立了无线链路。刚开始时,移动台从基站 A 处收到最强的信号。当它开始向小区 B 移动时,来自于基站 A 的信号逐渐变弱,而来自于基站 B 的信号逐渐变强,直到系统决定断开到基站 A 的链路为止。软切换在移动台靠近两个小区边界时能够显著改善系统性能,这是因为它提供了可以克服大尺度衰落和小尺度衰落的分集(宏分集)。不利的一面,软

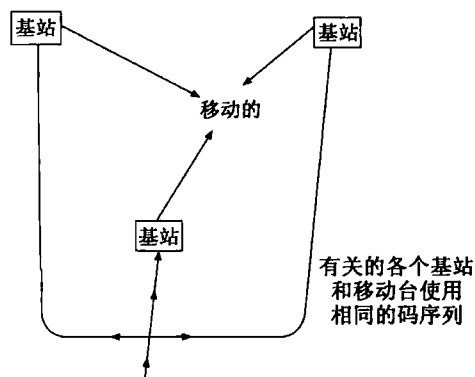


图 18.9 软切换原理。引自 Oehrvik[1994]© Ericsson AB

切换降低了下行链路的可用容量:软切换过程中,移动台需要同时占用两个小区的资源(Walsh-Hadamard 码),而用户却只进行了一次通话及付费。此外,软切换增加了基站之间所需的信令流量。

18.3.3 提升容量的方法

- 语音传输的静默期(Quiet period)。就语音传输而言,码分多址自然而然地利用了如下事实:一个人在整个通话期间并不是连续不停地讲话,而是只在约 50% 的时间里讲话,剩下的时间里他(她)在接听另一方的讲话。此外,人讲话时,在单词之间甚至音节之间都是有停顿的,所以“讲话时间”与“一次通话的总时间”的比值约为 0.4。在静默期期间,不必发送信号或者只需传送数据速率非常低的信号^②。在一个码分多址系统中,不进行信息发送就可以使总发送功率有所降低,因而系统中的干扰也会有所减弱。而前面已经看到,干扰功率的降低将支持更多的用户进行呼叫。小区中的所有用户同时处于通话状态无疑是可能发生的最差情况,然而从统计学角度讲,这种情况几乎不可能发生,尤其是在用户数很大时。因此,为提高系统容量,码分多址可以非常有效地利用交谈过程中的停顿[可以与时分多址系统中的非连续发送(DTX)进行比较]。
- 灵活多变的数据速率。在频分多址(或时分多址)系统中,用户或者占用一个频率(/时隙),或者占用整数倍数目频率(/时隙)。在码分多址系统中,可以通过扩频序列的适当选择来实现任意数据速率的传输。这一点对工作于固定数据速率的语音通信并不重要。然而,就数据传输而言,灵活的数据速率为更好地利用可用频谱创造了条件。

^① 注意,不同的小区也可能采用不同的码。这并不成为问题,而仅仅意味着(对下行链路而言)各 Rake 接收机分支上的不同相关器必须采用不同的扩频序列。

^② 实际上,大多数系统在此期间发送舒适的噪声(comfort noise),即某种背景噪声。如果在用户通话期间(即对方处于静默期期间)听不到任何声音,讲话的人会感到不适(以为通信连接已经中断了)。

- 软容量(soft capacity)。码分多址系统的容量可能因小区而异。如果一个给定小区增加了更多的用户,也就增加了对其他小区的干扰。因此就可能出现一些小区容量较高而另一些小区容量较低的状况;此外,这种状况可能会随着业务量的变化而动态改变。这一概念称为呼吸小区(breathing cells)。
- 纠错编码。引入纠错编码的缺点在于传输的数据速率会有所提高,从而降低了频谱效率。另一方面,码分多址有意识地提高了传送数据的数量^①,从而为引入差错编码而又不降低频谱效率提供了可能。换句话说,不同的用户以不同的差错控制码(通过扩频来实现编码)来区分。然而,值得注意的是,通用移动通信系统(UMTS,见第26章)之类的商用系统中不采用这一方式,而是分别进行差错控制和扩频。

18.3.4 与其他多址方式的结合

码分多址与时分多址和频分多址相比有许多优点,特别是在灵活性方面,然而它也存在一些缺点,例如复杂性。因此,为了获得“两全其美”的效果,将码分多址与其他多址方式结合起来的办法就是显而易见的了。最普及的解决方案就是码分多址与频分多址的结合,整个可用带宽被划分为多个子频带,在每个子频带上采用码分多址作为多址方式。很明显,这种系统的频率分集程度要比在整个可用带宽上进行频率扩展的情形低一些。有利的一面是发射机和接收机的处理速度可随码片速率的降低而有所下降。例如,IS-95(1.25 MHz宽的带宽)和占用5 MHz宽的子频带的通用移动通信系统(UMTS)都采用这一方式。

另一种结合方式是码分多址和时分多址相结合。可以像在时分多址系统中那样给每个用户分配一个时隙,而不同小区中的用户以不同的扩频码(而不是不同的频率)来区分。另一可能的实现形式是合并几个时隙,并用它们来构建一个窄带码分多址系统。在给现有的时分多址系统中添加了一个码分多址组件以后(如UMTS的时域双工模式),这种系统可以取得最佳的工作效果。

18.4 多用户检测

18.4.1 简介

多用户检测的基本思路

多用户检测基于干扰检测的思想,并利用检测所得的信息来减轻干扰对欲接收信号的负面影响。直到1986年,人们所形成的成见仍然是来自其他用户的干扰不可减轻,而且充其量干扰的表现能够类似于加性高斯白噪声就不错了。人们认为,在强干扰存在的情况下,正确的检测和解调是做不到的,就如同在强噪声环境下无法正确地检测信号一样。Poor和Verdu及其他人所开展的工作在Verdu[1998]中进行了总结,表明利用多用户干扰的结构来战胜干扰的负面影响实际上可以做到。如果采用这种策略,干扰就不会比高斯噪声更有害。人们在20世纪90年代对多用户检测进行了深入研究,在随后的10年里,采用这一方式的第一批实用系统得以进入市场。

^① 这里作者应该是指由直接序列扩频方式所决定的扩频后码片速率相对于信源数据速率的显著提高。——译者注

概念上最简单的多用户检测形式是串行干扰抵消,即逐个干扰抵消(SIC)。考虑一个这样的系统:该系统中存在一个(单个的)比欲接收信号更强的干扰信号。接收机首先检测和解调这一最强的信号。这一信号具有很好的信干噪比,因此就有希望将它无差错地检测出来。然后从总的接收信号中去除(抵消掉)其影响。这样,接收机就可以从“清理过的”信号中检测出希望接收的信号。由于这个“清理过的”信号中只含有所希望接收的信号和噪声,信干噪比会比较好,检测也就可以正确地完成。这个例子说明,采用多用户检测以后,在信干比低于 0 dB 的条件下实现有效检测是可能的。

其他的多用户结构包括最大似然序列估计检测器,这种方法试图同时实现对所有用户信号的最优检测。这些接收机呈现出非常好的性能,但由于运算强度随着被检测用户数的增加而呈指数增长,所以其复杂性也常常令人望而却步。最大似然序列估计的性能也可以用采用 Turbo 原理的接收机来逼近(见 14.5 节)。

一般将多用户检测器分为线性的和非线性的。前一类包括解相关接收机和最小均方误差接收机;后一类包括最大似然序列估计、干扰抵消、判决反馈接收机和 Turbo 接收机。

若干假设

针对这里将要进行的基本描述,为简化起见,先给出如下几个假设。

- 接收机具备了从干扰机到接收机的信道的完整信息。这显然是最好的状况。考虑串行干扰抵消接收机的情况:只有当接收机拥有了干扰信号的完整信道信息,它才能从总的信号中将其完全去除掉。干扰越强,任何信道估计误差在去除过程中的影响就越大。
- 所有用户都采用码分多址的多址方案。然而,要强调的是,对于其他多址方式(如时分多址),同样可以进行多用户检测,比如在时分多址网络中,可以用多用户检测来减小重用距离。后面也将看到,多用户检测和空分复用系统(见 20.2 节)的检测之间有着很强的相似性。
- 所有用户都是同步的。
- 这里只探讨接收机处的多用户检测。一个相关的主题是设计可以缓和不同接收机处干扰的发送方案。信息论的一个令人惊奇的结论是:如果发射机确知了信道和干扰,则有干扰信道的容量就与无干扰信道的容量完全相等。通过适当的编码策略可以达到这一容量,这种策略称为“在脏纸上书写”(writing on dirty paper)编码[Peel 2003]。可以利用这种编码来提升蜂窝系统的下行链路容量。

18.4.2 线性多用户检测

图 18.10 绘制了一个线性多用户系统的框图。该系统首先对来自不同用户的信号进行估计,实现估计的方式是用不同用户的扩频序列进行解扩处理^①。注意,这里需要许多并行的解扩器,其中每个解扩器都采用一个不同的扩频序列来实现解扩操作。之后,将这些解扩器的输出线性合并。这一合并步骤可以看成用矩阵滤波器来进行滤波,并可通过这种处理来消除干扰。这一方式与用于消除符号间干扰的线性均衡有很强的相似性。因此,第 16 章讨论过的迫零和维纳滤波等概念,在这里也会遇到。

^① 这一步骤可能要对来自多个 Rake 分支的信号进行合并,还要对来自天线阵列的多个单元的信号进行合并。

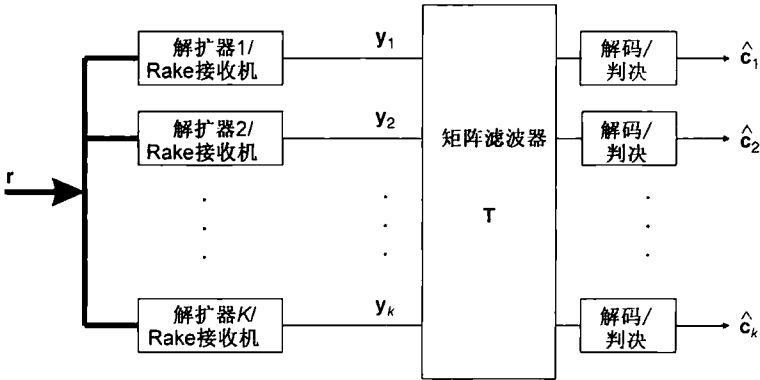


图 18.10 线性多用户检测器

解相关接收机

解相关接收机是最简单的多用户检测方法，等价于迫零均衡器。我们将接收滤波器(解扩以后)的输出写成

$$\mathbf{y} = \mathbf{R}\mathbf{c} + \mathbf{n} \tag{18.37}$$

其中，相关矩阵 $\mathbf{R}$ 包括了可能存在的天线分集和(/或)延迟分集支路； $\mathbf{n}$ 是噪声向量。对符号的估计可以通过用矩阵滤波器 $\mathbf{T} = \mathbf{R}^{-1}$ 滤波来简捷地获得：

$$\hat{\mathbf{c}} = \mathbf{R}^{-1}\mathbf{y} = \mathbf{c} + \mathbf{R}^{-1}\mathbf{n} \tag{18.38}$$

这一方法的优点在于其简单性，并且无须知晓接收信号的幅度，要确定的只是相关矩阵 $\mathbf{R}$ 。缺点在于噪声增强作用(也可以与迫零均衡器相类比)。相关矩阵调适的程度越差，噪声增强效果就越明显。

最小均方误差接收机

正如最小均方误差均衡器，最小均方误差多用户检测器在干扰抑制和噪声增强之间求取平衡。总扰动量的度量值是均方误差 $E\{|\mathbf{c} - \hat{\mathbf{c}}|^2\}$ 。因此，矩阵滤波器为

$$\mathbf{T} = [\mathbf{R} + \sigma_n^2\mathbf{I}]^{-1} \tag{18.39}$$

最小均方误差无法做到完全消除干扰，但由于相对更小的噪声增强作用，所引起的信号失真与解相关接收机相比还是要小一些。

18.4.3 非线性多用户检测器

线性多用户检测器忽略了发送信号的结构要素，允许发送信号的估计值取任何连续数值，因此就忽略了如下事实：发送信号仅可能包含有限发送符号集中的成员。非线性检测器则同时利用到了这一信息。

多用户最大似然序列估计

多用户最大似然序列估计器的结构与传统的最大似然序列估计器(常使用维特比检测器进行解码)相同。如果存在 K 个用户，则这些用户可能发送符号的组合数为 M^K 个。因此，网格图中的可能状态数也将随着用户数的增加而呈指数增长。由于这个原因，多用户最大似然序列估计器并未实用化。然而，它却给出了关于多用户检测性能限的至关重要的见解(更多细节参见 Verdu[1998])。

逐个干扰抵消

逐个干扰抵消(即串行干扰抵消)依照信号强度顺序检测用户。在检测下一个信号之前,将已检测出的每个用户的信号从总信号中减掉(见图 18.11)。因此,逐个干扰抵消是判决反馈接收机的特例。这种接收机按如下方式工作:接收到所有信号之和,并用每个用户的不同扩频码进行解扩。然后,最强的信号被检测出来并进行判决,从而得到不受噪声或干扰影响的原始比特流。随后,这个比特流被再次扩展,并从总信号中减去。“清理过”的信号再次通过解扩器,这一新信号中最强的用户信号被检测出来,再次扩展然后被减掉。重复这一过程直至最后一个用户被检测出来。值得注意的是,误差传播会严重影响逐个干扰抵消的性能:如果接收机误判了一个比特,则存在错误成分的信号将从总信号中减去,这样等待进一步处理的剩余信号将遭受更多而非更少的干扰。

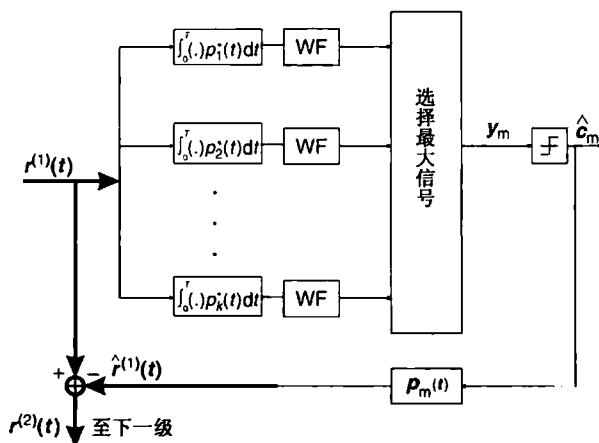


图 18.11 逐个干扰抵消接收机中的一级。来自 K 个用户的总接收信号 $\mathbf{r}$ 与第 k 个用户的扩频序列做相关;最大的信号被挑选出来,而其影响(即再次扩频信号)将从接收信号中减掉。WF 代表白化滤波器。图中没有显示信道延迟扩展的影响

干扰去除存在两种可能方式:“硬”去除和“软”去除。就硬去除而言,干扰被完全减掉;而软去除只减掉信号的一个按比例缩减版本。这样做无法完全消除干扰信号,却使误差传播问题的严重程度有所缓解。另一种抑制误差传播的方法是保证所得到的比特数据不仅是经过解调处理的,而且还经过了译码处理(即经历了纠错译码),然后在去除之前再次进行纠错编码、调制和扩频。因为译码后的错误概率比之前大大降低了,这就显著地减少了误差的传播。不利的一面是译码过程增加了判决过程的时延。

并行干扰抵消

如果不采用串行(逐个用户)方式去除干扰,也可以同时抵消掉所有用户。为达此目的,第一个步骤就是基于总的接收信号对所有用户进行(硬或软)判决。随后又将这些信号再次扩频,并从总信号中减去来自所有干扰用户的成分。注意,对每个用户而言,有着不同的干扰用户群:对于用户 1,干扰用户群是用户 2, ..., 用户 K ;对于用户 2,干扰用户群是用户 1 和用户 3, ..., 用户 K ;以此类推。接着,干扰抵消器中的下一级将采用这些“清理过”的信号作为判决的基础,并再次进行扩频和去除。重复这一过程直至前后两次迭代所得的判决结果不再变化,或者直至达到一定的迭代次数为止(见图 18.12)。

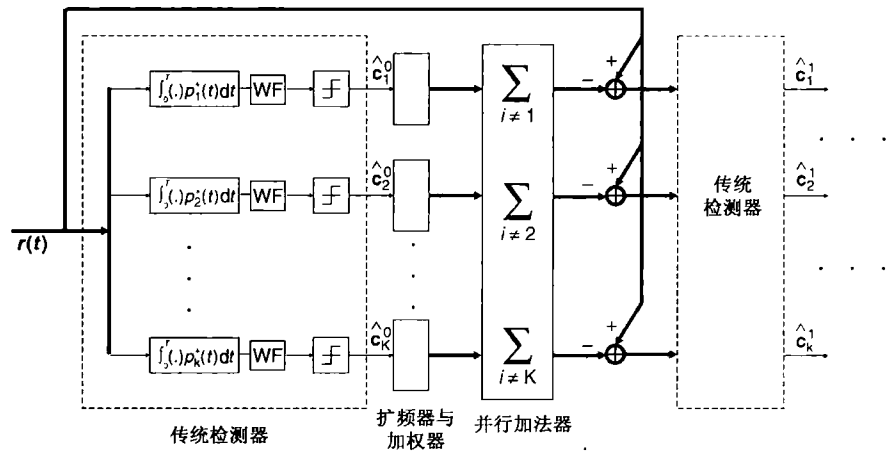


图 18.12 并行干扰抵消

误差传播对于并行干扰去除来说同样是一个重要问题。如果所有判决都正确无误，就只需要干扰抵消器中的第一级了。可是，第一级中每个信号的信噪比都比较好，所以只有最强的那个信号才具有较高的获得正确判决的概率。而且，错误的判决将引起更严重的信号失真，因此后续各级的表现会比第一级更差。

缓解这些问题的一项措施是将信号依功率等级归类，用于干扰抵消的只是那些能被可靠检测出来的信号。按照它们的信干噪比，众多到达信号被分成若干个类。对于 m 类中的某个判决，采用从 1 类到 m 类的反馈(即采用那些更可靠的判决结果)，而不用 $m + 1$ 类和 $m + 2$ 类等。

此外，还可以采用部分抵消的方式。在干扰抵消器的每一级，信号都将通过一个映射关系为 $x \rightarrow \tanh(\lambda x)$ 的映射器(以替代硬判决设备)，映射曲线的陡度(即 λ)将逐级递增。因此，只有最后一级进行了实际的硬判决。Turbo 多用户检测器针对不同比特来反馈对数似然比，从而进一步改善了这一原理。这种检测方式大大降低了误差传播的概率，可能发生错判的那些比特往往有着相当大的不确定性(噪声的强度不见得会大到足以让人确信发生了错判)。Turbo 检测器通过为每比特数据指定一个小的对数似然比来有效地利用这种不确定性。事实上，可以指出，Turbo 多用户检测器在性能上能够逼近于多用户最大似然序列估计检测器。

18.5 跳时脉冲无线电

当人们希望扩频带宽 W 能够达到近 500 MHz 或者更宽时，就开始对通过发送若干窄脉冲(用户所需的信息包含在其位置或幅度中)来实现扩频的方式产生了兴趣。通常把这种方式称为“脉冲无线电”(impulse radio)，利用简易的发射机和接收机就可以实现这种通信方式。大多数情况下，它被用在所谓的“超宽带”通信系统中[diBenedetto et al. 2005]。也可参考 21.8 节。

18.5.1 简易脉冲无线电

首先从最简单可行的脉冲无线电开始讨论。发射机每次发出的单个脉冲都代表着一个比特^①。暂且假定调制方式是正交脉冲位置调制(见第 11 章)。脉冲或者于时刻 t 开始发送，或者于时刻 $t + T_d$ 开始发送，这里 T_d 大于脉冲持续时间 T_c 。在加性高斯白噪声信道下，检测过程

① 原文为 symbol，即符号，译者认为表达为比特更容易使读者理解，且与下文中用词是一致的。——译者注

极其简单：只需用一个能量检测器来进行能量检测，以确定在两个可能的时间间隔

$$[t + \tau_{\text{run}}, t + \tau_{\text{run}} + T_C] \quad (18.40)$$

或

$$[t + \tau_{\text{run}} + T_d, t + \tau_{\text{run}} + T_C + T_d] \quad (18.41)$$

中，哪一个的持续时间里能得到更多的能量。此处， τ_{run} 是发射机和接收机之间的电波传播时间，通过同步过程来确定它。

显然，这种脉冲化的传输可以实现扩频，这是因为发送信号带宽由脉冲间隔的倒数 $1/T_C$ 给定。而“解扩”则是以非常简单的方式来的：只要在由式(18.40)和式(18.41)给定的间隔中，对到达信号进行记录和处理，就能判定哪一个比特被发送了。接收机的输入信噪比表示为 E_B/N_0 ：如果接收机是能量检测器，则它所检测到的峰值信号功率等于 $\bar{P}T_S/T_C$ ，而输出的噪声功率等于 N_0/T_S ，结果输出信噪比为 $(\bar{P}T_S/N_0)(T_S/T_C) = (E_B/N_0)(T_S/T_C)$ 。这表明(宽带)噪声抑制因子为 T_S/T_C 。

但是，这种简易方案存在如下两个重大缺陷。

1. 发送信号的峰平比(峰值因数)很高，造成发送和接收电路设计方面的种种问题。
2. 这种方案对不同类型的干扰都显得不够强健。特别地，在有多址干扰的情况下，此方案存在着诸多问题。换句话说，假定目标发射机发出了一个 0，0 所对应的脉冲发出时刻为 t ，该脉冲到达接收机的时刻为 $t + \tau_{\text{run, desired}}$ ，其中 $\tau_{\text{run, desired}}$ 是目标发射机与接收机之间的信号传播时间；而另一个用户也发出了一个 0，对应的脉冲于 $t + \tau_{\text{run, interference}}$ 时刻到达同一接收机，其中 $\tau_{\text{run, interference}}$ 是干扰用户的发射机与接收机之间的信号传播时间。那么将会发生什么呢？可能这一干扰脉冲正好在如下时刻到达接收机：如果目标发送用户发出一个 1，这一时刻本该是相应的代表 1 的脉冲(脉冲发出时刻为 $t + T_d$) 在接收机处出现的时刻。更确切地说，这一情况发生在 $t + \tau_{\text{run, interference}} = t + \tau_{\text{run, desired}} + T_d$ 的时候。因为很难以非常精准的方式对不同用户信号的传播时间施加影响，上述情况总是会以一定的概率发生的。我们用“灾难性碰撞”的术语来命名它，因为接收机无法对接收到的符号进行恰当的判决，并且错误概率将会非常高；也可以用它与跳频系统(见 18.1 节)中的灾难性碰撞类比。

为解决这些问题，我们应该就每个符号发出多个脉冲。首先能想到的就是发送一个规则脉冲序列，脉冲之间的间隔为 T_f 。这就解决了峰平比的问题。然而，很容易就能看出它并没有降低灾难性碰撞的概率：如果脉冲序列中的第一个脉冲与一个干扰脉冲序列中的脉冲发生了碰撞，则后续脉冲也会碰撞。此问题的一种解决方案将通过跳时脉冲无线电的概念提供出来，下文会加以描述。

18.5.2 跳时脉冲无线电

跳时脉冲无线电(TH-IR)的基本思想是采用间隔不规则的脉冲序列来代表单个符号。更确切地说，将可用的符号时间划分成一定数目的持续时间为 T_f 的“帧”，并在每个帧的时间范围内发送一个脉冲(见图 18.13)^①。这样，其核心思想就是在帧的范围内改变脉冲出现的位置。

① 注意，在脉冲无线电文献里所采用的名词“帧”(frame)可能会引起混淆。在脉冲无线电中，一个数据符号包含几个帧。对于时分多址或分组预约多址(PRMA)，一个数据块通常也称为一个“帧”；当用于这种场合时，一个帧包含几个符号。

置。例如,在第一个帧中,在该帧的第三个片段(chip)发送脉冲;在第二个帧中,在该帧的第八个片段发送脉冲,等等。各个脉冲在帧里出现的位置由称为“跳时序列”的伪随机序列来决定。发送信号的数学表达式为

$$s(t) = \frac{\sqrt{E_s}}{\sqrt{N_f}} \sum_i \sum_{j=1}^{N_f} g(t - jT_f - c_j T_c - b_i T_d - iT_s) = \frac{\sqrt{E_s}}{\sqrt{N_f}} \sum_i p(t - b_i T_d - iT_s) \quad (18.42)$$

其中, $g(t)$ 是持续期为 T_c 的单位能量发送脉冲, N_f 是用来代表一个信息符号(时长为 T_s)的帧的个数(因此也是脉冲的个数),而 b 是所发送的信息符号。跳时序列向信号的第 j 个脉冲提供了 $c_j T_c$ 秒的额外的时移,其中 c_j 是伪随机序列的元素,取 0 至 $N_f - 1$ 之间的整数值(见图 18.13)。接收机中,进行匹配滤波(与整个发送波形 $p(t)$ 相匹配)以后,在时刻 $t = T_s + \tau_{\text{run}}$ 和 $t = T_s + \tau_{\text{run}} + T_d$ 对匹配滤波器的输出进行取样。通过对这两个时刻的取样值进行比较来判定已发送的是哪个符号。如论述码分多址接收时提到过的,匹配滤波器操作也可以解释成与发送序列做相关。

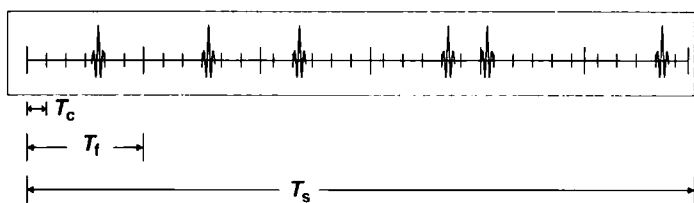


图 18.13 跳时脉冲无线电中,一个符号的发送波形 $p(t)$, 图中显示了符号间隔 (T_s)、帧间隔 (T_f) 和帧内片段间隔 (T_c)

在这种脉冲无线电 (TH-IR) 中,可以得到与“简易”脉冲无线电一样的噪声抑制力,即 T_s/T_c ; 但此时这一增益源于两方面,其中一部分源于仅在较短的时间段观测噪声,即这部分增益与简易脉冲无线电中所得的增益属于同一类型,它的值现在是 T_f/T_c , 因此比简易脉冲无线电中的值更小。另一种类型的增益来自于不同帧中脉冲的合并,不同帧中的信号成分相干地累加起来,而噪声成分则不相干地叠加在一起。这两部分增益合起来就得到了总增益 T_s/T_c 。

如此一来,相比于规则脉冲串,这种方式的好处何在呢? 不同用户可以使用不同的跳时序列。如后面将要指出的,这些序列可以按如下方式来构造: 不同用户的脉冲碰撞只发生在少数几个帧上(假定发生碰撞的帧数为 N_{collide}),而在其他帧上都不发生碰撞。当接收机用希望接收的用户的跳时序列与接收信号做相关时,只有在那些会真正发生碰撞的帧上才出现干扰。这就获得了抑制因子为 N_f/N_{collide} 的干扰抑制能力。

前文曾经指出,所希望接收的用户和施加干扰的用户之间可能是不同步的,从而彼此之间可能有任意的时移。所以,应以如下准则来构造跳时序列: 不管不同序列之间的相对平移是多少,脉冲之间的碰撞数不应超过一个门限值 λ (通常选 $\lambda = 1$)^①。这样,系统设计者就不必担心传播时间的影响或用户之间的同步状况; 对多址干扰 (MAI) 的良好抑制总能得到保证。设计这种序列是比较困难的,尤其是在想得到具有很少碰撞的大量序列时; 通常,获得这些序列的最佳办法就是穷尽型计算机搜索。

① 有趣的是,好跳时序列的构造问题与跳频扩频系统中好跳频序列的构造问题具有很强的相似性(见 18.1 节)。

例 18.3 某跳时脉冲无线电系统的帧数 $N_f=6$ ，其中采用的一个跳时序列为 $[1, 2, 4, 6, 3, 5]$ 。找出在任意平移情况下至多只与上述序列发生一次碰撞的另一跳时序列。

解：

经过一番全面有序的搜索，我们找到了满足上述要求的序列： $[1, 3, 5, 2, 6, 4]$ ；从图 18.14 可以看出它满足题设的要求。

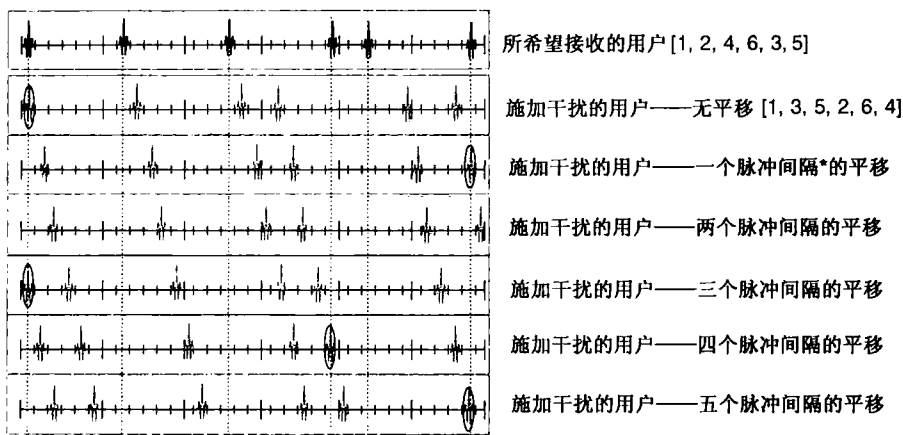


图 18.14 两用户之间的干扰，图中给出了两用户之间所有可能的（整数）平移情况。周围带圈的脉冲表示有碰撞发生^①

对于脉冲无线电，还可以给出一种非常有用的不同解释：可以将它视为一种直接序列码分多址系统（正如 18.2 节看到的那种），只不过其扩频序列中有大量的 0，而只有少量的 1。可以拿它与几乎具有相等数目的 +1 和 -1 的传统直接序列码分多址系统进行比较。这一解释的引人之处在于，人们在传统直接序列码分多址系统方面已经进行过比脉冲无线电多得多的研究，通过将跳时脉冲无线电解释为一种直接序列码分多址系统，许多直接序列码分多址方面的结论就能直接应用于脉冲无线电。

最后要指出，跳时脉冲无线电不仅可以与脉冲位置调制结合起来使用，还可以与脉冲幅度调制结合起来。另外，可以将一个符号内的发送脉冲极性随机化，从而可将发送信号表示为

$$s(t) = \frac{\sqrt{E_s}}{\sqrt{N_f}} \sum_i b_i \sum_{j=-\infty}^{\infty} d_{ij} g(t - jT_f - c_j T_c - iT_s) \quad (18.43)$$

其中，每个脉冲与一个伪随机变量 d_j 相乘， d_j 等概率地取 +1 或 -1。这种极性随机化对发送信号的频谱成形有好处。就这种系统而言，与直接序列码分多址之间的相似性将会更引人注目：现在扩频序列是三值序列，具有（几乎）相等数目的 +1 和 -1，而在其间有大量的 0。于是，干扰的抑制力不仅源于较少数目的脉冲碰撞，而且还可以归因于来自不同帧（但取相同的符号）的干扰成分可能相互抵消。

18.5.3 时间色散信道下的脉冲无线电

至此已讨论了加性高斯白噪声信道下的脉冲无线电。已知这种情形下可以以非常简单的

^① 原文为 chip，涉及码分多址时通常译为“码片”；涉及脉冲无线电时，译为码片似不太合适，故前面将其译为“帧内片段”或者“片段”，实际上也就是在一个帧里，脉冲可能出现的一小段时间，一个帧包含 N_c 个 chip，所以又译为“脉冲间隔”。——译者注

形式构造发射机和接收机。然而,跳时脉冲无线电的使用环境永远也不会是加性高斯白噪声信道。这种系统的出发点是采用非常宽的带宽(通常是 500 MHz 或更宽),这也就意味着在这样宽的带宽上信道必将有所变化。于是,构造可以在时间色散信道下很好地发挥作用的接收机就显得非常必要了。

首先来考虑到达信号的相干接收。这种情况下需要一个 Rake 接收机,正如 18.2.4 节所讨论到的那种。本质上, Rake 接收机包括多个相关器或者分支。在每一分支中,到达信号与跳时序列的一个时延版本一起完成相关运算,其中具体的时延值等于想要“分离出来”的那个多径分量的时延。然后,各 Rake 分支的输出被加权(依照最大比值合并或最优合并原理)后再相加。由于脉冲无线电系统通常占用非常宽的带宽,它们总是采用分支数少于可用多径分量数目的结构(SRake 和 PRake, 见 18.2.4 节)。为获得满意的性能,即使是在室内传播环境中, Rake 接收机可能也需要 20 个或更多个 Rake 分支。因此,差分相干(发送参考)或非相干方案就成为了吸引人的替代方案。

要理解发送参考(Transmitted Reference, TR)方案的原理,应当记住,时延色散信道中,理想匹配滤波器的匹配对象是跳时序列与信道冲激响应的卷积结果。就相干接收而言,首先让一个滤波器与跳时序列相匹配,后面再跟一个 Rake 接收机,该接收机与信道的冲激响应匹配。所谓的发送参考方案是以不同的方式来产生这一复合匹配滤波器。一个发送参考发射机每帧发出两个脉冲:一个未调制的(参考)脉冲和一个固定间隔 T_{pd} 之后发送的已调制(数据)脉冲。当通过无线信道发送时,参考脉冲序列将与信道冲激响应做卷积,因此这一信号(卷积结果)就成为了系统冲激响应(发送基本波形与信道冲激响应的卷积)的一个有噪声版本。为实现接收信号的数据载荷部分的匹配滤波,接收机只要将接收信号与收到的参考信号相乘就行了。

现在来看发送信号中的一个符号的数学表达式:

$$p(t) = \sqrt{\frac{1}{2}} \sum_{j=0}^{N_f} d_j [g(t - jT_f - c_j T_C) + b \cdot g(t - jT_f - c_j T_C - T_{pd})] \quad (18.44)$$

仔细观察上式就会发现,与收到的参考信号之间的相关运算可以通过以下处理来实现:只要将总接收信号与其自身的延迟(延迟时间为 T_{pd})版本相乘并积分,就可以得到相关结果。更确切地说,接收机处的第一步是用接收滤波器 $h_R(t)$ 进行滤波,这一滤波器应该宽到不至于引入任何信号失真,而只是尽可能地抑制噪声 $\hat{r}(t)$ 。滤波后的接收信号与自己的延迟版本相乘,并在有限间隔 T_{in} 上积分,积分间隔应该长到足以搜集尽量多的多径能量,而又短到不至于搜集太多的噪声能量。

发送参考方案的优点就在于极其简单(尽管在接收机中实现延迟并非易事)。不利的一面是,它的性能要比相干方案差,原因有两个:(i) 在其不承载信息的意义上,参考脉冲浪费了能量(这导致 3 dB 的损失);(ii) 与信号的数据承载部分一样,信号的参考部分也会受到噪声影响。接收机中的这两个信号相乘会产生噪声与噪声的交叉项,从而劣化了信噪比。由于脉冲无线电是扩展频谱系统,接收信号的信噪比是负值,于是噪声与噪声的交叉项可能变得很大。

最后要指出,非相干检测也可以成为一种具吸引力的替代方案。仅服务于单个用户时尤其如此。可是,非相干接收机在处理多址干扰方面存在诸多问题。由于多径传播,能量被扩散到几个相连的帧内片段上。因此,能量检测将会遇到更多的干扰(见图 18.15)。

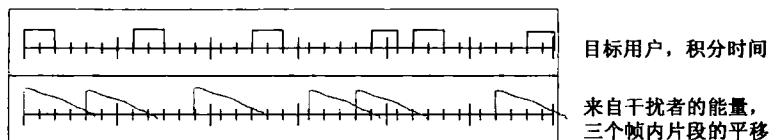


图 18.15 时间色散干扰信号(下面一行)对目标信号的干扰,目标信号的积分时间(每帧的两个帧内片段时间)在上面一行画出

深入阅读

扩展频谱通信的一般性概述可以在经典专著 Dixon[1994]和 Simon et al.[1994]中找到,它们很好地涵盖了跳频和直接序列系统。更专门地针对蜂窝系统特别是码分多址的著作是 Glisic and Vucetic[1997], Goiser[1998], Li and Miller[1998], Viterbi[1995]和 Ziemer et al.[1995];概述性论文 Kohno et al.[1995]给出了更简短的描述。Scholtz[1982]给出了扩展频谱历史的概述(尽管更多是从一种军事应用视角来展开,这是因为文章发表的时期还没有考虑蜂窝的应用)。Milstein[1988]讨论了不同扩频技术的干扰抑制能力。

Maric and Titlebaum[1992]中设计了跳频码。对码分多址系统容量的估计最早是在被广泛引用的论文 Gilhousen et al.[1991]中发表的。多径对码分多址的影响,即 Rake 接收机的应用在 Goiser et al.[2000]和 Swarts et al.[1998]中有所评论;特别地,Win and Chrisikos[2000]中论及有限数目 Rake 分支的影响。其他关于 Rake 接收的重要论文包括 Holtzman and Jalloul[1994]和 Bottomley et al.[2000]等。就同步方面来说,论文 Polydoros and Weber[1984]仍然值得一读。Dinan and Jabbari[1998]中评论了扩频码。针对雷达和通信中序列设计的权威性阐述由 Golomb and Gong[2005]给出。虽然高斯近似被广泛用于小区之间干扰的分析,但当存在阴影时这种近似可能会造成背离实际情况的相当大的偏差, Singh et al.[2010]中给出了一个更为精确的模型。Molisch[2004]中描述了一个完整的跳时脉冲无线电系统。

就多用户检测而言,Verdu[1998]著作是必读的。概述文章 Duel-Hallen et al.[1995], Moshavi[1996] and Poor[2001]都给出了不少有价值的观点。盲多用户检测这样的高级概念在 Wang and Poor[2003]和 Honig[2009]中有深入描述;Poor[2004]探讨了多用户检测和解码的相互作用。

跳时脉冲无线电在 20 世纪 90 年代由 Win and Scholtz 进行了开拓性的研究。基本原理的描述可以参考 Win and Scholtz[1998],而更多细节则可以在 Win 和 Scholtz[2000]中找到。超宽带系统设计的不同侧面的概述可以在专著 Shen et al.[2006], Ghavami et al.[2006], Reed[2005], diBenedetto et al.[2005]和 Roy et al.[2004]中找到;发送参考接收机和其他简化接收机结构在概述性文章 Witrisal et al.[2009]中有所讨论。

第 19 章 正交频分复用

19.1 引言

正交频分复用(OFDM)是特别适合在时延色散环境下进行高数据速率传输的一种调制方案。该方案将一个高速数据流转换成若干个低速数据流,并采用一组并行的易于均衡的窄带信道进行传输。

首先来分析在甚高速率数据传输中,传统的调制方法会产生的问题。当所需的数据速率提高时,为了获得该数据速率,符号周期 T_s 必须非常小,此时系统带宽变得非常大^①。无线信道自然产生时延色散,其数值取决于环境,而不是传输系统。因此,如果符号周期很小,那么冲激响应(和均衡器所需的长度)就符号周期而言,会变得比较长。这样一个长均衡器的计算量非常大(见第 16 章),并且会增加系统的不稳定概率。例如, GSM 系统(见第 24 章)所设计的数据速率达 200 kbps,所用带宽为 200 kHz,而 IEEE 802.11 系统(见第 29 章)的数据速率达 55 Mbps,所用带宽为 20 MHz。在一个最大接入时延为 $1\text{ }\mu\text{s}$ 的信道中,前者需要一个 2 抽头的均衡器,而后者则需要 20 个抽头。正交频分复用则从另一个角度,与单载波系统相比增加了各个载波上的符号周期,使每个子载波具有更简单的均衡器。

正交频分复用技术可以追溯到 40 多年前,这项技术在 20 世纪 60 年代中期被申请为专利 [Chang, 1966]。几年后,引入了一种非常重要的改进循环前缀,它有助于消除剩余时延色散。Cimini[1985]首先提出将正交频分复用技术应用于无线通信。但是直到 20 世纪 90 年代初期,数字信号处理硬件的发展才使正交频分复用技术成为无线系统中可选用的实现方案,此外非常适于采用正交频分复用的高数据速率的应用近几年才出现。现在,正交频分复用技术已被应用于数字音频广播(DAB)、数字视频广播(DVB)和无线局域网(IEEE 802.11a 和 IEEE 802.11g)。它也将用于第四代蜂窝系统,包括第三代合作伙伴计划长期演进(3GPP)和 WiMAX。

19.2 正交频分复用的原理

正交频分复用将高速数据流分为 N 个并行的数据流,这些数据流调制到 N 个不同的载波(下文称为子载波或单音)上,然后发射出去。每个子载波的符号周期增大为原来的 N 倍。为了能让接收机分离不同子载波携带的信号,子载波必须是正交的。如同 17.1 节所介绍的及图 19.1 所示,传统的频分多址通过在载波之间设置较大的(频率)间隔来达到信号分离的目的,但是这样做会浪费宝贵的频谱资源。子载波间隔可以设得更窄。例如,令子载波的频率 $f_n = nW/N$,其中 n 为整数, W 为总的可用带宽;最简单的情况为 $W = N/T_s$ 。进一步假设每个子载波上采用矩形基带脉冲的脉冲幅度调制。由关系式

^① 可通过多址接入模式复合,如时分多址接入,它具有较高峰值速率的原因是将数据压缩成脉冲串(见第 17 章)。

$$\int_{iT_s}^{(i+1)T_s} \exp(j2\pi f_k t) \exp(-j2\pi f_n t) dt = \delta_{nk} \quad (19.1)$$

可以很容易地发现子载波间是相互正交的。

图 19.1 给出了频域的原理。因为时域为矩形脉冲，所以每个已调载波在频域具有 $\sin(x)/x$ 形状。不同的已调载波谱相互重叠，但是每个载频都处在所有其他载波的零点处。因此只要接收机进行合适的解调(乘以 $\exp(-j2\pi f_n t)$ ，并在符号周期内积分)，那么任何两个子载波的数据流都不会发生串扰。

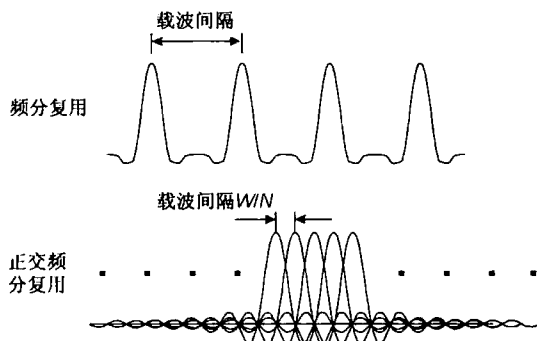


图 19.1 正交频分复用的原理: N 个载波, 带宽为 W

19.3 收发信机实现

正交频分复用可以用两种方式来描述: 一种为“模拟”形式, 如图 19.2(a) 所示。正如 19.2 节所讨论的, 先把原始的数据流分为 N 个并行的数据流, 每个数据流具有相对较低的数据速率。同时使用多个本地振荡器, 每个振荡器的振荡频率为 $f_n = nW/N$, 其中 $n = 0, 1, \dots, N-1$ 。然后每个并行的数据流调制到其中一个载波上。图 19.2(a) 使其原理易于理解, 但是这种方式不适用于实际实现, 因为多个本地振荡器的硬件成本太高。

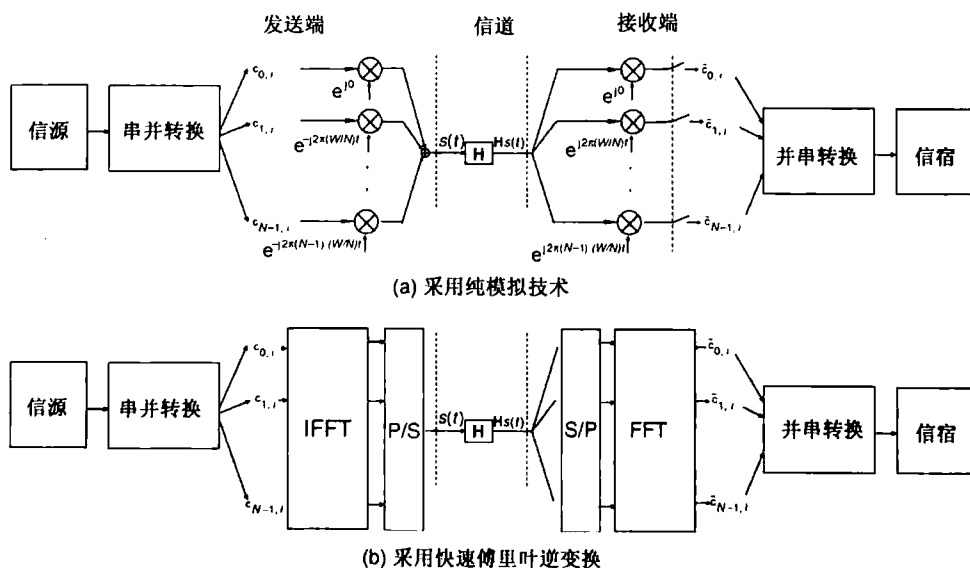


图 19.2 正交频分复用收发信机结构

另一种可选用的实现方案是数字的,它将传输数据分成 N 个符号长的数据块,对这个数据块进行快速傅里叶逆变换(IFFT),然后再传输,如图 19.2(b) 所示。这种方法用集成电路实现起来容易得多。下面将证明这两种方法是等价的。

首先考虑模拟形式。令第 n 个载波在第 i 个时刻的复发送符号为 $c_{n,i}$, 则发送信号为

$$s(t) = \sum_{i=-\infty}^{\infty} s_i(t) = \sum_{i=-\infty}^{\infty} \sum_{n=0}^{N-1} c_{n,i} g_n(t - iT_S) \quad (19.2)$$

其中,基带脉冲 $g_n(t)$ 为归一化的频移矩形脉冲:

$$g_n(t) = \begin{cases} \frac{1}{\sqrt{T_S}} \exp\left(j2\pi n \frac{t}{T_S}\right), & 0 < t < T_S \\ 0, & \text{其他} \end{cases} \quad (19.3)$$

不失一般性,现在仅考虑 $i=0$ 时信号的情况,在 $t_k = kT_S/N$ 时刻对其采样:

$$s_k = s(t_k) = \frac{1}{\sqrt{T_S}} \sum_{n=0}^{N-1} c_{n,0} \exp\left(j2\pi n \frac{k}{N}\right) \quad (19.4)$$

可以发现,上式恰好为发送符号的离散傅里叶逆变换。因此,发射机可以通过对发送符号块进行一次离散傅里叶逆转换来实现(块的大小必须与子载波个数相同)。几乎在所有的实际情况中,样本数 N 都选为 2 的幂次,而离散傅里叶逆变换用快速傅里叶逆变换来实现。下文中只讨论快速傅里叶逆变换和快速傅里叶变换。

注意,快速傅里叶逆变换的输入由 N 个样本值组成(不同子载波的符号),因此其输出也有 N 个数值,这 N 个数值必须按采样的时间顺序逐个发射出去,这就是在快速傅里叶逆变换后直接进行 P/S(并串)转换的原因。在接收机中,可以进行相反过程的处理:对我们接收信号进行采样,将 N 个采样数据组成一个向量,也就是一次 S/P(串并)转换,然后对这一向量进行快速傅里叶变换。变换的结果就是原始数据 c_n 的估计值 $\hat{c}_n$ 。

正交频分复用的模拟实现需要多个本地振荡器,而为了保持不同子载波之间的正交性,每个振荡器的相位噪声和漂移必须很小,所以这通常并不是一种实用的解决方案。正交频分复用的成功应用基于上面描述的数字实现,这使得收发信机更简单和廉价,尤其是在数字实现中,存在高效的快速傅里叶变换运算结构(所谓的“蝶形结构”),并且进行快速傅里叶逆变换操作的运算量(每比特)仅随 $\log(N)$ 增加。

正交频分复用还可以在时频空间内进行说明。每个下标 i 对应于一个(时域)脉冲;每个下标 n 对应于一个载波频率。全体函数集分布于时频空间的一个网格内。

19.4 频率选择性信道

前一节解释了正交频分复用发射机和接收机在加性高斯白噪声信道中的工作原理。现在对这种方案不做任何改动,就可让它工作在频率选择性信道中。从直觉上,预计时延色散对正交频分复用的性能仅会产生微小的影响:把系统转化为由一些窄带信道组成的并行系统,因此每个载波上的符号周期比时延扩展大得多。但是,正如第 12 章所介绍的,即使 $S_c/T_s < 1$, 时延色散也会导致相当大的误差。此外,下面将要详细讨论到,时延色散还会使子载波之间的正交性丧失,从而引起载波间干扰(ICI)。幸运的是,这些负面的影响都可以用一种特殊的保护间隔来消除,这种保护间隔称为循环前缀(CP)。这一节将说明如何构造循环前缀,循环前缀如何起作用,以及在频率选择信道中能获得怎样的性能。

19.4.1 循环前缀

首先为传输定义一个新的基函数：

$$g_n(t) = \exp \left[j 2 \pi n \frac{W}{N} t \right], \quad -T_{cp} < t < \hat{T}_s \quad (19.5)$$

其中, W/N 仍然为载波的频率间隔, 且 $\hat{T}_s = N/W$ 。符号周期 T_s 在这里为 $T_s = \hat{T}_s + T_{cp}$ 。这个基函数的定义意味着当 $0 < t < \hat{T}_s$ 时, 发送“常规”的正交频分复用符号(见图 19.3)。将 $g_n(t) = g_n(t + N/W)$ 代入式(19.5)很容易看出。因此, 当 $-T_{cp} < t < 0$ 时, 发送符号最后部分的一份副本。从线性角度来看, 它按照总发射信号 $s(t)$ 在时间 $-T_{cp} < t < 0$ 内的信号为在时间 $\hat{T}_s - T_{cp} < t < \hat{T}_s$ 内 $s(t)$ 信号的副本。信号的这个预置部分称为“循环前缀”。

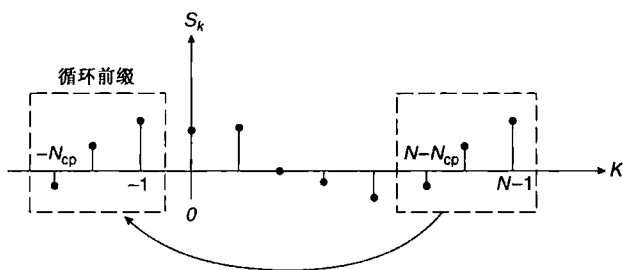


图 19.3 循环前缀原理。 $N_{cp} = NT_{cp}/(N/W)$ 是循环前缀中的样点数

现在我们知道了什么是循环前缀, 再分析一下为什么它对时延色散信道是有好处的。当任何数据流通过一个时延色散信道传输时, 接收信号等于发送信号与信道的冲激响应的线性卷积。循环前缀将线性卷积转化为循环卷积。当 $-T_{cp} < t < -T_{cp} + \tau_{\max}$ (其中 $\tau_{\max}$ 为信道的最大附加时延) 时, 当前符号和它前一符号后面部分的拖尾混叠, 使接收信号受到“实的”符号间干扰^①。通过丢弃这段时间内接收到的信号, 这种“规则”的符号间干扰就能被消除。在剩下的符号周期内, 还存在着循环符号间干扰, 特别是当前(而不是前一个)符号的后一部分干扰了当前符号的前一部分。下面将说明如何用一种极其简单的数学运算来消除这种循环卷积的影响。

在下面的数学推导中, 假定冲激响应的持续时间严格等于循环前缀的长度; 此外为了简化算式, 假设(不失一般性) $i=0$ 。在接收机中, 有一组滤波器与不带循环前缀的基函数相匹配:

$$\bar{g}_n(t) = \begin{cases} g_n^*(\hat{T}_s - t), & 0 < t < \hat{T}_s \\ 0, & \text{其他} \end{cases} \quad (19.6)$$

这一操作从检测过程中去除了接收信号的第一部分 (T_{cp} 时段); 正如前面所讨论的, 符号剩余部分的匹配滤波可以通过快速傅里叶变换操作实现。因此匹配滤波器输出端的信号为发送信号与信道和接收滤波器冲击响应的卷积:

$$r_{n,0} = \int_0^{\hat{T}_s} \left[\int_0^{T_{cp}} h(t, \tau) \left(\sum_{k=0}^{N-1} c_{k,0} q_k(t - \tau) \right) d\tau \right] g_n^*(t) dt + n_n \quad (19.7)$$

其中, n_n 为匹配滤波器输出端的噪声。注意, g_k 可以从 $-T_{cp}$ 到 $\hat{T}_s$ 取值, 这一取值范围为表达式(19.5)的定义区间。如果信道在 T_s 期间可以认为是不变的, 那么 $h(t, \tau) = h(\tau)$, 于是可得

① 在下面的内容中, 假设 $\tau_{\max} \leq T_{cp}$ 。

$$r_{n,0} = \sum_{k=0}^{N-1} c_{k,0} \int_0^{\hat{T}_s} \left[\int_0^{T_{cp}} h(\tau) (g_k(t - \tau)) d\tau \right] g_n^*(t) dt + n_n \quad (19.8)$$

内积分可以写为

$$\exp \left[j2\pi tk \frac{W}{N} \right] \int_0^{T_{cp}} h(\tau) \exp \left(-j2\pi \tau k \frac{W}{N} \right) d\tau = g_k(t) H \left(k \frac{W}{N} \right) \quad (19.9)$$

其中, $H\left(k \frac{W}{N}\right)$ 是在频率 kW/N 处的信道传递函数。另外, 基函数 $g_n(t)$ 在 $0 < t < \hat{T}_s$ 期间是正交的:

$$\int_0^{\hat{T}_s} g_k(t) g_n^*(t) dt = \delta_{kn}(t) \quad (19.10)$$

接收信号采样值 r 可写为

$$r_{n,0} = H \left(n \frac{W}{N} \right) c_{n,0} + n_n \quad (19.11)$$

因此, 正交频分复用系统可以由一组并行的非色散平坦衰落信道组成, 每个信道的复衰减为 $H\left(n \frac{W}{N}\right)$ 。系统的均衡因此变得极其简单: 只需要除以子载波频率处的传输函数, 且各个子载波是独立的。换句话说, 循环前缀恢复了子载波的正交性。

要注意两点警示: (i) 在推导中假设在正交频分复用信号持续期间, 信道是静态的。如果不能满足这个假设条件, 各个子载波信号仍有可能发生干扰(见 19.7 节); (ii) 接收信号的丢弃部分不但会降低频带利用率, 而且会降低信噪比。对于通常的工作参数(循环前缀约为符号周期的 10%), 这种损失是可接受的。

包含循环前缀的正交频分复用系统框图如图 19.4 所示。原始数据流进行串并转换。每个数据块由 N 个数据符号组成, 对其进行快速傅里叶逆变换, 接下来最后的 NT_{cp}/T_s 个样点被复制后前置。这样得到的信号调制到(单)载波上再发射出去。在信道传输过程中, 信号产生畸变, 并引入噪声。在接收机中, 信号被划分为数据块。对于每个数据块, 去掉循环前缀, 剩余部分再进行快速傅里叶变换。在每个载波上, 得到的样点(可以认为是频域样点)采用单抽头均衡器进行“均衡”, 即除以复信道衰减系数。

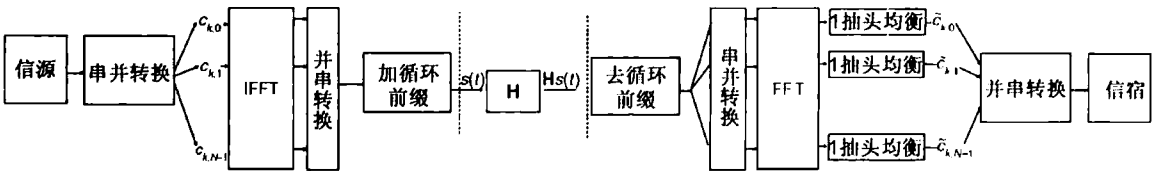


图 19.4 带有循环前缀和单抽头均衡器的正交频分复用传输链路的结构

19.4.2 频率选择性信道中的性能

循环前缀将频率选择性信道转化为一组并行的平坦衰落信道。从它消除了困扰时分多址和码分多址系统的符号间干扰这一点来讲, 它是有利的。不利的一面是未编码正交频分复用系统根本显示不出任何频率分集。如果子载波处于深衰落点, 那么该子载波上的错误概率将会非常高, 并且即使在高信噪比的条件下仍然决定了整个系统的误比特率。

例 19.1 未编码正交频分复用的误比特率。

解:

图 19.5 给出了一个具体频率选择性信道的传输函数和二进制相移键控正交频分复用系统在其中传输的误比特率。很明显,在深衰落点的误比特率最大。注意结果是用对数坐标表示的,虽然一些“好的”子载波上的误比特率会低到 10^{-4} ,但一些处于深衰落点的子载波上的误比特率却会高达 0.5。这对平均差错率也会产生明显的影响,因为坏子载波上的误比特率决定了总的误比特率。图 19.6 给出了频率选择性信道的平均误比特率的仿真(在许多信道实现上)。可以看出,随着信噪比的增加,误比特率只是随着信噪比增大呈线性减小;进一步的观察显示结果与图 12.6 中的情况一致。

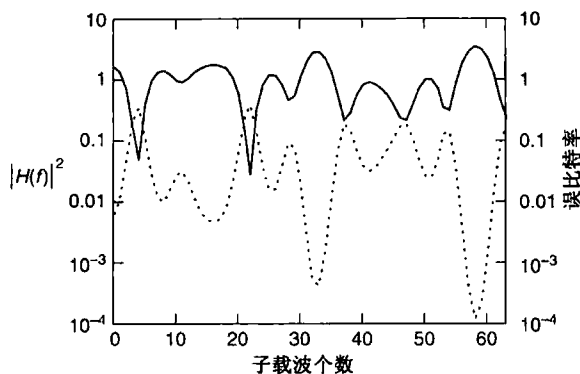


图 19.5 传输函数的归一化幅度的平方(实线)和误比特率(虚线)。信道抽头系数为 $[0, 0.89, 1.35, 2.41, 3.1]$,幅度为 $[1, -0.4, 0.3, 0.43, 0.2]$ 。接收机的平均信噪比为 3 dB,调制方式为二进制相移键控,子载波频率 $f_k = 0.05k$, $k = 0, \dots, 63$

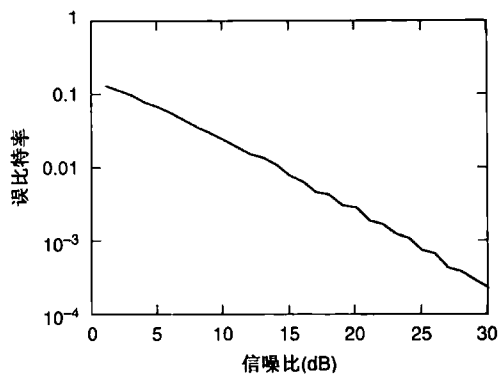


图 19.6 抽头系数为 $[0, 0.89, 1.35, 2.41, 3.1]$,平均功率为 $[1, 0.16, 0.09, 0.185, 0.04]$ 的信道误比特率。每个抽头相互独立,服从瑞利衰落。调制方式为二进制相移键控,子载波频率 $f_k = 0.05k$, $k = 0, \dots, 63$

更一般地,可以看出,未编码的正交频分复用具有相同的平均误比特率,而与信道的不同频率选择性无关。也可以这么理解:频率选择性在不同的子载波上给出了不同的信道实现;时变性在不同时间上给出了不同的信道实现。因而双选择性信道在不同时间和不同频率具有不同的实现形式。但是,只要衰落具有相同的分布(如瑞利分布),且子载波数足够大^①,那么如何产生不同的信道实现,对于平均误比特率的计算没有任何影响。

从这些例子可以看出,主要问题在于低信噪比的载波决定了系统的性能。下面的方法都可以解决这个问题。

- 对不同的子载波进行编码。这种编码用一个子载波上的大信噪比来补偿另一个子载波上的深衰落点。这种方法在 19.4.3 节中详细介绍。
- 将信号扩展到所有子载波上。在这种方法中,每一个符号都被扩展到所有的载波上,因此,信噪比为信号所在的所有子载波上的均值。该方法的详细介绍见 19.10 节和 19.11 节。
- 自适应调制。如果发射机知道每个子载波上的信噪比,那么它可以自适应地选择其调制符号集和编码率。因此,在低信噪比的载波上,发射机可以选择纠错能力更强的编码

^① 注意,在时不变的频率选择性信道中,独立的信道实现的个数取决于系统带宽和信道相关带宽的比值。如果该数值比较小,也许就没有一个足够大的总体来获得好的平均。

方式和更小的调制符号集。而且，每个子载波上的分配功率也可以不同。这种方法将在 19.8 节详细介绍。

19.4.3 编码正交频分复用

正如编码可以非常有效地提高单载波系统在衰落信道中的性能，将其用于正交频分复用系统也同样有效，但因为是在不同时间及不同频率发送数据，所以就产生了在正交频分复用系统中如何应用数据编码的问题。

为了对不同子载波上的编码有直观的认识，再次想象重复编码的简单情形：每一个发送符号在 K 个不同的子载波上重复。只要在不同子载波上的衰落相互独立，就可以实现 K 重分集。在最简单的情况下，接收机首先对每个子载波上的符号进行一个硬判决，然后对这 K 个与发送比特有关的接收符号进行大数判决。当然，实际的系统并不使用重复编码，但原理是相同的。

现在可以讨论正交频分复用系统编码的全部理论。不过，将时域和频域进行类比，问题会变得更简单。回忆 14.8 节的主要内容：必须采用足够的交织，以使编码比特的衰落相互独立。换句话说，只需要独立的信道状态来传输编码比特，这将自动产生一个高的分集阶数。信道状态无论是从信道的瞬时变化产生，还是作为频率选择性信道中子载波的不同传输函数，都没有关系。因此，没有必要为正交频分复用定义新的码字，但问题是如何设计合适的映射器和交织器，以在时频空间中分配不同的编码比特。这种映射取决于信道的时间选择性和频率选择性。如果信道具有高度频率选择性，那么也许仅在所用的频率上编码就足够了，而不用在时间轴上进行任何编码或交织。这样做有两个好处：一方面，这个方案同样适用于静态信道，静态信道在无线局域网和其他高速率数据传输中很常见；另一方面，在时域中不进行交织，会使传输和译码过程的时延更短。

图 19.7 给出了一个实例的性能曲线。可以看出，对于加性高斯白噪声信道，1/3 和 3/4 码率的编码系统均呈现良好的性能，仅有约 1 dB 的差别。然而在衰落信道中，性能却有显著的差异。1/3 码率编码有很好的分集，因此误比特率随着信噪比的增大而迅速下降，而 3/4 码率的编码系统仅有非常低的频率分集，因此性能很差。

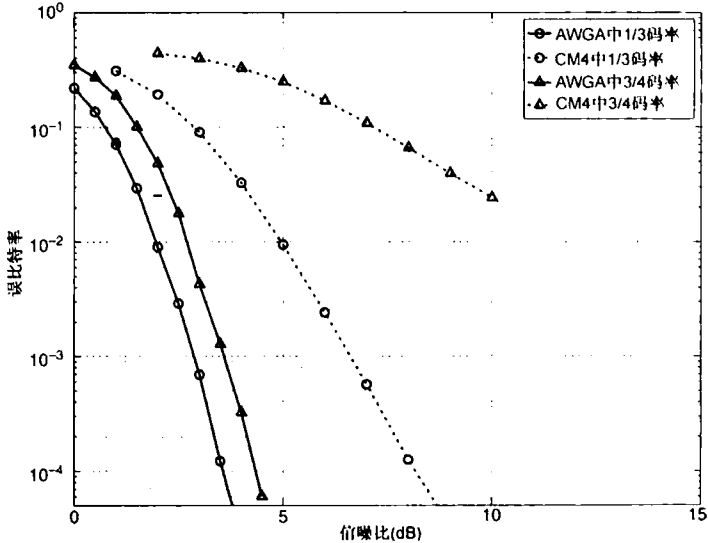


图 19.7 1/3 码率和 3/4 码率的编码正交频分复用系统的误比特率与信噪比的函数关系。信道为加性高斯白噪声信道或 IEEE802.15.3a 的 4 号信道模型。正交频分复用系统服从 WiMedia 标准的规定。引自 Ramachandran et al. [2004] © IEEE

19.5 信道估计

和任何其他相干无线系统一样,正交频分复用系统的正常工作都需要估计信道的传递函数或者信道的冲激响应(两者等价)。由于正交频分复用系统是在一组并行的窄带子载波上工作,我们自然要在频域对信道进行估计。更精确地说,要获得 N 个子载波的复信道增益。把信道衰减记为 $h_{n,i}$,其中 n 为表示子信道的下标, i 为表示时间的下标。假设知道信道衰减的统计特性和正交频分复用信号的结构,就能推导出性能良好的信道估计器。

接下来将讨论 3 种估计算法:(i)导频符号法,此算法主要适用于信道的初始估计;(ii)分散导频法,该算法可以跟踪信道随时间的变化;(iii)基于特征值分解的算法,此算法用来减小前两种算法的复杂度。

19.5.1 基于导频符号的方法

在正交频分复用中,最直接的信道估计就是指定的导频符号仅包含已知数据的情况。换句话说,每个子载波所携带的数据已知。该方法适用于在突发传输的起始阶段来获取信道的初始信息。将 i 时刻第 n 个子载波上的已知数据记为 $c_{n,i}$,可以得到信道的最小二乘(LS)估计:

$$h_{n,i}^{\text{LS}} = r_{n,i} / c_{n,i}$$

其中, $r_{n,i}$ 为第 n 个子信道的接收值。

通过考虑不同频率之间信道衰减的相关性,可以改进信道估计。将最小二乘估计值放入一个向量 $\mathbf{h}_i^{\text{LS}} = (h_{1,i}^{\text{LS}} \ h_{2,i}^{\text{LS}} \cdots h_{N,i}^{\text{LS}})^T$ 中,线性最小均方误差估计的相应向量变为

$$\mathbf{h}_i^{\text{LMMSE}} = \mathbf{R}_{hh}^{\text{LS}} \mathbf{R}_{h^{\text{LS}}h^{\text{LS}}}^{-1} \mathbf{h}_i^{\text{LS}} \quad (19.12)$$

其中, $\mathbf{R}_{hh}^{\text{LS}}$ 为信道增益与信道增益的最小二乘估计之间的互协方差矩阵, $\mathbf{R}_{h^{\text{LS}}h^{\text{LS}}}$ 为最小二乘估计的自协方差矩阵。假定每个子载波上有方差为 σ_n^2 的加性高斯白噪声, $\mathbf{R}_{hh}^{\text{LS}} = \mathbf{R}_{hh}$, 并且 $\mathbf{R}_{h^{\text{LS}}h^{\text{LS}}} = (\mathbf{R}_{hh} + \sigma^2 \mathbf{I})$ 。将信道衰减写成向量的形式: $\mathbf{h}_i = (h_{1,i} \ h_{2,i} \cdots h_{N,i})^T$, 可以确定如下关系式:

$$\mathbf{R}_{hh} = E\{\mathbf{h}_i \mathbf{h}_i^H\} = E\{\mathbf{h}_i^* \mathbf{h}_i^T\}^* \quad (19.13)$$

如果信道是广义平稳信道,那么上式与时间 i 无关。

这种估计方法给出了良好的估计值,但是当子载波数很大时,计算复杂度相当大:需要 N^2 次乘法运算,即估计每个信道增益需要 N^2 次乘法(假定所有的相关矩阵及其逆矩阵都已经预先计算出来了)。即使只在每次突发传输的起始阶段采用这种基于导频符号的估计,复杂度也是很高的。鉴于此,可采用其他一些次最佳方案,例如为了利用相邻子信道的互相关性,采用有限长度(远小于 N)的平滑有限冲激响应(FIR)滤波器对信道衰减的最小二乘估计值进行处理。

19.5.2 基于离散导频的方法

得到信道的初始估计之后,还需要跟踪信道随时间的变化。在这种情况下,应该做两件事情:(i)减少正交频分复用符号中的已知比特数(这样可以提高频谱效率);(ii)利用信道的时域相关性,即信道特性只随时间缓慢变化。跟踪信道的一种引人注目的方法是将导频符号分散在正交频分复用时频栅格上,如图 19.8 所示,其中导频间隔 N_f 个子载波和 N_t 个正交频分复用符号^①。

^① 在图示中采用的是矩形导频图案,但是也可采用其他导频图案。

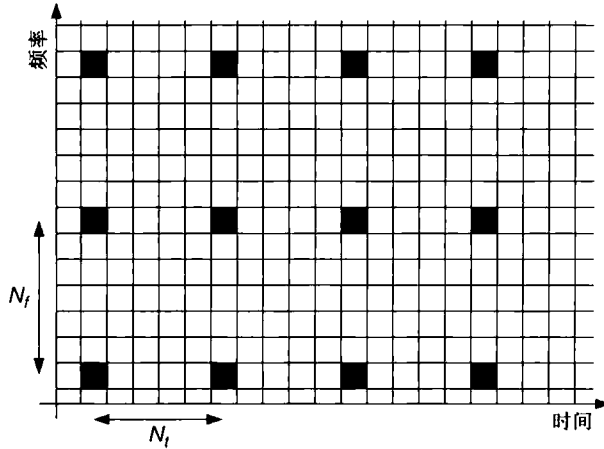


图 19.8 正交频分复用时频栅格上的离散导频。这里采用矩形图案,导频距离在频率上为 N_f 个子载波,在时间上为 N_t 个正交频分复用符号

当使用离散导频进行信道估计时,可以先在导频处进行信道的最小二乘估计,即 $h_{n,i}^{LS} = r_{n,i}/c_{n,i}$ 为接收信号的值, $c_{n,i}$ 为导频处 (n, i) 的已知导频信号数据。从导频位置的初始估计出发,主要是想利用插值来得到信道其他所有位置的估计。把导频看成一个二维空间的采样,可以利用标准采样定理,对所要求的导频图样密度加以限制[Nilsson et al. 1997]:

$$N_f < \frac{N}{N_{cp}}$$
$$N_t < \frac{1}{2(1 + N_{cp}/N)v_{max}}$$

由于不仅需要减小导频带来的噪声影响,还要降低估计算法的复杂度,所以有人认为一种比较好的折中方案是在每个传输方向上放置比采样定理所要求的多一倍的导频[Nilsson et al. 1997]。

从理论上讲,导频位置之间的信道插值可以利用与全导频符号估计理论相同的插值方法。当使用一个包括 K 个导频位置 $(n_j, i_j) (j = 1, \dots, K)$ 的集合来估计某个确定的信道衰减 $h_{n,i}$ 时,可将最小二乘估计的结果写成导频向量 $\mathbf{p} = (h_{n,i_1}^{LS} \quad h_{n,i_2}^{LS} \quad \dots \quad h_{n,i_K}^{LS})^T$, 并计算线性最小均方误差估计值:

$$h_{n,i}^{LMMSE} = \mathbf{r}_{hp} \mathbf{R}_{pp}^{-1} \mathbf{p}$$

其中, $\mathbf{r}_{hp}$ 为相关(行)向量 $E\{h_{n,i} \mathbf{p}^\dagger\}$, 且 $\mathbf{R}_{pp}$ 为 $E\{\mathbf{p} \mathbf{p}^\dagger\}$ 。这种估计器的复杂度随着估计中采用导频数目的增加而增加,每个衰减的估计需要 K 次乘法,这里仍假设所有相关矩阵及其逆矩阵已经预先计算好了。

在上面介绍的二维滤波法中,在频率上和时间上同时使用导频,另一种可选方案是使用分离滤波器。这意味着要使用两个一维滤波器,一个用在时域,另一个用在频域。因此在给定的估计器复杂度下,将会有更多的导频影响每个信道衰减的估计。在从通常的二维滤波器向基于两个一维滤波器的分离滤波器过度的优化过程中,其带来的性能提升超过了损失。

19.5.3 基于特征值分解的算法

正交频分复用的结构为高效的信道估计器结构提供了可能性。我们知道,在任何设计合理的系统中,信道冲激响应的长度要比正交频分复用符号的长度短。这一事实可以用于减小

估计问题的维数。从本质上讲,在采用式(19.12)中的线性最小均方误差估计器时,希望利用信道的统计特性使矩阵乘法运算的效率更高。这可以用估计理论中的最优降阶理论完成。根据估计理论,特征值分解(EVD) $\mathbf{R}_{xx} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ 会使式(19.12)的计算效率更高。空间的维数近似为 $N_{cp} + 1$, 即比循环前缀的样点数多 1。因此可以期望,在 $\mathbf{\Lambda}$ 的前 $N_{cp} + 1$ 个对角元素之后,幅值应该会快速下降。利用奇异值重写式(19.12),可得

$$\mathbf{h}^{\text{LMMSE}} = \mathbf{U}\mathbf{\Delta}\mathbf{U}^H\mathbf{h}^{\text{LS}}$$

其中, $\mathbf{\Delta}$ 为一个对角矩阵,其对角数值为 $\delta_i = \lambda_i / (\lambda_i + 1/\gamma)$ 。由于 λ_i 的作用,在前 $N_{cp} + 1$ 个对角元素之后, δ_i 会快速变小。通过将除了前 p 个之外的所有 λ_i 设为 0, 即当 $i > p$ 时,令 $\delta_i = 0$, 可以得到信道增益的最佳 p 阶估计器。估计器的计算复杂度为 $2Np$ 次乘法运算,即每个衰减系数的估计需 $2p$ 次乘法运算。可将其与原估计器式(19.12)中的每个衰减估计需 N 次乘法运算相比较。该估计器的原理如图 19.9 所示。

当自相关矩阵 $\mathbf{R}_{xx}$ 为一个循环矩阵时,所得到的最佳 $\mathbf{U}^H$ 变换和 $\mathbf{U}$ 变换分别为离散傅里叶逆变换和离散傅里叶变换,并且只有 N_{cp} 个非零奇异值。基本的估计器结构与图 19.10 所示的结构一致,同时正交频分复用接收机中已有的快速傅里叶变换处理器也可以用于信道估计。

在许多情况下,当信道自相关矩阵不是循环矩阵时,基于离散傅里叶变换的估计器的运算效率可能会超过降阶最佳估计器的运算效率。在进行二维估计时,这种估计器的通用结构也被用于两个一维估计器之一(见上文)。其中的好处在于可以在两个变换之间完成时域的平滑,这使得需要并行使用的滤波器数量更小。相比于最优估计器的 N 个滤波器(每个子载波一个), p 阶估计器只需要 p 个滤波器(见图 19.11)。

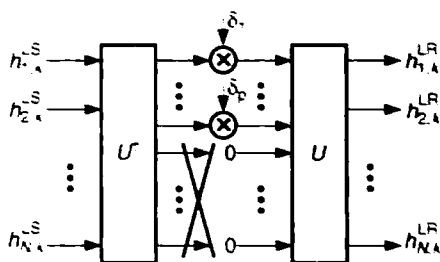


图 19.9 最佳 p 阶信道估计器, 可以看成是一个 $\mathbf{U}^H$ 变换之后接一个 p 级乘法和第二个 $\mathbf{U}$ 变换

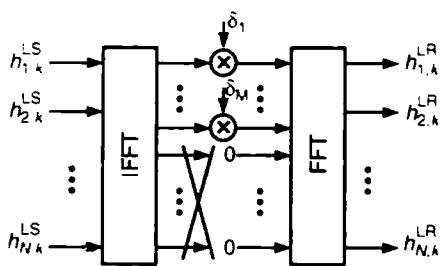


图 19.10 具有循环自相关函数, 并用快速傅里叶变换实现的信道低阶估计器

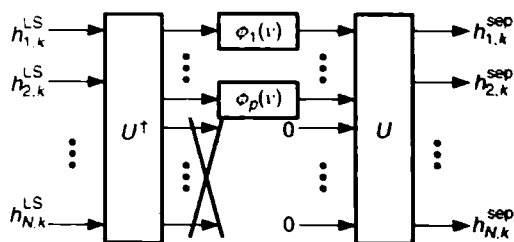


图 19.11 时域平滑在传输域完成的二维(可分离的)信道估计。这样可以将所需的并行滤波器的数量从 N 个减少到 p 个

19.6 峰平比

19.6.1 峰平比问题的起因

正交频分复用的主要问题之一就是发射信号的峰值幅度比平均幅度大得多。这个峰平比(PAR)问题由不同子载波上的 N 个正弦信号的叠加引起。平均来说,发射功率与 N 成正比。